Biochemistry **2004,** *43,* 7143−7150

Thermodynamic Parameters Based on a Nearest-Neighbor Model for DNA Sequences with a Single-Bulge Loop

Fumiaki Tanaka,*,‡ Atsushi Kameda,§ Masahito Yamamoto,‡,§ and Azuma Ohuchi‡,§

Graduate School of Engineering, Hokkaido University, North 13, West 8, Kita-ku, Sapporo 060-8628, Japan, and CREST, Japan Science and Technology Corporation, Tokyo, Japan

Received December 5, 2003; Revised Manuscript Received March 25, 2004

ABSTRACT: All 64 possible thermodynamic parameters for a single-bulge loop in the middle of a sequence were derived from optical melting studies. The relative stability of a single bulge depended on both the type of bulged base and its flanking base pairs. The contribution of the single bulge to helix stability ranged from 3.69 kcal/mol for a T**A**T bulge to −1.05 kcal/mol for an A**C**C bulge. Thermodynamics for 10 sequences with a G**T**G bulge were determined to test the applicability of the nearest-neighbor model to a single-bulge loop. Thermodynamic parameters for the G**T**G bulge and Watson−Crick base pairs predict ΔG°_{37}, ΔH°, ΔS°, and T_M(50 μM) values with average deviations of 3.0%, 4.3%, 4.7%, and 0.9 °C, respectively. The prediction accuracy was within the limits of what can be expected for a nearest-neighbor model. This certified that the thermodynamics for single-bulge loops can be estimated adequately using a nearest-neighbor model.

Engineers have tried to utilize DNA and its chemical reactions in several technological applications such as biosensors (*1*−*3*), in nanofabrications (*4*, *5*), and in computations (*6*, *7*). The success of these applications requires an accurate prediction of duplex formation. Therefore, an accurate prediction of duplex stability is more important than before because of its expanded applicability. The most useful strategy for prediction is modeling the duplex stabilities by thermodynamics based on the nearest-neighbor model. Nearest-neighbor thermodynamics have already been investigated for Watson−Crick base pairs (*8*, *9*), internal mismatches (*9*−*16*), and dangling ends (*17*, *18*). However, little is known about the thermodynamics of other DNA motifs, including bulge loops, internal loops, hairpins, and various loop structures. In this paper, we derive all 64 possible thermodynamic parameters for a single-bulge loop in DNA from optical melting studies. Single-bulge loops are one of the most common structures, where one unpaired nucleotide interrupts the double helix on only one strand. Some research estimated the thermodynamics of a single-bulge loop in DNA (*19*−*23*) and RNA (*23*−*25*), but these findings are not sufficient. More accurate thermodynamic parameters for single bulges facilitate prediction of a DNA secondary structure from a sequence and the design of DNA sequences without formation of unwanted structures.

EXPERIMENTAL PROCEDURES

Materials. Oligonucleotides were supplied by Hokkaido System Science and were synthesized using column purification. All oligonucleotides were dissolved in a buffer containing 1 M NaCl, 10 mM Na_2HPO_4, and 1 mM Na_2EDTA[1] with a pH of 7.0.

Melting Curves. Absorbance versus temperature profiles (melting curves) were measured at 260 and 320 nm with a heating rate of 1.0 °C/min on a Shimadzu UV-1650PC spectrophotometer. The absorbance at 260 nm was for the oligonucleotides, while that at 320 nm was for the background. Therefore, the melting curves were plotted by an absorbance difference ($\epsilon_{260} - \epsilon_{320}$) between the absorbance at 260 nm (ϵ_{260}) and that at 320 nm (ϵ_{320}). Oligonucleotide concentrations of each sample were determined by the absorbance difference between 260 and 320 nm using extinction coefficients calculated from the nucleotides and dinucleoside phosphates based on a nearest-neighbor approximation (*26*). The DNA sequence and its complementary sequence with or without a single bulge were mixed at a 1:1 concentration ratio. Oligonucleotide samples were hybridized by increasing the temperature to 90 °C for 10 min and lowering the temperature to 20 °C at heating rates of 5.0 and 1.0 °C/min, respectively.

Sequence Design. The sequences were designed to minimize the possibility of formation of an undesired hairpin or a slipped duplex using a program described previously (*27*). Furthermore, because our previous data showed that the duplex stability of bulge loops depends on the loop position (*28*), we inserted a single-bulge base into the middle of the sequence to produce the same effect of the loop position. The procedure to design the sequences was as follows.

* To whom correspondence should be addressed. Telephone: +81-11-716-2111, ext 6498. Fax: +81-11-706-7834. E-mail: fumiaki@dna-comp.org.

‡ Hokkaido University.

§ Japan Science and Technology Corp.

[1] Abbreviations: Na_2EDTA, disodium ethylenediaminetetraacetate; eu, entropy units (calories per kelvine per mole); SVD, singular-value decomposition; C_T, total strand concentration; T_M, melting temperature; NMR, nuclear magnetic resonance.

10.1021/bi036188r CCC: $27.50
Published on Web 05/07/2004

First, we chose even lengths of sequences and determined the flanking bases of the bulge loop in the middle of the sequence (e.g., NNNNNGANNNNN, where G and A are the flanking bases and N represents a not yet decided base). Then, the program described above designed the sequence that minimized the penalty of sequence by using the search method, called simulated annealing. The penalty was calculated by the sum of three terms, "similarity", "self-complementary", and "GC content". Similarity was defined as the maximum similarity between the sequence and its shifted sequence. Because a large similarity induces the formation of a slipped duplex between a sequence and its complementary sequence, a reduction of the penalty of similarity can reduce the possibility of forming an undesired slipped duplex. The second term, self-complementary, was defined as the maximum similarity between a sequence and its complementary sequence with and without the shift. In other words, self-complementary represents the maximum number of complementary base pairs between the sequence and itself. Therefore, to minimize the penalty of self-complementary is to reduce the possibility of formation of an undesired intramolecular structure (e.g., a hairpin structure). The third term, GC content, was calculated by the square value of the deviation from 50% GC content. This term was used to uniform the T_M of the sequence. Finally, we obtained the sequences for minimizing the penalty on sequences without a single-bulge loop (e.g., GGTCAGATACAC) and for those with a single bulge by inserting a bulged base between the flanking bases (e.g., GGTCAG**T**ATACAC, in which the bold base is a bulged base).

For more details of the process for minimizing the sequence penalty, refer to ref *27*.

Determination of Thermodynamic Parameters for Analyzed Sequences. Thermodynamic parameters were derived from absorbance versus temperature optical melting curves using plots of the reciprocal melting temperature (called a van't Hoff plot) according to the following equation (*29*)

$$T_M{}^{-1} = \frac{R}{\Delta H^\circ} \ln \frac{C_T}{\alpha} + \frac{\Delta S^\circ}{\Delta H^\circ} \qquad (1)$$

For self-complementary sequences, α in eq 1 equaled 1, and for non-self-complementary sequences, α equaled 4. The slope of the van't Hoff plot is $R/\Delta H^\circ$; the intercept is $\Delta S^\circ/\Delta H^\circ$, and the error in the thermodynamic parameters was estimated by the linearity of the regression line (*30*). A regression line was drawn by using the least-squares method with all data points weighted equally and with the error in $\ln C_T/4$ (all sequences in this paper are non-self-complementary) made negligible compared with the error in $1/T_M$. The melting temperature was estimated using the following procedure. First, the upper and lower baselines were determined by the regression line to the post-transition and pretransition domains, respectively. Second, the T_M was determined by the intersection point between the melting curve and the median line between the upper and lower baselines. The 10 T_M values of the ACACTCATCGATCTCGAGCT sequence with a standard deviation of 0.31 °C obtained by this method were more precise than those with a standard deviation of 0.62 °C obtained by calculating them from the maximum point of the first derivative of the absorbance versus temperature profile.

Although we presented the thermodynamics for the sequences by another method called the "curve fitting procedure", these values were not used for calculating nearest-neighbor thermodynamics (see the Discussion).

Calculation of the Watson−Crick Nearest-Neighbor Parameters. We derived Watson−Crick nearest-neighbor parameters for estimating the thermodynamic parameters for the single bulge from thermodynamic data combined with those from other laboratories to facilitate the estimation of thermodynamic parameters consistent with them. According to the nearest-neighbor model, the thermodynamics for the complementary duplex can be predicted with two initiation parameters, a symmetry parameter, and Watson−Crick nearest-neighbor parameters. Two initiation parameters contain contributions from the terminal base pairs (*9*). Thermodynamic contributions for the terminal GC pair are calculated with the term "init_GC", while those for the terminal AT pair are calculated with the term "init_AT". The symmetry parameters for the entropy and free energy are applied when the sequence is self-complementary. The entropy and free energy changes for self-complementary sequences were assumed to be −1.4 eu and 0.4 kcal/mol, respectively (*9*). The number of all possible thermodynamic parameters for the Watson−Crick nearest neighbor is 10. Therefore, we needed 12 parameters (two initiations and 10 Watson−Crick nearest-neighbor parameters) to predict the thermodynamics for the complementary duplex. The thermodynamics for the Watson−Crick nearest neighbor were determined by multiple linear regression using Octave (*31*). The thermodynamic parameters for 141 complementary duplexes from the Supporting Information for ref *9* (106 sequences), from our laboratory (31 sequences), and from others (*17*, *20*) (four sequences) were used to construct 141 simultaneous equations with 12 unknowns. These overdetermined simultaneous equations were solved by using singular-value decomposition (SVD), which minimizes the error-weighted squares of the residuals (*32*). The experimental uncertainties of ΔG°_{37}, ΔH°, and ΔS° were assumed to be 5, 10, and 10%, respectively. The errors in the 12 nearest-neighbor parameters were obtained from the variance−covariance matrix given by the SVD analysis propagated from the experimental uncertainties (*32*). A detailed procedure was described in the literature (*9*, *33*, *34*). The validity of our calculation procedure was checked by using data in the literature (*35*). The same nearest-neighbor thermodynamics were reproduced from thermodynamics for 45 sequences.

Determination of the Nearest-Neighbor Parameters for a Single Bulge. Throughout this paper, the nearest neighbors for the single bulges are represented, from left to right, by three letters indicating a flanking base in a 5′-end side, a bulged base, and a flanking base in a 3′-end side (e.g., C**A**T indicates a 5′-CAT-3′ sequence paired with a 3′-GA-5′ sequence in which the bold base is a bulged base). On the basis of the nearest-neighbor model, the thermodynamics of the sequence with the single bulge were derived like the following equation (only the top strand is given):

$$5'\text{-GAC}\mathbf{A}\text{TAT-3}' = \text{initiation} + \text{symmetry} + \text{GA} + \text{AC} + \text{C}\mathbf{A}\text{T} + \text{TA} + \text{AT} \qquad (2)$$

Table 1: Thermodynamics for Complementary Sequences[a]

sequence	$1/T_M$ vs ln $C_T/4$			fits of individual melting curves		
	$\Delta H°$ (kcal/mol)	$\Delta S°$ (eu)	$\Delta G°$ (kcal/mol)	$\Delta H°$ (kcal/mol)	$\Delta S°$ (eu)	$\Delta G°$ (kcal/mol)
CGCTATGAACCACTTG	−133.6 ± 5.2	−368.3 ± 15.3	−19.40 ± 0.44	−99.0 ± 5.0	−265.8 ± 14.4	−16.56 ± 0.57
GACCTAGACACCATTC	−129.6 ± 8.9	−356.5 ± 26.2	−18.99 ± 0.75	−111.4 ± 6.5	−302.8 ± 19.2	−17.46 ± 0.58
TTAGCAGAGTCCATCG	−116.1 ± 4.3	−314.2 ± 12.5	−18.63 ± 0.38	−108.9 ± 8.6	−293.0 ± 25.5	−18.06 ± 0.71
CACCATTCAGAGCCTA	−128.0 ± 4.8	−349.1 ± 14.2	−19.74 ± 0.44	−102.3 ± 4.2	−273.6 ± 12.8	−17.43 ± 0.36
GTCATATCCAGCGTTC	−152.5 ± 4.8	−424.1 ± 14.1	−20.97 ± 0.41	−103.5 ± 11.5	−279.5 ± 34.0	−16.85 ± 0.94
CTAAGGTCGTGTATCC	−137.9 ± 5.5	−382.3 ± 16.2	−19.32 ± 0.44	−98.4 ± 4.5	−264.8 ± 13.3	−16.23 ± 0.48
CGAAATGCTAGTGACC	−127.1 ± 3.2	−347.7 ± 9.6	−19.25 ± 0.28	−122.4 ± 6.7	−334.0 ± 19.8	−18.82 ± 0.54
GTCGTTACTGGGCTAT	−126.4 ± 2.5	−344.4 ± 7.4	−19.61 ± 0.23	−114.4 ± 5.1	−309.0 ± 14.9	−18.54 ± 0.49
AGCATTAGACGGACCT	−128.6 ± 5.6	−349.3 ± 16.3	−20.28 ± 0.53	−119.1 ± 3.8	−321.1 ± 11.0	−19.47 ± 0.38
CACTAAGGCTGTTCAC	−135.7 ± 5.8	−373.6 ± 17.1	−19.83 ± 0.49	−111.1 ± 1.9	−300.8 ± 5.1	−17.78 ± 0.31
CATCCGAGTTTACAGC	−130.0 ± 9.4	−356.6 ± 27.5	−19.41 ± 0.82	−117.8 ± 4.9	−320.5 ± 14.8	−18.35 ± 0.39
GACAGTAGCTCGGAAT	−135.5 ± 6.3	−372.1 ± 18.7	−20.08 ± 0.55	−103.7 ± 4.3	−278.3 ± 12.9	−17.43 ± 0.43
CTCAAATGGTACGCAG	−144.0 ± 4.6	−398.9 ± 13.5	−20.32 ± 0.40	−97.5 ± 2.7	−261.6 ± 7.6	−16.41 ± 0.41
GACTTTAGGAGCGATG	−128.0 ± 1.9	−351.7 ± 5.6	−18.94 ± 0.16	−110.9 ± 2.9	−301.2 ± 8.2	−17.52 ± 0.34
TCGATTAGGACACAGG	−127.0 ± 3.4	−348.6 ± 10.1	−18.89 ± 0.29	−111.3 ± 2.1	−302.2 ± 6.3	−17.60 ± 0.18
CTAGAGCGGATGACTT	−125.5 ± 3.1	−341.9 ± 9.2	−19.48 ± 0.28	−95.5 ± 3.2	−253.4 ± 9.1	−16.92 ± 0.49
GTCAACAGGCATAGCA	−121.6 ± 2.3	−331.1 ± 6.9	−18.92 ± 0.21	−109.6 ± 2.2	−295.6 ± 6.4	−17.92 ± 0.21
GACTTGTGGACGGTAT	−132.0 ± 4.2	−361.1 ± 12.4	−19.96 ± 0.38	−108.4 ± 2.8	−291.8 ± 7.9	−17.89 ± 0.35
ACCTGACGGAGACTAT	−112.5 ± 3.4	−304.6 ± 9.9	−17.99 ± 0.30	−96.2 ± 2.5	−256.6 ± 7.7	−16.65 ± 0.15
GATGCCAGGTCGATAA	−121.6 ± 3.0	−329.4 ± 8.7	−19.49 ± 0.28	−105.4 ± 3.3	−281.7 ± 9.4	−18.01 ± 0.38
GCGGTAAGGATCTAGT	−134.8 ± 3.3	−372.0 ± 9.8	−19.46 ± 0.28	−108.2 ± 3.1	−293.3 ± 9.0	−17.24 ± 0.33
GTATGACGGAACTGGT	−135.0 ± 4.1	−370.8 ± 12.1	−19.99 ± 0.36	−102.1 ± 2.6	−273.8 ± 7.3	−17.15 ± 0.34
GGTCAGATACAC	−93.6 ± 1.4	−259.1 ± 4.1	−13.19 ± 0.08	−85.2 ± 6.2	−233.7 ± 19.1	−12.76 ± 0.27
GCATTCTGTACCTCGA	−150.2 ± 13.2	−418.1 ± 39.0	−20.50 ± 1.09	−92.4 ± 8.8	−246.8 ± 27.1	−15.81 ± 0.45
TTGAAGATACGCTGGC	−123.4 ± 6.3	−335.6 ± 18.4	−19.34 ± 0.57	−114.5 ± 3.3	−309.2 ± 9.9	−18.60 ± 0.29
CTACCACTAAGCCTTG	−119.7 ± 4.4	−328.6 ± 13.0	−17.80 ± 0.36	−93.3 ± 2.5	−249.8 ± 7.4	−15.78 ± 0.28
CGATGTTCCAGTCA	−112.7 ± 1.9	−310.3 ± 5.6	−16.48 ± 0.14	−93.1 ± 1.6	−251.3 ± 4.6	−15.14 ± 0.19
GCTGAATGTCCTAG	−109.8 ± 2.9	−303.1 ± 8.6	−15.78 ± 0.20	−90.0 ± 4.7	−243.6 ± 13.6	−14.45 ± 0.44
GTACGCCTTGATTCTC	−137.7 ± 4.4	−380.5 ± 12.9	−19.70 ± 0.37	−92.8 ± 4.4	−247.3 ± 12.9	−16.11 ± 0.48
GTTGCGGTCTTACATG	−128.7 ± 6.0	−352.1 ± 17.6	−19.54 ± 0.54	−104.9 ± 2.2	−282.1 ± 6.4	−17.41 ± 0.22
CACGACTATCTGAACC	−123.5 ± 5.7	−337.6 ± 16.7	−18.74 ± 0.49	−122.3 ± 7.1	−334.1 ± 20.6	−18.65 ± 0.71

[a] Only the top strand is given (5′-end → 3′-end from left to right). Solutions include 1 M NaCl, 10 mM Na_2HPO_4, and 1 mM Na_2EDTA (pH 7.0).

Then, the contribution to the thermodynamics for the single bulge was derived using the following equation by rearranging eq 2.

$$C\mathbf{A}T = GAC\mathbf{A}TAT - \text{initiation} - \text{symmetry} - GA - AC - TA - AT \quad (3)$$

The number of all possible nearest neighbors for the single bulge (such as C**A**T in the example above) is 64. Thus, 64 thermodynamic parameters for a single bulge allow the prediction of the thermodynamics for all sequences with a single bulge. The frequency of the 64 nearest-neighbor parameters in this paper is 1, except as follows: 2 for A**T**A, 3 for A**C**A, 2 for A**G**A, 2 for T**A**T, 3 for T**C**T, 2 for T**G**T, 2 for C**A**T, 2 for C**T**T, 2 for C**C**T, 2 for C**G**T, 3 for G**A**G, 10 for G**T**G, 3 for G**C**G, and 3 for G**G**G. The thermodynamics of sequences with the G**T**G bulge were also used to test the applicability of the nearest-neighbor model to the single-bulge loop (see the Discussion).

RESULTS

Thermodynamic Parameters for Analyzed Sequences. Plots of T_M^{-1} versus ln $C_T/4$ were linear (coefficient of determination ≥ 0.98) over the entire 50-fold range of concentrations. Thermodynamic parameters with errors for complementary sequences and for sequences with a single bulge are shown in Tables 1 and 2, respectively. $\Delta H°$ values derived from plots of T_M^{-1} versus ln $C_T/4$ for 60 of 114 sequences disagreed by more than 15% with those derived from fits of individual melting curves. This implies that the transition in forming a duplex does not stringently proceed in a two-state manner (*34*). Therefore, we adopted thermodynamics derived from plots of T_M^{-1} versus ln $C_T/4$ because of their robustness (see the Discussion).

Thermodynamic Parameters for Watson−Crick Nearest Neighbors. The nearest-neighbor thermodynamic parameters with errors are shown in Table 3. The nearest-neighbor parameters for a Watson−Crick base pair derived from 141 sequences (110 from references and 31 from ours) predict $\Delta G°_{37}$, $\Delta H°$, and $\Delta S°$ values with average deviations of 4.2, 6.4, and 7.3%, respectively. These predictions are as good as those obtained from previous parameters (*8*, *9*) and are within the limits of what can be expected for a nearest-neighbor model.

Thermodynamic Parameters for Single Bulges. The contributions of single bulges to the nearest-neighbor thermodynamic parameters are listed in Table 4. The relative stability of a single bulge depended on both the type of bulged base and its flanking base pairs. The contribution of a single bulge to helix stability, $\Delta G°$, ranged from 3.69 kcal/mol for the T**A**T bulge to −1.05 kcal/mol for the A**C**C bulge.

Effect of the Bulged Base and the Flanking Base Pairs. Traditional studies presented data on the effects of the bulged base and the flanking base pairs on $\Delta G°_{37}$ (*19*, *22*, *36*). However, the data from these studies presented inconsistent results. LeBlanc *et al*. (*19*) concluded that bulged pyrimidines (i.e., T and C bulges) were more stable than bulged purines (i.e., A and G bulges), while Ke *et al*. (*22*) found that bulged purines were more stable than bulged pyrimidines. We assumed that this inconsistency was due to limited data. Therefore, we systematically investigated the effect of the

Table 2: Thermodynamics for Sequences with a Single Bulge[a]

sequence	$1/T_M$ vs ln $C_T/4$			fits of individual melting curves		
	$\Delta H°$ (kcal/mol)	$\Delta S°$ (eu)	$\Delta G°$ (kcal/mol)	$\Delta H°$ (kcal/mol)	$\Delta S°$ (eu)	$\Delta G°$ (kcal/mol)
CGCTATGA**A**ACCACTTG	−125.0 ± 5.3	−349.5 ± 15.9	−16.60 ± 0.35	−96.4 ± 3.9	−263.0 ± 12.0	−14.79 ± 0.26
GACCTAGA**A**CACCATTC	−106.6 ± 6.2	−295.7 ± 18.7	−14.87 ± 0.38	−108.0 ± 6.0	−300.1 ± 18.7	−14.94 ± 0.25
TTAGCAGA**A**GTCCATCG	−117.5 ± 6.3	−325.2 ± 18.9	−16.65 ± 0.45	−99.3 ± 4.1	−270.6 ± 12.2	−15.40 ± 0.39
CACGACTA**A**TCTGAACC	−133.9 ± 6.6	−375.5 ± 19.9	−17.40 ± 0.45	−116.7 ± 7.7	−321.2 ± 18.9	−17.09 ± 2.03
CGCTATGA**C**ACCACTTG	−120.9 ± 6.4	−338.6 ± 19.3	−15.86 ± 0.41	−104.3 ± 14.2	−289.0 ± 43.9	−14.69 ± 0.66
GACCTAGA**C**CACCATTC	−145.9 ± 4.7	−412.8 ± 14.2	−17.87 ± 0.31	−113.1 ± 3.3	−313.9 ± 9.7	−15.73 ± 0.37
TTAGCAGA**C**GTCCATCG	−118.5 ± 5.4	−328.0 ± 16.1	−16.82 ± 0.39	−103.9 ± 7.3	−283.9 ± 22.2	−15.84 ± 0.40
CACGACTA**C**TCTGAACC	−123.8 ± 1.9	−346.2 ± 5.7	−16.45 ± 0.13	−111.7 ± 5.1	−309.7 ± 15.9	−15.67 ± 0.22
CGCTATGA**G**ACCACTTG	−123.5 ± 7.5	−347.5 ± 22.7	−15.76 ± 0.45	−97.1 ± 3.7	−267.5 ± 11.3	−14.18 ± 0.33
GACCTAGA**G**CACCATTC	−124.0 ± 2.3	−349.9 ± 7.0	−15.49 ± 0.14	−104.6 ± 4.7	−291.4 ± 14.1	−14.28 ± 0.36
TTAGCAGA**G**GTCCATCG	−111.6 ± 6.6	−306.2 ± 19.6	−16.60 ± 0.49	−107.1 ± 3.3	−293.1 ± 10.1	−16.24 ± 0.19
CACGACTA**G**TCTGAACC	−115.1 ± 2.8	−320.3 ± 8.5	−15.77 ± 0.18	−105.6 ± 2.8	−291.7 ± 8.5	−15.17 ± 0.15
CGCTATGA**T**ACCACTTG	−122.5 ± 3.4	−343.4 ± 10.3	−16.05 ± 0.22	−99.9 ± 2.0	−274.8 ± 6.1	−14.68 ± 0.22
GACCTAGA**T**CACCATTC	−121.6 ± 4.3	−341.5 ± 13.0	−15.65 ± 0.26	−102.6 ± 6.3	−284.0 ± 19.7	−14.54 ± 0.24
TTAGCAGA**T**GTCCATCG	−111.7 ± 3.1	−307.7 ± 9.3	−16.32 ± 0.22	−105.5 ± 11.4	−289.4 ± 35.1	−15.76 ± 0.48
CACGACTA**T**TCTGAACC	−139.9 ± 16.5	−392.5 ± 49.3	−18.14 ± 1.19	−116.6 ± 2.1	−322.9 ± 5.6	−16.47 ± 0.36
CACCATTC**A**AGAGCCTA	−128.6 ± 8.2	−357.9 ± 24.6	−17.63 ± 0.60	−100.7 ± 3.8	−274.1 ± 12.0	−15.65 ± 0.08
GTCATATC**A**CAGCGTTC	−116.7 ± 3.3	−326.7 ± 10.1	−15.40 ± 0.20	−102.7 ± 7.6	−284.1 ± 22.7	−14.56 ± 0.57
CTAAGGTC**A**GTGTATCC	−118.1 ± 2.8	−331.6 ± 8.5	−15.22 ± 0.17	−96.9 ± 7.6	−267.7 ± 23.3	−13.87 ± 0.43
CGAAATGC**A**TAGTGACC	−127.2 ± 2.6	−356.6 ± 7.7	−16.63 ± 0.17	−114.1 ± 1.7	−317.1 ± 5.3	−15.71 ± 0.09
GTCGTTAC**A**TGGGCTAT	−126.4 ± 7.1	−353.0 ± 21.4	−16.94 ± 0.51	−103.6 ± 3.8	−284.4 ± 11.7	−15.37 ± 0.18
CACCATTC**C**AGAGCCTA	−108.4 ± 5.0	−297.9 ± 14.9	−16.00 ± 0.36	−103.2 ± 1.7	−282.2 ± 5.1	−15.68 ± 0.17
GTCATATC**C**CAGCGTTC	−123.6 ± 2.2	−344.9 ± 6.7	−16.62 ± 0.15	−101.8 ± 3.7	−279.3 ± 10.9	−15.17 ± 0.34
CTAAGGTC**C**GTGTATCC	−106.1 ± 3.5	−294.2 ± 10.5	−14.86 ± 0.22	−105.1 ± 3.6	−291.2 ± 11.0	−14.85 ± 0.18
CGAAATGC**C**TAGTGACC	−111.5 ± 4.1	−308.0 ± 12.1	−16.00 ± 0.29	−114.3 ± 2.2	−316.3 ± 6.6	−16.19 ± 0.18
GTCGTTAC**C**TGGGCTAT	−132.2 ± 2.4	−368.1 ± 7.1	−18.06 ± 0.18	−108.1 ± 2.1	−296.2 ± 5.9	−16.26 ± 0.28
CACCATTC**G**AGAGCCTA	−128.9 ± 4.9	−360.9 ± 14.7	−16.95 ± 0.34	−98.3 ± 3.0	−269.0 ± 8.7	−14.90 ± 0.33
GTCATATC**G**CAGCGTTC	−125.4 ± 4.5	−353.3 ± 13.5	−15.84 ± 0.27	−96.9 ± 8.8	−266.7 ± 27.2	−14.21 ± 0.40
CTAAGGTC**G**GTGTATCC	−131.6 ± 8.1	−370.2 ± 24.4	−16.78 ± 0.52	−102.6 ± 1.0	−282.9 ± 3.0	−14.92 ± 0.22
CGAAATGC**G**TAGTGACC	−115.4 ± 2.4	−318.9 ± 7.2	−16.51 ± 0.17	−105.2 ± 4.2	−288.1 ± 12.8	−15.83 ± 0.26
GTCGTTAC**G**TGGGCTAT	−128.4 ± 7.7	−358.4 ± 23.1	−17.25 ± 0.54	−103.7 ± 1.7	−284.3 ± 5.1	−15.54 ± 0.24
CACCATTC**T**AGAGCCTA	−132.9 ± 10.6	−373.0 ± 32.0	−17.24 ± 0.72	−96.9 ± 4.6	−264.5 ± 14.6	−14.90 ± 0.21
GTCATATC**T**CAGCGTTC	−135.5 ± 4.3	−383.6 ± 13.0	−16.56 ± 0.27	−99.0 ± 3.9	−272.9 ± 11.6	−14.37 ± 0.35
CTAAGGTC**T**GTGTATCC	−119.9 ± 4.3	−337.4 ± 13.2	−15.26 ± 0.25	−99.3 ± 6.8	−275.2 ± 20.7	−13.97 ± 0.39
CGAAATGC**T**TAGTGACC	−130.7 ± 6.7	−365.6 ± 20.2	−17.29 ± 0.46	−116.0 ± 2.2	−321.5 ± 6.6	−16.28 ± 0.22
GTCGTTAC**T**TGGGCTAT	−119.6 ± 3.1	−331.9 ± 9.4	−16.72 ± 0.22	−103.8 ± 6.4	−284.0 ± 19.6	−15.70 ± 0.37
AGCATTAG**A**ACGGACCT	−121.6 ± 5.7	−336.0 ± 17.0	−17.37 ± 0.43	−105.3 ± 9.7	−287.0 ± 29.6	−16.28 ± 0.53
CACTAAGG**A**CTGTTCAC	−121.8 ± 6.5	−342.2 ± 19.7	−15.70 ± 0.41	−105.7 ± 12.3	−294.0 ± 38.1	−14.56 ± 0.49
CTCAAATG**A**GTACGCAG	−119.7 ± 4.3	−334.1 ± 13.1	−16.12 ± 0.28	−93.0 ± 5.2	−253.1 ± 16.0	−14.47 ± 0.29
TCGATTAG**A**GACACAGG	−119.7 ± 3.8	−335.0 ± 11.5	−15.81 ± 0.25	−106.5 ± 4.8	−295.3 ± 14.3	−14.95 ± 0.33
CTAGAGCG**A**GATGACTT	−146.9 ± 7.3	−414.2 ± 22.0	−18.45 ± 0.51	−96.4 ± 6.4	−262.3 ± 20.0	−15.09 ± 0.26
CATCCGAG**A**TTTACAGC	−129.4 ± 5.1	−364.2 ± 15.5	−16.46 ± 0.33	−110.6 ± 4.4	−307.2 ± 13.7	−15.29 ± 0.23
AGCATTAG**C**ACGGACCT	−116.9 ± 2.7	−323.3 ± 8.1	−16.60 ± 0.19	−99.6 ± 10.9	−271.1 ± 33.4	−15.50 ± 0.57
GACAGTAG**C**CTCGGAAT	−115.6 ± 1.2	−320.2 ± 3.5	−16.32 ± 0.08	−105.5 ± 2.5	−289.8 ± 7.6	−15.65 ± 0.22
CTCAAATG**C**GTACGCAG	−93.7 ± 2.0	−258.3 ± 6.2	−13.63 ± 0.12	−85.5 ± 7.9	−233.6 ± 24.7	−13.07 ± 0.30
TCGATTAG**C**GACACAGG	−103.5 ± 3.4	−285.7 ± 10.3	−14.84 ± 0.22	−109.6 ± 3.3	−304.3 ± 10.5	−15.25 ± 0.16
CTAGAGCG**C**GATGACTT	−124.1 ± 3.9	−344.8 ± 11.7	−17.19 ± 0.29	−95.8 ± 3.1	−259.9 ± 8.6	−15.15 ± 0.42
CATCCGAG**C**TTTACAGC	−123.1 ± 5.5	−344.7 ± 16.7	−16.23 ± 0.37	−107.6 ± 2.3	−297.9 ± 7.4	−15.23 ± 0.10
AGCATTAG**G**ACGGACCT	−112.1 ± 5.4	−307.2 ± 16.0	−16.83 ± 0.42	−105.5 ± 5.0	−287.4 ± 15.4	−16.38 ± 0.28
GACAGTAG**G**CTCGGAAT	−117.7 ± 2.1	−326.4 ± 6.1	−16.44 ± 0.15	−101.5 ± 1.5	−278.0 ± 4.7	−15.29 ± 0.14
CTCAAATG**G**GTACGCAG	−120.0 ± 5.4	−333.5 ± 16.3	−16.54 ± 0.38	−95.9 ± 5.1	−261.1 ± 15.1	−14.88 ± 0.45
TCGATTAG**G**GACACAGG	−97.8 ± 5.6	−267.8 ± 17.0	−14.75 ± 0.36	−106.7 ± 6.8	−294.7 ± 21.2	−15.26 ± 0.29
CTAGAGCG**G**GATGACTT	−124.5 ± 5.5	−344.9 ± 16.5	−17.48 ± 0.41	−95.2 ± 4.7	−257.6 ± 13.4	−15.32 ± 0.57
GCATTCTG**G**TACCTCGA	−118.5 ± 6.9	−329.9 ± 20.6	−16.14 ± 0.47	−96.9 ± 1.8	−265.0 ± 5.2	−14.75 ± 0.28
GGTCAG**T**ATACAC	−91.0 ± 2.3	−260.0 ± 7.3	−10.31 ± 0.06	−61.7 ± 11.9	−169.0 ± 37.4	−9.29 ± 0.31
CACTAAGG**T**CTGTTCAC	−129.3 ± 10.4	−363.9 ± 31.5	−16.39 ± 0.66	−109.6 ± 9.3	−304.5 ± 28.5	−15.12 ± 0.46
CTCAAATG**T**GTACGCAG	−132.2 ± 5.0	−370.7 ± 14.9	−17.20 ± 0.34	−97.1 ± 4.4	−265.1 ± 13.5	−14.87 ± 0.34
GACTTTAG**T**GAGCGATG	−119.2 ± 1.8	−332.8 ± 5.4	−16.03 ± 0.12	−105.7 ± 2.3	−292.0 ± 6.8	−15.10 ± 0.19
TCGATTAG**T**GACACAGG	−123.7 ± 3.3	−346.4 ± 9.9	−16.26 ± 0.21	−108.4 ± 0.8	−300.1 ± 2.6	−15.30 ± 0.08
CTAGAGCG**T**GATGACTT	−124.8 ± 1.6	−346.9 ± 4.7	−17.19 ± 0.11	−98.7 ± 8.5	−268.6 ± 25.8	−15.40 ± 0.54
GTCAACAG**T**GCATAGCA	−113.7 ± 7.6	−315.1 ± 22.7	−16.01 ± 0.52	−98.0 ± 2.1	−267.9 ± 6.4	−14.93 ± 0.10
GACTTGTG**T**GACGGTAT	−119.1 ± 2.5	−330.6 ± 7.4	−16.57 ± 0.17	−105.2 ± 2.5	−288.9 ± 7.5	−15.63 ± 0.25
ACCTGACG**T**GAGACTAT	−104.1 ± 4.8	−287.0 ± 14.3	−15.10 ± 0.31	−85.8 ± 3.0	−232.0 ± 9.2	−13.88 ± 0.15
GATGCCAG**T**GTCGATAA	−114.8 ± 2.0	−315.2 ± 5.8	−17.03 ± 0.15	−107.7 ± 2.5	−294.1 ± 7.3	−16.50 ± 0.26
GCGGTAAG**T**GATCTAGT	−118.3 ± 2.3	−330.5 ± 6.8	−15.77 ± 0.14	−106.9 ± 7.6	−295.8 ± 23.1	−15.13 ± 0.44
GTATGACG**T**GAACTGGT	−122.6 ± 4.2	−341.6 ± 12.6	−16.63 ± 0.29	−103.9 ± 4.3	−285.2 ± 13.2	−15.41 ± 0.29
GCATTCTG**T**TACCTCGA	−115.5 ± 2.7	−321.2 ± 8.0	−15.90 ± 0.18	−95.7 ± 2.3	−261.2 ± 6.8	−14.70 ± 0.27
TTGAAGAT**A**ACGCTGGC	−105.2 ± 3.7	−288.2 ± 11.0	−15.80 ± 0.27	−103.9 ± 2.7	−284.2 ± 7.8	−15.73 ± 0.28
GTGTAT**A**CTGACC	−85.9 ± 3.6	−245.7 ± 11.5	−9.67 ± 0.08	−60.1 ± 18.7	−165.8 ± 59.4	−8.68 ± 0.38
GCTGAAT**A**GTCCTAG	−84.2 ± 7.0	−233.2 ± 21.5	−11.82 ± 0.31	−85.1 ± 2.4	−236.2 ± 7.5	−11.89 ± 0.11
GTACGCCT**A**TGATTCTC	−115.2 ± 3.8	−322.1 ± 11.4	−15.29 ± 0.23	−92.7 ± 7.3	−253.7 ± 22.6	−14.06 ± 0.35
CTACCACT**C**AAGCCTTG	−115.4 ± 9.8	−323.3 ± 29.8	−15.10 ± 0.59	−97.6 ± 8.8	−269.3 ± 26.3	−14.09 ± 0.68

Table 2 (Continued)

sequence	$1/T_M$ vs ln $C_T/4$			fits of individual melting curves		
	$\Delta H°$ (kcal/mol)	$\Delta S°$ (eu)	$\Delta G°$ (kcal/mol)	$\Delta H°$ (kcal/mol)	$\Delta S°$ (eu)	$\Delta G°$ (kcal/mol)
GTTGCGGT**C**CTTACATG	−137.1 ± 7.1	−384.1 ± 21.4	−17.95 ± 0.50	−107.7 ± 3.0	−296.3 ± 8.4	−15.85 ± 0.38
GCTGAAT**C**GTCCTAG	−99.9 ± 1.8	−280.6 ± 5.6	−12.86 ± 0.08	−90.4 ± 6.1	−251.5 ± 18.7	−12.38 ± 0.31
GTACGCCT**C**TGATTCTC	−124.4 ± 4.3	−349.6 ± 12.9	−15.94 ± 0.27	−93.1 ± 3.0	−255.1 ± 9.2	−14.01 ± 0.25
CTACCACT**G**AAGCCTTG	−129.3 ± 1.5	−366.4 ± 4.7	−15.71 ± 0.09	−88.6 ± 2.7	−242.7 ± 8.7	−13.36 ± 0.13
CGATGTT**G**CCAGTCA	−99.6 ± 4.4	−279.0 ± 13.7	−13.03 ± 0.21	−84.8 ± 3.5	−233.5 ± 11.0	−12.36 ± 0.13
GCTGAAT**G**GTCCTAG	−108.4 ± 4.4	−305.0 ± 13.5	−13.85 ± 0.23	−83.9 ± 10.4	−229.5 ± 32.6	−12.69 ± 0.25
GTACGCCT**G**TGATTCTC	−124.3 ± 6.7	−349.8 ± 20.4	−15.80 ± 0.41	−88.5 ± 2.1	−241.2 ± 6.6	−13.66 ± 0.23
TTGAAGAT**T**ACGCTGGC	−123.2 ± 2.8	−342.7 ± 8.3	−16.94 ± 0.19	−97.8 ± 1.9	−264.4 ± 5.5	−15.85 ± 0.64
GTTGCGGT**T**CTTACATG	−113.1 ± 3.3	−312.2 ± 9.8	−16.28 ± 0.24	−106.2 ± 2.0	−291.4 ± 6.0	−15.79 ± 0.17
GCTGAAT**T**GTCCTAG	−92.2 ± 3.0	−256.1 ± 9.1	−12.80 ± 0.15	−90.8 ± 3.5	−251.7 ± 10.4	−12.79 ± 0.29
GTACGCCT**T**TGATTCTC	−128.8 ± 7.8	−360.7 ± 23.3	−16.98 ± 0.54	−94.3 ± 3.5	−256.7 ± 10.1	−14.69 ± 0.40

[a] Only the top strand is given (5′-end → 3′-end from left to right). Solutions include 1 M NaCl, 10 mM Na_2HPO_4, and 1 mM Na_2EDTA (pH 7.0).

Table 3: Nearest-Neighbor Thermodynamic Parameters for Watson−Crick Base Pairs[a]

sequence	$\Delta H°$ (kcal/mol)	$\Delta S°$ (eu)	$\Delta G°$ (kcal/mol)	$\Delta G°$ (Allawi *et al.*)[b] (kcal/mol)
AA/TT	−8.3 ± 1.1	−23.3 ± 3.1	−1.04 ± 0.07	−1.00 ± 0.01
AT/TA	−8.5 ± 0.8	−24.6 ± 2.4	−0.91 ± 0.04	−0.88 ± 0.04
AC/TG	−9.5 ± 1.1	−26.0 ± 2.7	−1.49 ± 0.09	−1.44 ± 0.04
AG/TC	−8.2 ± 0.9	−22.3 ± 2.7	−1.33 ± 0.06	−1.28 ± 0.03
TA/AT	−6.5 ± 0.6	−19.0 ± 1.7	−0.60 ± 0.04	−0.58 ± 0.06
TC/AG	−8.7 ± 0.8	−23.6 ± 2.4	−1.35 ± 0.03	−1.30 ± 0.03
TG/AC	−8.9 ± 0.4	−24.3 ± 1.3	−1.51 ± 0.05	−1.45 ± 0.06
CC/GG	−8.8 ± 0.6	−21.7 ± 1.7	−1.90 ± 0.04	−1.84 ± 0.04
CG/GC	−11.4 ± 0.9	−29.7 ± 2.5	−2.25 ± 0.06	−2.17 ± 0.05
GC/CG	−10.1 ± 1.0	−25.4 ± 2.6	−2.28 ± 0.06	−2.24 ± 0.03
init_AT	4.4 ± 1.4	10.7 ± 3.9	1.19 ± 0.09	1.03 ± 0.05
init_GC	1.7 ± 0.5	2.4 ± 1.3	1.12 ± 0.03	0.98 ± 0.05
symmetry	0 ± −	−1.4 ± −	0.4 ± −	0.4 ± −

[a] These were derived from thermodynamics for complementary sequences from the literature and plots of $1/T_M$ vs ln $C_T/4$ in Table 1. Note that thermodynamics from fits of individual melting curves in Table 1 were not used (see the Discussion). [b] These parameters presented in the literature (*9*) are shown for comparison.

bulged base and the flanking base pairs on $\Delta G°_{37}$. The contribution of each type of bulged base to $\Delta G°_{37}$ is revealed in Figure 1. Our comparison shows no significant differences between each bulged base.

The contribution of each type of flanking base pair is shown in Figure 2. Papanicolaou *et al*. (*36*) showed that a single bulge between two pyrimidines was observed to be less stable than others because of a poor stacking between the pyrimidines. However, our results do not reveal this tendency. The type of flanking base pair also shows no significant differences, although we may assume the contribution of flanking base pairs has a more profound effect than that of a bulged base.

$\Delta G°_{37}$ for a single bulge cannot be classified by the type of bulged base or flanking base pairs despite the previous trial, as illustrated in Figures 1 and 2. Therefore, $\Delta G°_{37}$ for a single bulge should be approximated at least on the basis of both the type of bulged base and flanking base pairs.

Thus, $\Delta G°_{37}$ for a single bulge is complicated, and this complexity is consistent with the results of conformational studies. From the results of nuclear magnetic resonance (NMR) studies, the experimental data indicated that a single-bulge loop in solution exists in various conformations, depending on the sequence context and temperature (*21*, *37*−*42*). For example, a G**T**G bulge is in a looped-out state at a low temperature or in a stacked state at a high temperature, while a C**T**C bulge is looped out independent of the temperature (*40*). Because the conformations of a duplex are related to their thermodynamics, the complexity of the conformation for a single bulge probably results in increased complexity in the thermodynamics. Therefore, because the conformation for the single bulge depended at least on both the bulged base and the flanking base pairs, we concluded that $\Delta G°_{37}$ for a single bulge depended on the bulged base and the flanking base pairs.

DISCUSSION

Applicability of the Nearest-Neighbor Model to Single-Bulge Loops. According to the nearest-neighbor model, the contribution of the thermodynamics for a nearest neighbor (such as an AC or A**C**T) is independent of the effect of nonadjacent base pairs. If this nearest-neighbor model is applicable to sequences with a single bulge, we can approximately estimate the thermodynamics for sequences containing the same single-bulge nearest neighbor but with different sequences by using a unique parameter for the single-bulge nearest neighbor. Thus, to test the applicability of the nearest-neighbor model to the sequences with a single-bulge loop, we determined the thermodynamics for 10 sequences with a G**T**G bulge. Thermodynamic parameters for a G**T**G bulge listed in Table 3 and those for Watson−Crick base pairs listed in Table 4 predict $\Delta G°_{37}$, $\Delta H°$, $\Delta S°$, and T_M(50 μM) values with average deviations of 3.0%, 4.3%, 4.7%, and 0.9 °C, respectively. The prediction accuracy was within the limits of what can be expected for a nearest-neighbor model. This certified that the thermodynamics for single-bulge loops can be estimated adequately using a nearest-neighbor model.

Validity of the Estimation of Thermodynamic Parameters Based on Two-State Approximation. Analysis of the van't Hoff plot is based on the two-state transition (i.e., random coil ↔ duplex) (*34*). If this two-state model is valid, $\Delta H°$ values from plots of $1/T_M$ versus ln $C_T/4$ generally agree with those from fits of individual melting curves (*19*, *33*, *43*) within 15%, although the deviations of 60 of 114 sequences in this study were greater than 15%. It follows that all our experimental data do not stringently proceed in a two-state manner. However, in cases where $\Delta H°$ values from these two methods marginally agree, the data from plots of $1/T_M$ versus ln $C_T/4$ appear to be more reliable (*44*) because a comparison with $\Delta H°$ values determined by using the model-independent calorimetry revealed T_M was relatively insensitive to non-two-state transitions (*45*). Another piece of

Table 4: Nearest-Neighbor Thermodynamic Parameters for Single Bulges[a]

sequence	ΔH° (kcal/mol)	ΔS° (eu)	ΔG° (kcal/mol)	sequence	ΔH° (kcal/mol)	ΔS° (eu)	ΔG° (kcal/mol)
A**A**A	−4.0 ± 7.8	−17.5 ± 22.2	1.66 ± 1.05	C**A**A	−14.4 ± 6.2	−46.6 ± 8.9	−0.11 ± 0.69
A**A**T	−13.5 ± 3.4	−44.6 ± 7.1	0.35 ± 0.61	C**A**T	−7.0 ± 3.9	−25.8 ± 8.9	1.06 ± 0.71
A**A**C	10.2 ± 4.5	25.8 ± 10.4	1.95 ± 0.93	C**A**C	4.3 ± 2.9	8.8 ± 6.2	1.73 ± 0.98
A**A**G	−0.2 ± 4.0	−4.6 ± 8.9	1.24 ± 1.10	C**A**G	−2.9 ± 4.8	−13.0 ± 14.7	0.83 ± 1.08
A**T**A[b]	9.8 ± 8.9	24.7 ± 11.4	2.89 ± 1.03	C**T**A	−18.7 ± 4.8	−61.7 ± 14.8	0.28 ± 1.17
A**T**T	−19.5 ± 2.7	−61.6 ± 8.2	−0.39 ± 1.51	C**T**T	−4.6 ± 5.7	−17.0 ± 12.0	0.83 ± 0.93
A**T**C	−4.8 ± 4.8	−20.0 ± 25.3	1.17 ± 0.84	C**T**C	−14.5 ± 5.6	−48.1 ± 17.2	0.57 ± 0.47
A**T**G	5.6 ± 3.8	12.9 ± 5.6	1.57 ± 0.93	C**T**G	−4.7 ± 6.5	−18.8 ± 11.7	0.79 ± 1.02
A**C**A[b,c]	6.1 ± 6.1	14.0 ± 25.1	2.51 ± 0.71	C**C**A	5.8 ± 3.8	13.4 ± 7.7	1.52 ± 0.89
A**C**T	−3.4 ± 7.4	−15.3 ± 11.1	1.30 ± 0.58	C**C**T	−5.3 ± 6.3	−24.8 ± 16.9	0.81 ± 1.07
A**C**C	−29.1 ± 7.5	−91.3 ± 15.9	−1.05 ± 0.97	C**C**C	−2.6 ± 5.8	−9.4 ± 15.3	0.51 ± 0.83
A**C**G	−1.2 ± 4.7	−7.4 ± 10.7	1.07 ± 0.94	C**C**G	9.1 ± 7.1	24.4 ± 19.6	1.19 ± 0.96
A**G**A[b]	15.5 ± 16.9	43.1 ± 7.6	3.33 ± 0.95	C**G**A	−14.7 ± 1.5	−49.6 ± 3.3	0.57 ± 0.98
A**G**T	5.3 ± 6.5	10.6 ± 17.2	1.98 ± 0.99	C**G**T	2.1 ± 3.8	6.5 ± 7.9	1.00 ± 0.31
A**G**C	−7.2 ± 5.2	−28.4 ± 32.6	1.33 ± 1.03	C**G**C	−4.4 ± 8.6	−17.8 ± 30.4	1.29 ± 0.71
A**G**G	5.7 ± 4.7	14.4 ± 20.6	1.29 ± 0.86	C**G**G	−16.4 ± 10.4	−51.6 ± 24.1	−0.73 ± 0.91
T**A**A	15.3 ± 8.8	38.5 ± 12.6	3.46 ± 0.93	G**A**A	−6.8 ± 6.9	−25.2 ± 18.2	0.96 ± 1.07
T**A**T[b]	19.8 ± 3.5	56.7 ± 20.5	3.69 ± 0.92	G**A**T	−9.8 ± 2.3	−35.6 ± 4.4	1.19 ± 0.55
T**A**C	−2.3 ± 5.6	−12.7 ± 16.1	1.67 ± 0.94	G**A**C	−4.8 ± 2.7	−18.3 ± 5.3	0.98 ± 0.45
T**A**G	15.0 ± 3.9	40.4 ± 8.7	2.36 ± 0.80	G**A**G	−6.5 ± 7.3	−21.0 ± 8.8	0.52 ± 0.80
T**T**A	−2.7 ± 11.2	−16.0 ± 14.6	2.32 ± 0.89	G**T**A	−7.4 ± 2.2	−27.0 ± 19.9	1.03 ± 1.16
T**T**T	−8.2 ± 6.1	−30.1 ± 20.8	0.96 ± 0.94	G**T**T	1.8 ± 3.9	−0.6 ± 4.3	1.99 ± 1.03
T**T**C	8.8 ± 5.5	23.2 ± 49.7	1.83 ± 0.98	G**T**C	−12.3 ± 4.1	−40.0 ± 32.0	0.29 ± 0.91
T**T**G	7.0 ± 3.8	17.5 ± 22.2	1.38 ± 0.58	G**T**G	−2.3 ± 11.0	−10.0 ± 9.1	0.87 ± 0.99
T**C**A	1.6 ± 7.5	−3.9 ± 14.3	2.74 ± 0.94	G**C**A	−2.1 ± 4.6	−12.5 ± 10.4	1.73 ± 1.12
T**C**T[b,d]	9.9 ± 4.5	30.5 ± 8.7	2.58 ± 1.04	G**C**T	−3.5 ± 4.2	−16.1 ± 10.1	1.42 ± 1.00
T**C**C	−15.2 ± 4.2	−48.7 ± 9.9	0.16 ± 0.96	G**C**C	0.4 ± 6.6	−1.0 ± 17.7	0.60 ± 0.93
T**C**G	−0.7 ± 4.1	−7.0 ± 10.3	1.32 ± 0.61	G**C**G	13.8 ± 5.0	42.9 ± 12.5	2.02 ± 0.94
T**G**A	−12.3 ± 7.9	−47.0 ± 6.4	2.13 ± 0.99	G**G**A	2.7 ± 4.5	3.6 ± 10.3	1.50 ± 0.92
T**G**T[b]	20.9 ± 5.5	59.8 ± 10.2	3.42 ± 0.94	G**G**T	−1.2 ± 7.7	−9.3 ± 21.5	1.75 ± 1.05
T**G**C	2.6 ± 3.2	1.6 ± 14.2	2.30 ± 0.96	G**G**C	−1.7 ± 5.5	−7.2 ± 14.4	0.48 ± 0.94
T**G**G	−9.2 ± 4.5	−31.4 ± 16.6	0.33 ± 0.59	G**G**G	3.5 ± 2.8	8.5 ± 6.1	0.98 ± 0.58

[a] These were derived from thermodynamics for sequences with a single bulge from the literature and plots of $1/T_M$ vs ln $C_T/4$ in Table 2. Note that thermodynamics from fits of individual melting curves in Table 2 were not used (see the Discussion). [b] These parameters were derived from the data in the literature (*19*) and ours (Table 2). [c] This parameter was derived from the data in the literature (*21*) and ours (Table 2). [d] This parameter was derived from the data in the literature (*20*) and ours (Table 2).

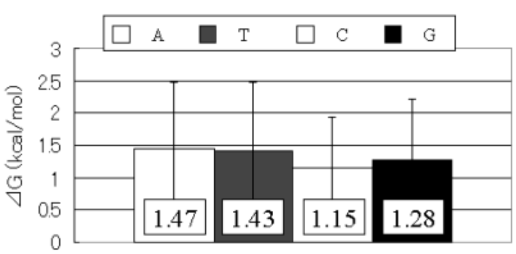

FIGURE 1: Effect of a bulged base on ΔG°_{37}. Each bar represents the average ΔG°_{37} with respect to the type of bulged base. The precise value is denoted by the number in the bar. The error bar represents the standard deviation per type of bulged base.

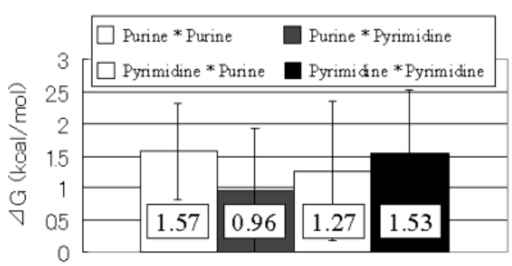

FIGURE 2: Effect of flanking base pairs on ΔG°_{37}. Each bar represents the average ΔG°_{37} with respect to the type of flanking base pairs. The precise value is denoted by the number in the bar. The error bar represents the standard deviation per type of flanking base pair.

evidence for the validity of the plots of $1/T_M$ versus ln $C_T/4$ was the linearity of the regression analysis, which produced expected results in the two-state model (coefficient of determination ≥ 0.98). Thus, we hypothesized that the thermodynamics from plots of $1/T_M$ versus ln $C_T/4$ were estimated more correctly than those from fits of individual melting curves. To verify this hypothesis, we investigated the average deviations between thermodynamics for complementary sequences from our laboratory and those predicted by using the nearest-neighbor parameters of Allawi *et al.* (*9*). If the aforementioned hypothesis is true, the thermodynamics from the plots of $1/T_M$ versus ln $C_T/4$ should be consistent with the predicted thermodynamics, while those from the fits of individual melting curves should be inconsistent with the predicted thermodynamics. In fact, the average deviations between the ΔG°_{37}, ΔH°, and ΔS° values of 31 complementary sequences estimated by plots of $1/T_M$ versus ln $C_T/4$ and the predicted values are 4.4, 6.5, and 7.0%, respectively. These were within the limits of what can be expected for a nearest-neighbor model. In contrast, the average deviations between the ΔG°_{37}, ΔH°, and ΔS° values of 31 complementary sequences estimated by fits of individual melting curves and the predicted values are 7.4, 15.1, and 16.6%, respectively. These were greater than the limits of what can be expected for a nearest-neighbor model. These results supported our hypothesis. Therefore, we concluded that the thermodynamics derived from plots of $1/T_M$ versus ln $C_T/4$ were estimated well because of their robustness.

Comparison with ΔG°_{37} Using the Traditional Model. A previous paper proposed the simplest model, which assigns a free energy penalty of 2.8 kcal/mol for each single-bulge loop (*46*). However, the contribution of a single bulge to duplex stability depended on both the type of the bulged base and the flanking base pairs and varied from 3.69 to −1.05 kcal/mol (see the Results). Therefore, this penalty model is an oversimplification.

Another model for predicting ΔG°_{37} for a single bulge was proposed by Zhu *et al*. (*23*). They modeled ΔG°_{37} based on the stacking energy between flanking base pairs. They defined a bulged base with no identical neighboring base as belonging to group I and those with at least one identical flanking base as belonging to group II. They noted that group I bulges tended to be less stable than group II bulges in the same nearest-neighbor environments because of more conformational freedom (they called this effect "position degeneracy"). They derived the following equation based on their experimental data.

$$\Delta G_{(\mathrm{XNZ})\cdot(\mathrm{X'-Z'})} = 2.72 + 0.48\Delta G^{S}_{(\mathrm{XZ})\cdot(\mathrm{X'Z'})} - \delta g \quad (4)$$

where δg is zero for group I bulges and 0.4 kcal/mol for group II bulges. The first term, 2.72 (kilocalories per mole), is the positive free energy due to insertion of a base bulge into the duplex. The second term is the free energy contribution of the stacking interaction caused by the flanking base pairs. The third term represents the effect of position degeneracy. The ΔG°_{37} from our results also reflected that effect. For example, the A**A**A bulge was the most stable structure among the A**X**A bulges (**X** being A, T, C, or G) (see Table 4). As described previously, however, this model has many limitations (unpublished experiments). It does not distinguish the type of bulged base, or in which strand (i.e., sense or antisense) a bulged base is inserted. For example, the stability of 5′-C**X**T-3′/3′-GA-5′ (X being A, T, C, or G) equals that of 5′-CT-3′/3′-G**X**A-5′ according to eq 4. However, the ΔG°_{37} values derived from our experimental data do not necessarily agree with those derived from eq 4. For example, the ΔG°_{37} of the G**A**G bulge was 0.52 kcal/mol, while that of the G**C**G bulge was 2.02 kcal/mol. This exception denotes that a bulged base affects ΔG°_{37} for a single-bulge loop. Moreover, ΔG°_{37} of the C**A**C bulge was 1.73 kcal/mol, while that of the G**A**G bulge (the flanking GXG base pairs are complementary to CXC) was 0.52 kcal/mol. This exception denotes that ΔG°_{37} for a single bulge depends on the strand in which a bulged base is inserted. Therefore, a model incorporating the type of bulged base and its flanking base pairs was needed to approximate the ΔG°_{37} for a single-bulge loop adequately. Conversely, the sufficiency of the nearest-neighbor model for a single bulge was confirmed by the experiment in which the thermodynamics for 10 sequences with the G**T**G bulge were determined (see above).

REFERENCES

1. Southern, E., Mir, K., and Shchepinov, M. (1999) Molecular interactions on microarrays, *Nat. Genet. 21*, 5−9.
2. Tillib, S. V., and Mirzabekov, A. D. (2001) Advances in the analysis of DNA sequence variations using oligonucleotide microchip technology, *Curr. Opin. Biotechnol. 12*, 53−58.
3. Gerry, N. P., Witowski, N. E., Day, J., Hammer, R. P., Barany, G., and Barany, F. (1999) Universal DNA microarray method for multiplex detection of low abundance point mutations, *J. Mol. Biol. 292*, 251−262.
4. Yan, H., Zhang, X., Shen, Z., and Seeman, N. C. (2002) A robust DNA mechanical device controlled by hybridization topology, *Nature 415*, 62−65.
5. Winfree, E., Liu, F., Wenzler, L. A., and Seeman, N. C. (1998) Design and self-assembly of two-dimensional DNA crystals, *Nature 394*, 539−544.
6. Adleman, L. M. (1994) Molecular computation of solutions to combinatorial problems, *Science 266*, 1021−1024.
7. Benenson, Y., Adar, R., Paz-Elizur, T., Livneh, Z., and Shapiro, E. (2003) DNA molecule provides a computing machine with both data and fuel, *Proc Natl Acad Sci U.S.A. 100*, 2191−2196.
8. Sugimoto, N., Nakano, S., Yoneyama, M., and Honda, K. (1996) Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes, *Nucleic Acids Res. 24*, 4501−4505.
9. Allawi, H. T., and SantaLucia, J., Jr. (1997) Thermodynamics and NMR of internal G•T mismatches in DNA, *Biochemistry 36*, 10581−10594.
10. Allawi, H. T., and SantaLucia, J., Jr. (1998) Thermodynamics of internal C•T mismatches in DNA, *Nucleic Acids Res. 26*, 2694−2701.
11. Peyret, N., Seneviratne, P. A., Allawi, H. T., and SantaLucia, J., Jr. (1999) Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A•A, C•C, G•G, and T•T mismatches, *Biochemistry 38*, 3468−3477.
12. Allawi, H. T., and SantaLucia, J., Jr. (1998) Nearest-neighbor thermodynamics of internal A•C mismatches in DNA: sequence dependence and pH effects, *Biochemistry 37*, 9435−9444.
13. Allawi, H. T., and SantaLucia, J., Jr. (1998) Nearest neighbor thermodynamic parameters for internal G•A mismatches in DNA, *Biochemistry 37*, 2170−2179.
14. Li, Y., Zon, G., and Wilson, W. D. (1991) Thermodynamics of DNA duplexes with adjacent G•A mismatches, *Biochemistry 30*, 7566−7572.
15. Li, Y., and Agrawal, S. (1995) Oligonucleotides containing G•A pairs: effect of flanking sequences on structure and stability, *Biochemistry 34*, 10056−10062.
16. Aboul-ela, F., Koh, D., Tinoco, I., Jr., and Martin, F. H. (1985) Base−base mismatches. Thermodynamics of double helix formation for dCA3XA3G + dCT3YT3G (X, Y = A, C, G, T), *Nucleic Acids Res. 13*, 4811−4824.
17. Bommarito, S., Peyret, N., and SantaLucia, J., Jr. (2000) Thermodynamic parameters for DNA sequences with dangling ends, *Nucleic Acids Res. 28*, 1929−1934.
18. Senior, M., Jones, R. A., and Breslauer, K. J. (1988) Influence of dangling thymidine residues on the stability and structure of two DNA duplexes, *Biochemistry 27*, 3879−3885.
19. LeBlanc, D. A., and Morden, K. M. (1991) Thermodynamic characterization of deoxyribooligonucleotide duplexes containing bulges, *Biochemistry 30*, 4042−4047.
20. Ohmichi, T., Nakamuta, H., Yasuda, K., and Sugimoto, N. (2000) Kinetic property of bulged helix formation: analysis of kinetic behavior using nearest-neighbor parameters, *J. Am. Chem. Soc. 122*, 11286−11294.
21. Morden, K. M., Chu, Y. G., Martin, F. H., and Tinoco, I., Jr. (1983) Unpaired cytosine in the deoxyoligonucleotide duplex $dCA_3CA_3G\cdot dCT_6G$ is outside of the helix, *Biochemistry 22*, 5557−5563.
22. Ke, S. H., and Wartell, R. M. (1995) Influence of neighboring base pairs on the stability of single base bulges and base pairs in a DNA fragment, *Biochemistry 34*, 4593−4600.
23. Zhu, J., and Wartell, R. M. (1999) The effect of base sequence on the stability of RNA and DNA single base bulges, *Biochemistry 38*, 15986−15993.
24. Longfellow, C. E., Kierzek, R., and Turner, D. H. (1990) Thermodynamic and spectroscopic study of bulge loops in oligoribonucleotides, *Biochemistry 29*, 278−285.
25. 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.
26. Gray, D. M., Hung, S. H., and Johnson, K. H. (1995) Absorption and circular dichroism spectroscopy of nucleic acid duplexes and triplexes, *Methods Enzymol. 246*, 19−34.
27. Tanaka, F., Nakatsugawa, M., Yamamoto, M., Shiba, T., and Ohuchi, A. (2002) Towards a general-purpose sequence design in DNA computing, *Proc. 2002 Congr. Evol. Comput.*, 73−78.
28. Tanaka, F., Kameda, A., Yamamoto, M., and Ohuchi, A. (2003) The effect of the bulge loop upon the hybridization process in DNA computing, *The 5th International Conference on Evolvable*

Systems: From Biology to Hardware, Lecture Notes in Computer Science 2606, 446−456.
29. Borer, P. N., Dengler, B., Tinoco, I., Jr., and Uhlenbeck, O. C. (1974) Stability of ribonucleic acid double-stranded helices, *J. Mol. Biol. 86*, 843−853.
30. Bevington, P. R., and Robinson, D. K. (2002) *Data Reduction and Error Analysis for the Physical Sciences*, 3rd ed., McGraw-Hill, New York.
31. Eaton, J. W. (2002) *GNU Octave*, version 2.1.36, http://www.octave.org/.
32. Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1988) *Numerical Recipes in C*, Cambridge University Press, Cambridge, U.K.
33. 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.
34. SantaLucia, J., Jr., Allawi, H. T., and Seneviratne, P. A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability, *Biochemistry 35*, 3555−3562.
35. 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.
36. Papanicolaou, C., Gouy, M., and Ninio, J. (1984) An energy model that predicts the correct folding of both the tRNA and the 5S RNA molecules, *Nucleic Acids Res. 12*, 31−44.
37. Morden, K. M., Gunn, B. M., and Maskos, K. (1990) NMR studies of a deoxyribodecanucleotide containing an extrahelical thymidine surrounded by an oligo(dA)•oligo(dT) tract, *Biochemistry 29*, 8835−8845.
38. Patel, D. J., Kozlowski, S. A., Marky, L. A., Rice, J. A., Broka, C., Itakura, K., and Breslauer, K. J. (1982) Extra adenosine stacks into the self-complementary d(CGCAGAATTCGCG) duplex in solution, *Biochemistry 21*, 445−451.
39. Kalnik, M. W., Norman, D. G., Swann, P. F., and Patel, D. J. (1989) Conformation of adenosine bulge-containing deoxytridecanucleotide duplexes in solution. Extra adenosine stacks into duplex independent of flanking sequence and temperature, *J. Biol. Chem. 264*, 3702−3712.
40. Kalnik, M. W., Norman, D. G., Li, B. F., Swann, P. F., and Patel, D. J. (1990) Conformational transitions in thymidine bulge-containing deoxytridecanucleotide duplexes. Role of flanking sequence and temperature in modulating the equilibrium between looped out and stacked thymidine bulge states, *J. Biol. Chem. 265*, 636−647.
41. Kalnik, M. W., Norman, D. G., Zagorski, M. G., Swann, P. F., and Patel, D. J. (1989) Conformational transitions in cytidine bulge-containing deoxytridecanucleotide duplexes: extra cytidine equilibrates between looped out (low temperature) and stacked (elevated temperature) conformations in solution, *Biochemistry 28*, 294−303.
42. Woodson, S. A., and Crothers, D. M. (1988) Structural model for an oligonucleotide containing a bulged guanosine by NMR and energy minimization, *Biochemistry 27*, 3130−3141.
43. McDowell, J. A., and Turner, D. H. (1996) Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: solution structure of $(rGAGGUCUC)_2$ by two-dimensional NMR and simulated annealing, *Biochemistry 35*, 14077−14089.
44. SantaLucia, J., Jr., and Turner, D. H. (1997) Measuring the thermodynamics of RNA secondary structure formation, *Biopolymers 44*, 309−319.
45. Albergo, D. D., Marky, L. A., Breslauer, K. J., and Turner, D. H. (1981) Thermodynamics of $(dG\text{-}dC)_3$ double-helix formation in water and deuterium oxide, *Biochemistry 20*, 1409−1413.
46. Wetmur, J. G. (1999) Physical chemistry of nucleic acid hybridization, *DNA Based Computers III, DIMACS Series in Discrete Mathematics and Theoretical Computer Science 48*, 1−23.

BI036188R