Differences between DNA Base Pair Stacking Energies Are Conserved over a Wide Range of Ionic Conditions†

Theodore Johnson, Jian Zhu, and Roger M. Wartell*

School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332-0230

Received May 11, 1998; Revised Manuscript Received July 8, 1998

ABSTRACT: Base pair stacking free energy parameters in a low ionic strength solvent were determined from an analysis of DNA fragments using temperature gradient gel electrophoresis (TGGE). Transition midpoint temperatures (T_u) were determined for the first melting domain (52 ± 4 bp) of 16, 339 bp DNAs that differed from each other by single base pair substitutions. The data were combined with previously obtained T_u data from 17 similar DNAs that had single base pair changes at different sites [Ke, S. H., and Wartell, R. M. (1995) *Biochemistry 34*, 4593−4599]. The T_u values were used to evaluate free energy differences ($\delta\Delta G$) between 31 pairs of DNAs. Linear equations relating the $\delta\Delta G$ values to changes in base pair stacking were analyzed by singular value decomposition (SVD) to determine the 10 nearest neighbor free energy parameters. The order of stability of the parameters, TA < AT < AA < AG < GT ≈ TC ≈ TG < CC < GC ≈ CG, was essentially the same as the hierarchy determined in 1 M Na^+ [Allawi, H. T., and SantaLucia, J., Jr. (1997) *Biochemistry 36*, 10581−10594]. The experimental free energy differences were in good agreement with predictions made using nearest-neighbor parameters determined from several previous studies conducted in medium or high salt concentrations. Conversely the parameters determined in the current study produced good predictions of free energy differences previously determined from 59 DNA oligomers in 1 M Na^+. The results indicate that differences between base pair stacking energies are conserved across a wide range of ionic conditions, and in both oligomer and polymer DNA contexts.

An accurate thermodynamic description of DNA duplex formation is of importance for understanding genetic processes that involve DNA unwinding (*1, 2*), and for molecular biological methods that utilize predictions of DNA stability (*3, 4*). The same basic model, with some differences in format and assumptions, has been employed by investigators in the analysis of DNA duplex−single strand transitions (*5*−*12*). In this model, a base pair's stability is assumed to be governed by its identity and by the identity and condition (paired or unpaired) of its nearest-neighbor base pairs. Ten nearest-neighbor stacking parameters are commonly used to quantify the free energy of these interactions. Additional parameters are required to describe the free energy of nucleating a duplex from single strands, and the free energy of internal loops of unpaired bases. The importance of the latter terms depends on DNA length and sequence, and its structure (e.g., hairpin or linear).

A number of studies have been made to determine the nearest-neighbor stacking parameters. UV absorbance temperature transitions of long DNA domains, and repeating DNA polymers were examined to evaluate the stacking parameters (*5*−*8*). Only 8 linearly independent combinations of the 10 stacking parameters can be obtained from these studies (*13*). Doktycz et al. (*10*) evaluated the nearest-neighbor parameters from transition curves of DNA dumbbell oligomers, thus avoiding the influence of the strands to duplex nucleation step. An analysis of previously published data indicated a general consensus for parameter sets evaluated from long DNAs. However, there was a lack of consensus between the parameters determined from the DNA oligomers and the parameters determined from the long DNAs.

Allawi and SantaLucia (*14*) recently examined thermodynamic data from 108 linear DNA oligomers that exhibited two-state transitions, and determined the10 nearest-neighbor stacking parameters in 1 M Na^+. Their analysis utilized two duplex initiation parameters, and assumed three distinct nucleation events depending on whether the oligomer had two G·C pairs, one G·C and one A·T, or two A·T pairs at the termini.

In the present investigation, temperature gradient gel electrophoresis (TGGE)[1] was used to obtain mobility transition curves of the end melting domain of DNA fragments differing by single base pair substitutions. Duplex DNAs differing in stability in their first melting domain unwind and decrease in mobility at different depths in a polyacrylamide gel with a superimposed temperature gradient. The differences in the transition midpoint temperatures were quantitatively measured and related to free energy differences. Since the transitions only involve an end melting domain, the free energy differences could be directly related to the

† This work was supported by a Georgia Tech-Emory Biomedical Center grant. T.J. was supported by a Howard Hughes Medical Institute undergraduate internship.

* To whom correspondence should be addressed.

[1] Abbreviations: nt, nucleotide(s); bp, base pair(s); TGGE, temperature gradient gel electrophoresis.

S0006-2960(98)01093-9 CCC: $15.00
Published on Web 08/13/1998

changes in base pair stacking within the domain. Duplex initiation parameters do not enter into the analysis.

Data from 31 pairs of DNAs were used to evaluate the nearest-neighbor free energy parameters in the low ionic strength solvent of the TGGE experiments (0.002 M Na^+). The parameters predicted the experimental free energy differences determined by TGGE with good accuracy ($R^2 = 0.91$). They also predicted, with similar accuracy, the free energy differences evaluated from DNA oligomer data obtained in 1 M Na^+ (*12*). In addition, predictions made using stacking parameter sets determined in 0.1 M Na^+ (*10*) and 1 M Na^+ (*14*) solutions produced good agreement with the TGGE free energy differences. The results are consistent with a recent analysis by SantaLucia that showed a previously unrecognized correlation between the stacking parameters determined from polymer studies with stacking parameters evaluated from oligomer studies when adjusted for salt dependence (*15*).

MATERIALS AND METHODS

Materials. The plasmid DNA used as a template for polymerase chain reactions (PCR) was pUC8-31 (*16*). This DNA contains a 130 bp fragment of the *ctc* promoter sequence from *Bacillus subtilis* inserted between the HindIII and *Eco*RI restriction sites of pUC8. Previous TGGE studies investigated base pair changes within the 373 bp DNA sequence between the *Eco*RI and *Rsa*I sites of pUC8-31 (*17, 18*). In the current work, 339 bp DNA fragments were generated by PCR from the 373 bp DNA region. The wild-type 339 bp sequence extends from the *Eco*RI site to 34 bp short of the *Rsa*I site, and the top strand is shown in Figure 1. Sixteen DNA molecules were produced containing the wild-type sequence or specific base pair substitutions at one of five positions indicated by solid circles. Upstream (UP) and downstream (DP) primers used to generate the DNAs are shown in Figure 1. They were obtained from Operon Inc. Base pair positions were numbered with respect to the *ctc* transcription initiation site (+1). The open circles above the template sequence denote positions −36, −38, −39, and −43 where the effects of base pair substitutions were previously determined using 373 bp fragments (*17*). Since the least stable melting domains for the 339 and 373 bp DNAs are identical, data from all nine sites were included in the analysis.

PCR. Polymerase chain reaction amplifications employed 100 μL mixtures containing 40−70 pg of the pUC8-31 DNA template, 0.6 μM of each primer, and 0.2 mM of each dNTP in a buffer of 10 mM Tris (pH 8.3), 50 mM KCl, and 2 mM $MgCl_2$. An additional 0.2 mM dATP was added to the reaction in order to obtain a single DNA band (see below); 100 μL of mineral oil was overlaid onto the mixture which was then placed in a thermal cycler and held at 80 °C for 10 min prior to adding 2.5 units of Taq DNA polymerase (Promega or BRL). The reaction mixture was subjected to 30−35 cycles of denaturation at 94 °C for 1 min, primer annealing at 55 °C for 1 min, and primer extension at 72 °C for 2 min. An additional 5 min at 72 °C was allowed for the completion of primer extension after the last cycle. The length and purity of the DNA product of each reaction were verified by gel electrophoresis. Each of the 339 bp PCR products exhibited identical mobility in a 6.5% polyacrylamide gel.

The doubling of the dATP concentration in the PCR eliminated closely spaced doublet bands which were sometimes observed in vertical TGGE for the 339 bp DNAs. Previous work has shown that Taq DNA polymerase can add a single nontemplated nucleotide, generally adenosine, to the blunt end of replicated DNA (*19*). We hypothesized that the doublet bands resulted from the slight difference in stability between the blunt end and base extended DNA. The extra 0.2 mM dATP apparently saturated the reaction causing A-additions since only a single band was observed for these PCR products.

TGGE. Most of the temperature gradient gel electrophoresis (TGGE) experiments were carried out using a gel apparatus described previously (*20*). The glass plates were 25 cm × 18.7 cm and were sandwiched between two aluminum blocks. In parallel TGGE, the fluid circulates through channels at the top and bottom of the blocks, establishing a temperature gradient parallel to the direction of mobility. In perpendicular TGGE, the fluid channels are along the left and right sides of the blocks so that the temperature gradient is perpendicular to the direction of mobility.

The gels were 0.75 mm thick and were composed of 6.5% polyacrylamide (with 37.5:1 acrylamide:bisacrylamide) containing 4.2 M urea and 24% (v/v) formamide in 0.5× TBE (0.045 M Tris−borate plus 1 mM Na-EDTA, pH 8.2). A mixed resin, AG501-X8D (BioRad), was used to deionize the formamide prior to its incorporation into the gel; 0.5× TBE was used as the gel running buffer. Parallel TGGE gels were run at 80−100 V for 13−16 h, and perpendicular TGGE gels were run at 120 V for 11−12 h. Temperature gradients were 28−33 °C for parallel TGGE gels, and about 26−32.5 °C for perpendicular TGGE gels. These temperatures were chosen to focus on the behavior of the DNAs' first melting domain. Ethidium bromide was used to stain the gels which were subsequently photographed.

The parallel TGGE experiments were used to determine the relative stability of a set of related DNA fragments. This information was used to assign the mobility transition curves of the DNAs when they were run simultaneously in a perpendicular TGGE experiment. A Bio-Rad Inc. mini-gel apparatus was employed for approximately one-third of the parallel TGGE experiments. It obtained results equivalent to the conventional TGGE apparatus. Instead of using an established temperature gradient, the temperature in the gel was increased with time by operating the power supply at constant current (9−11 ma). To generate a temperature gradient in the range from 29 to 33 °C, the buffer in the gel tank was preheated to 29 °C. After loading samples, the temperature in the tank was monitored periodically and the current adjusted as needed to produce a slow quasi-linear increase in temperature. Although band separations were smaller, this method took only 1 h and was adequate for ordering the relative stability of a number of DNA species.

The perpendicular TGGE format was used to obtain mobility transition curves of the first melting domain of the DNAs. Three or four DNA samples were mixed and loaded into the single 'well' that spanned the width of the gel. Each DNA exhibited a sigmoidal decrease in mobility from left to right (Figure 2C). Approximately 1 μg of each DNA was required to visualize by ethidium bromide staining. Temperatures at several locations across the gel were measured

by inserting a needlelike thermocouple probe (TMTSS-020-6; Omega Inc). This was done at the end of a run with the voltage off. Dark lines in the photographs indicated the positions of the temperature measurements. The mobility transition temperatures, T_u, were determined as the midpoint of the mobility curve after making linear base line corrections. This method differed from the earlier approach (*17*) which used the peak of the derivative of the transition curve. It was simpler to apply and yielded the same values for temperature shifts between pairs of DNAs differing by a base pair substitution.

Evaluating Nearest-Neighbor Stacking Free Energies. The most commonly used version of the nearest neighbor model assigns a stacking free energy, ΔG_i°, to the process of adding a stacked H-bonded base pair onto the end of a base paired segment. The subscript '*i*' designates 1 of the 10 specific base pair stacks: AA/TT, AT/AT, TA/TA, CA/TG, GT/AC, CT/AG, GA/TC, CG/CG, GC/GC, GG/CC. For the 339 or 373 bp DNA duplexes used in this study, the first melting domain melts from the *Eco*RI end in a two-state manner, leaving a partial duplex (*17*, *18*). The total free energy change for melting the domain at temperature T is

$$\Delta G_T^\circ = \Sigma_{(i=1-10)} n_i \Delta G_i^\circ = \Delta H^\circ - T\Delta S^\circ \qquad (1)$$

with ΔH° and ΔS° corresponding to the domain's enthalpy and entropy change. The symbol n_i is the number of nearest-neighbor stacking interactions of the i^{th} type present in the domain. For a two-state melting transition, the domain's enthalpy and entropy are related through its melting temperature:

$$\Delta H^\circ = T_m \Delta S^\circ \qquad (2)$$

If a single base substitution occurs within the domain that does not alter its size, the free energy difference in melting the original (ΔG_T°) vs the altered ($\Delta G'_T{}^\circ$) domain is linearly correlated to their difference in melting temperatures, δT_m. This follows from eqs 1 and 2 assuming ΔS° is essentially unchanged.

$$\delta\Delta G \equiv \Delta G'_T{}^\circ - \Delta G_T^\circ = T_m' \Delta S^\circ - T_m \Delta S^\circ = \delta T_m \Delta S^\circ \qquad (3)$$

Assuming that $\delta T_m \approx \delta T_u$, eqs 3 and 1 can relate the stacking energy difference created by a base pair substitution to the experimental δT_u. ΔS° can be estimated by multiplying the number of base pair doublets opened in the first melting domain, N, times ΔS_{AV}, the average entropy change per base pair doublet. Using these substitutions:

$$\delta\Delta G = \Sigma(n'_i - n_i)\Delta G_i^\circ = \delta T_u (N\Delta S_{AV}) \qquad (4)$$

The value for ΔS_{AV} has been determined from DNA polymer melting analysis as -24.85 ± 1.7 cal·K^{-1}·(mol of bp)$^{-1}$ (*9*). A detailed study of DNA oligomers has yielded an average entropy change of -22.4 ± 2.2 cal·K^{-1}·(mol of bp)$^{-1}$ (*14*). Since the DNAs examined are beyond oligomer length, the former value was assumed in eq 4. The size of the melting domain, N, was previously estimated to be 52.5 $\pm$ 3.5 bp for the DNA molecules studied (*17*, *18*).

For M+1 molecules differing by single base pairs, M equations analogous to eq 4 can be written. This system of equations can be expressed in the matrix notation:

$$\mathrm{A}\cdot\mathrm{x} = \mathbf{b} \qquad (5)$$

The matrix **A** is a $M \times 10$ matrix in which each row contains the 10 terms $(n'_i - n_i)$ representing the differences in the numbers of *i*th-type base pair doublet between the altered and wild-type sequence. The vector **x** is a 1×10 column vector with elements corresponding to the free energy parameters, ΔG_i°. The vector **b** represents the $M \times 1$ column vector containing the experimental values of $N\Delta S_{AV}\delta T_u$. The database consisted of $M = 31$ equations. Fifteen were generated from the 16 DNA species used in this investigation (the wild-type species and all base pair substitutions at positions -27, -28, -33, -34, and -35). Sixteen equations were based on previously published data on single base pair substitutions and deletions at sites -36, -38, -39, and -43 (*17*).

The singular value decomposition (SVD) algorithm (*21*) was used to determine the 10 stacking free energy parameters from eq 5. As has been previously described (*13*, *23*), constraints on base pair doublet frequencies for a system of long DNA sequences prohibit obtaining a unique solution to the 10 parameters. This is the case for the data used to evaluate the 10 parameters in this study. As indicated below, eight linearly independent combinations of the parameters are unique.

The experimental errors of the terms in matrix **b** were collected in an $M \times 1$ column vector, **e**, and served as weighting factors in the least-squares analysis. Goodness of fit values of the stacking parameters **x** were assessed from the squares of the weighted residuals, $\chi^2 = \Sigma[(b_i - A_{ij}x_j)/e_i]^2$, and the probability Q that a value of χ^2 as poor as the experimental value should occur by random chance alone (*22*). A typical value of χ^2 for a 'moderately good fit' (*21*) is $\chi^2 \approx \nu$, number of degrees of freedom. For the data examined, $\nu = 31 - 10 = 21$. Based on the criteria underlying the application of Q (*21*), a value exceeding 0.001 may still indicate that the apparent error of χ^2 is due to random chance alone. For $Q < 0.001$, the apparent error is more likely to result from inaccurate assumptions of the model.

RESULTS

Mobility transition temperatures were determined for the first melting domain of 16 DNAs differing by single base pair substitutions. The base pair substitutions were at the five sites indicated in Figure 1 by filled circles. Each set of DNA molecules, corresponding to all base pair changes at a given site, were subjected to parallel TGGE and perpendicular TGGE. The parallel TGGE experiments determined the relative stabilities of the DNAs. These results were used to identify the DNA species associated with mobility transition curves generated by perpendicular TGGE.

Figure 2A shows a parallel TGGE mini-gel experiment of five DNAs differing by a base pair substitution or deletion at site -35. The temperature of the gel increased from 29.5 to 34.0 °C as a function of run time. The lanes contained, from left to right, 339 bp DNAs with the base pairs C·G, G·C, and T·A, A·T and the deletion at site -35. The observed order of stability was C·G > G·C > -/- > T·A > A·T. Figure 2B shows results from a conventional parallel

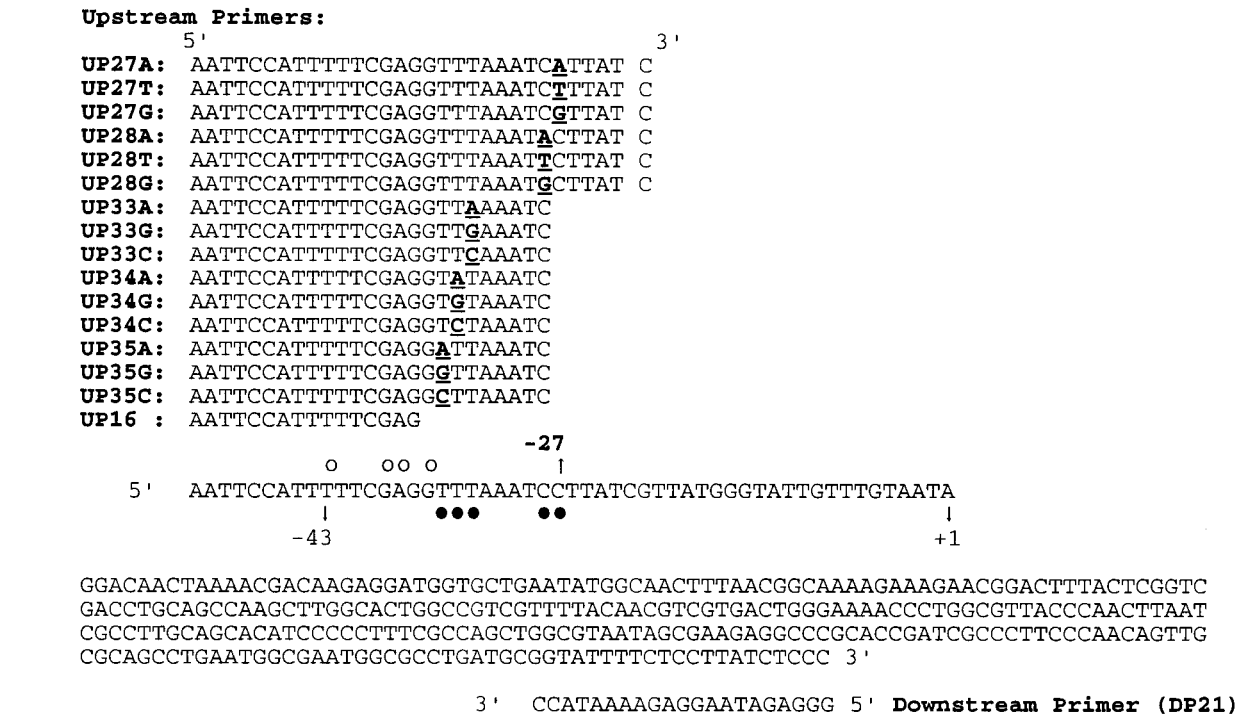

FIGURE 1: The 339 bp sequence of the PCR-amplified restriction fragment between the *Eco*RI and *Rsa*I sites on the pUC8-31 plasmid. Positions +1 and −43 are shown as are the various upstream PCR primers and the downstream PCR primer. Substitution sites within the primers are underlined. Black circles denote the sites under investigation in this study (−27, −28, −33, −34, −35). White circles denote the sites previously investigated by Ke and Wartell, 1993 (−36, −38, −39, −43).

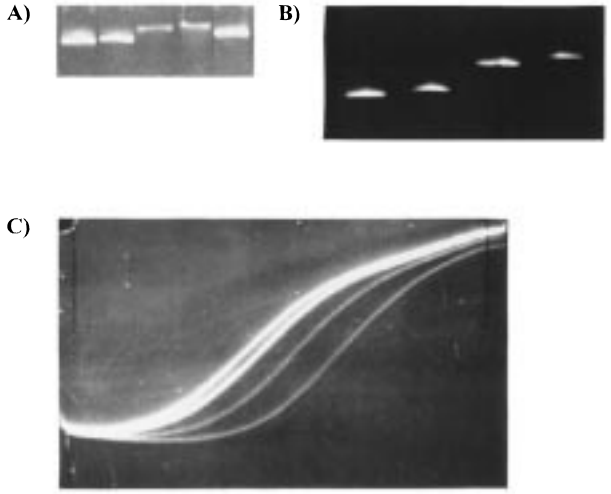

FIGURE 2: Temperature gradient gel experiments of 339 bp DNAs. (A) Separation of five DNAs using a minigel with a temperature increase from 29.5 to 34 °C over 1 h. Samples in lanes from left to right are DNAs with base pairs C·G, G·C, T·A, and A·T and deletion at position −35. (B) Conventional parallel TGGE results with temperature gradient from 28 to 33 °C. Four bands from left to right are DNAs with base pairs C·G, G·C, T·A, and A·T at position −35. (C) Perpendicular TGGE experiment of DNAs with gradient from 27.5 to 32.5 °C. Mobility curves from left to right correspond to DNAs with base pairs A·T, T·A, C·G, and G·C at position −27.

TGGE experiment using the 339 bp DNAs with the four possible base pairs at site −35. The lanes contained, from left to right, the DNA species with C·G, G·C, T·A, and A·T, respectively, at site −35. The relative stabilities of the DNAs were the same as in the mini-gel experiment (Figure 2A). The mini-gel produced smaller band separation, but experiments could be completed in 1 h. The longer gels were used to resolve ambiguities.

Table 1: δT_u Measurements from Perpendicular TGGE Experiments

site	nearest-neighbor environment[a]	δT_u (°C)[b] with X·Y representing the base pair A·T	T·A	G·C	C·G	-/-[c]
−27	CXT/AYG	−0.6	−0.4	0.5	0.0[d]	−
−28	TXC/GYA	−1.0	−0.6	0.3	0.0[d]	−
−33	TXA/TYA	0.0	0.0[d]	1.3	1.1	−
−34	TXT/AYA	−0.3	0.0[d]	0.8	0.6	−
−35	GXT/AYC	−0.2	0.0[d]	0.7	0.8	−
−36	GXT/AYC	−0.9	−0.8	0.0[e]	0.0	−0.5
−38	GXG/CYC	0.0[e]	0.3	0.8	1.6	0.5
−39	CXA/TYG	−1.2	−1.7	0.0[e]	−0.2	−0.8
−43	TXT/AYA	−0.6	0.0[e]	0.6	0.1	0.2

[a] Described as 5′→3′/5′→3′ with X·Y the base pair at the stated site. [b] δT_u are the temperature differences, [T_u(given DNA) − T_u(wt DNA)]. Estimated precision is ±0.1 °C for $\delta T_u \leq 1.0$ °C and ± 0.2 °C for $\delta T_u > 1.0$ °C. [c] -/- represents a deletion at the indicated site. [d] DNA sequence is wild type. $\langle T_u \rangle$ of first domain was 29.9 ± 0.5 °C based on 7 experiments. [e] DNA sequence is wild type. $\langle T_u' \rangle$ of first domain was 31.8 ± 0.6 °C based on 19 experiments (*17*).

Figure 2C shows a typical perpendicular TGGE experiment of four 339 bp DNAs with base pair differences at site −27. The temperature gradient was from 27.5 to 32.5 °C, increasing from left to right. From the order of stability determined by parallel TGGE, the mobility curves from left to right were assigned to the DNA molecules with base pairs A·T, T·A, C·G, and G·C, respectively.

Table 1 lists the δT_u values determined from the DNAs examined in this study relative to the wild-type sequence DNA, as well as the δT_u data from the work by Ke and Wartell (*17*). The latter data were evaluated from 373 bp DNAs at the sites −36,−38, −39, and −43. Base pair substitutions at site −36 were reexamined with the 339 bp DNAs. The δT_u values were identical or within 0.1 °C of the previously obtained values. We note that the average T_u value of the wild-type first melting domain in the current

Table 2: Stacking Free Energy Parameters and Test of the Robustness of Data Set

stacking base pair doublet[a]	stacking parameters (kcal/mol) with the following modifications to **A** matrix[b]									
	using all sites	w/o site −27	w/o site −28	w/o site −33	w/o site −34	w/o site −35	w/o site −36	w/o site −38	w/o site −39	w/o site −43
AA/TT	0.26 ± 0.08	0.29	0.27	0.25	0.21	0.26	0.27	0.27	0.27	0.22
AT/AT	0.51 ± 0.09	0.51	0.51	0.49	0.53	0.50	0.50	0.52	0.53	0.43
AG/CT	0.15 ± 0.07	0.17	0.17	0.11	0.17	0.15	0.16	0.15	0.16	0.01
TA/TA	0.76 ± 0.09	0.77	0.80	0.67	0.80	0.76	0.77	0.77	0.77	0.65
TG/CA	−0.26 ± 0.07	−0.24	−0.26	−0.23	−0.29	−0.26	−0.26	−0.25	−0.23	−0.35
TC/GA	−0.16 ± 0.07	−0.14	−0.16	−0.13	−0.15	−0.16	−0.16	−0.15	−0.16	−0.27
GT/AC	−0.05 ± 0.07	−0.03	−0.03	−0.10	−0.07	−0.05	−0.04	−0.04	−0.06	−0.18
GC/GC	−1.02 ± 0.07	−1.03	−1.03	−1.03	−1.05	−1.03	−1.00	−1.01	−0.99	−1.07
CG/CG	−1.07 ± 0.07	−1.08	−1.08	−1.08	−1.10	−1.07	−1.05	−1.06	−1.06	−1.12
CC/GG	−0.68 ± 0.07	−0.69	−0.70	−0.69	−0.68	−0.68	−0.66	−0.69	−0.64	−0.74
χ^2	34.19	29.48	32.31	30.57	25.78	33.93	33.21	29.51	29.87	17.80
Q	0.03	0.04	0.02	0.03	0.10	0.01	0.01	0.03	0.03	0.40

[a] Arranged as 5′→3′/5′→3′. [b] w/o site: rows of **A** representing operations at site were removed, and SVD regression performed.

study was 1.9 °C lower than the average value obtained previously under ostensibly identical conditions (Table 1). Although we are uncertain of the cause of this difference, verification of the previously obtained δT_u values at site −36 supports combining all δT_u data in the analysis. Recent mobility curve T_u values of the wild-type 373 bp DNA first melting domain are consistent with the 339 bp DNA. Replacement of the temperature probe and/or differences in heating block insulation material during the past 5 years may have been responsible for the difference.

Determination of Base Pair Stacking Free Energy Parameters. The 10 nearest-neighbor stacking free energy parameters for helix formation were determined from the data in Table 1 as described under Methods. The first column in Table 2 lists the stacking parameters along with their associated standard deviations. Data from all nine sites were employed. As expected, the rank of matrix **A** was 8, indicating that only 8 linear combinations of the 10 parameters can be uniquely determined from the data set. The average T_u for all measurements was close to 31 °C. An adjustment of the free energies to 37 °C can be made by subtracting 0.15 kcal/mol from each value. This adjustment utilizes the relation of free energy, enthalpy, and entropy, and assumes a constant $\Delta S = 24.8$ cal K^{-1} mol^{-1} for each stacking interaction.

The relative order of stability of the free energy stacking parameters is TA/TA < AT/AT < AA/TT < AG/CT < GT/AC ≈ GA/TC ≈ CA/TG < GG/CC < CG/CG ≈ GC/GC. This order is essentially identical to the order of stability obtained from an analysis of melting curves of oligomer duplexes in 1.0 M Na^+ (*14*): TA/TA < AT/AT < AA/TT < AG/CT ≈ GA/TC < GT/AC ≈ CA/TG < GG/CC < CG/CG ≈ GC/GC. A more meaningful comparison of the hierarchies is to compare eight linearly independent combinations of the stacking parameters. If we employ the eight combinations of stacking parameters proposed by Goldstein and Benight (*13*), the hierarchies from the two studies are again almost the same. They differ only in the ranking of stability for the combination 1/3 ($\Delta G_{AG} + \Delta G_{GC} + \Delta G_{CA}$). These results indicate that the order of DNA stacking energies is preserved over a large change in ionic conditions.

Analysis of Parameter Fit. The standard deviations and goodness of fit terms in Table 2 were based on experimental errors in the **b** terms. Their average was 12%. The χ^2 value of 34 and Q value of 0.035 indicate that the nearest-neighbor model produces a 'moderately good fit' to the data (*21*). Varying the experimental error from 10% to 15% has a small effect on χ^2 and Q. A test of the robustness (*21*) of the solution set was made by running nine sequential SVD regressions in which data from each of the nine sites under study were eliminated from consideration (Table 2). The rank values of the reduced **A** matrixes with three or four less rows, and the changes, or lack thereof, in the solution set provide a measure of the completeness of the data set. This type of analysis was carried out by Allawi and SantaLucia (*14*) to evaluate the uncertainties of thermodynamic parameters of 10 sequences with a G·T mismatch.

Columns 2 through 10 of Table 2 show the results of removing each site on the analysis. In all cases, the rank of the modified **A** matrix remained at 8. Thus, the number of independent combinations of stacking parameters that can be determined remains the same. With the exception of the regression that excluded site −43, the stacking parameters showed only small changes. When site −43 was excluded, approximately half the parameters were altered to values outside stated errors. The Q values of the regressions that excluded site −34 or site −43 were 0.1 and 0.4, respectively, notably higher than the other cases. The increase in Q suggests that $(A)_n$·$(T)_n$ tracts, which occur at sites −34 and −43, decrease the quality of fit when they are included in the analysis. This suggests that the stacking interaction of an (AA)·(TT) doublet within an $(A)_n$·$(T)_n$ tract differs from the same doublet in a heterogeneous sequence. This is consistent with previous work that shows $(A)_n$·$(T)_n$ tracts create an unusual duplex structure (*24*).

Comparison of Experimental Free Energy Differences with Predictions. Table 3 summarizes the operations employed to generate the rows of the complete **A** matrix. It also compares the experimental free energy differences, $\delta(\Delta G)$, with values predicted by the stacking parameters from the present study, Alawari and SantaLucia (*14*), Doktycz et al. (*10*), and Delcourt and Blake (*9*). The three base sequences in the column marked "fragments" represent the DNA duplexes whose free energies were subtracted to give the experimental $\delta(\Delta G)$'s used in eq 3. Each sequence is the three top-strand bases of the DNA with the center base at the stated site. A "∼" indicates a deleted base pair at the site.

Table 3: Comparison of Experimental and Calculated Free Energy Differences

			δΔG(kcal/mol)			
site	fragments[a]	exptl δΔG (kcal/mol)[b]	current[c] analysis	Allawi and Santa Lucia (1997)	Doktycz et al. (1992)	Delcourt and Blake (1991)
	CAT-CGT	1.42	1.37 (−0.05)	1.28 (−0.14)	1.08 (−0.34)	1.34 (−0.08)
−27	CTT-CGT	1.19	1.54 (0.35)	1.33 (0.14)	1.21 (0.02)	1.32 (0.13)
	CCT-CGT	0.65	0.59 (−0.06)	0.49 (−0.16)	0.47 (−0.17)	0.44 (−0.21)
	TAC-TGC	1.72	1.99 (0.26)	1.67 (−0.05)	1.48 (−0.24)	1.05 (−0.67)
−28	TTC-TGC	1.25	1.38 (0.13)	1.39 (0.14)	1.21 (−0.04)	1.24 (−0.01)
	TCC-TGC	0.44	0.44 (0.00)	0.55 (0.11)	0.47 (0.03)	0.36 (−0.08)
	TAA-TGA	1.71	1.44 (−0.27)	1.17 (−0.54)	1.19 (−0.52)	0.94 (−0.77)
−33	TTA-TGA	1.71	1.44 (−0.27)	1.17 (−0.54)	1.19 (−0.52)	0.94 (−0.77)
	TTA-TCA	1.44	1.44 (0.00)	1.17 (−0.27)	1.19 (−0.25)	0.94 (−0.50)
	TAT-TGT	1.38	1.58 (0.20)	1.43 (0.05)	1.25 (−0.13)	1.14 (−0.24)
−34	TTT-TGT	1.01	0.83 (−0.17)	0.89 (−0.12)	0.92 (−0.09)	1.13 (0.12)
	TCT-TGT	0.25	0.30 (0.05)	0.31 (0.06)	0.19 (−0.06)	0.18 (−0.07)
	GAT-GCT	1.27	1.22 (−0.05)	1.34 (0.07)	1.08 (−0.19)	1.27 (0.00)
−35	GTT-GCT	1.04	1.08 (0.04)	1.08 (0.04)	1.02 (−0.02)	1.07 (0.03)
	GGT-GCT	0.13	0.14 (0.01)	0.24 (0.11)	0.28 (0.15)	0.18 (0.05)
	GAT-GGT	1.17	1.08 (−0.09)	1.10 (−0.07)	0.80 (−0.37)	1.09 (−0.08)
−36	GTT-GGT	1.04	0.94 (−0.10)	0.84 (−0.20)	0.74 (−0.30)	0.88 (−0.16)
	GTT-GCT	1.04	1.08 (0.04)	1.08 (0.04)	1.02 (−0.02)	1.07 (0.03)
	G∼T-GGT	0.70	0.68 (−0.02)	−	−	−
	GAG-GCG	2.09	2.08 (−0.01)	1.83 (−0.26)	1.5 (−0.59)	1.43 (−0.66)
−38	GTG-GCG	1.70	1.78 (0.08)	1.52 (−0.18)	1.31 (−0.39)	1.26 (−0.44)
	GGG-GCG	1.04	0.73 (−0.31)	0.73 (−0.31)	0.75 (−0.29)	0.62 (−0.46)
	G∼G-GCG	1.47	1.41 (−0.06)	−	−	−
	CAA-CGA	1.57	1.23 (−0.34)	1.02 (−0.55)	1.02 (−0.55)	1.14 (−0.43)
−39	CTA-CGA	2.22	2.14 (−0.08)	1.61 (−0.61)	1.48 (−0.74)	1.13 (−1.09)
	CCA-CGA	0.26	0.29 (0.03)	0.18 (−0.08)	0.28 (0.02)	0.26 (0.00)
	C∼A-CGA	1.11	0.97 (−0.14)	−	−	−
	TAT-TGT	1.57	1.58 (0.01)	1.43 (−0.14)	1.25 (−0.32)	1.14 (−0.43)
−43	TTT-TGT	0.78	0.83 (0.05)	0.89 (0.11)	0.92 (0.14)	1.13 (0.35)
	TCT-TGT	0.65	0.30 (−0.35)	0.31 (−0.34)	0.19 (−0.46)	0.18 (−0.47)
	T∼T-TGT	0.52	0.57 (0.05)	−	−	−
standard deviation			0.16	0.22	0.24	0.34

[a] Fragments are the two DNA species for which a free energy difference was determined. Each 3-base sequence represents a duplex DNA. It is the sequence of the forward strand (Figure 1) with the middle base at the stated site. [b] Experimental $\delta\Delta G$'s were calculated using eq 4 with δT_u's from Table 1. $\Delta S_{AV} = -24.8$ cal·K^{-1}·mol^{-1}. [c] Calculated $\delta\Delta G$'s using stacking parameters from Table 2. Difference between calculated and experimental values in parentheses.

The predicted $\delta(\Delta G)$ values were generated using the left side of eq 4. Comparisons of experimental vs predicted $\delta(\Delta G)$ values were omitted for the three other studies when operations involved a DNA with a base pair deletion. A deletion-containing operation can produce a row in matrix **A** with only one nonzero term, and results in a direct comparison of a stacking parameter with the current data. A direct comparison is not meaningful due to the different experimental conditions under which the other parameter sets were obtained. The standard deviations between the predicted and experimental $\delta(\Delta G)$ values are given at the bottom of Table 3. They are similar for the four different stacking parameter sets, ranging from 0.16 to 0.34 kcal/mol.

Figure 3 displays the data of Table 3 in graphical form. Linear regression analysis produced the lines shown in the figure. A correlation coefficient (R^2) of 0.91 was obtained for the fit between the predicted and experimental $\delta(\Delta G)$ values using the stacking parameters of the present study (Figure 3A). Stacking parameters from the works of Doktycz et al. (*10*) (Figure 3B) and Allawi and SantaLucia (*14*) (Figure 3C) produced fits with R^2 values of 0.85 and 0.83, respectively. The good agreement predicted by the latter two parameter sets occurs despite the large difference in conditions used for evaluating these parameters vs the conditions of our experiments. Allawi and SantaLucia and Doktycz et al. determined their stacking parameters from DNA oligomer melting curves in 1.0 and 0.1 M Na^+, respectively. Regression analysis comparing the experimental $\delta(\Delta G)$ values with predictions using several other nearest-neighbor parameter sets in different salt conditions also gave similar results. R^2 values of 0.80, 0.80, and 0.81 were obtained with the parameters obtained by Sugimoto et al. (*11*), Vologodskii et al. (*6*), and Gotoh and Tagashira (*5*), respectively (not shown). The predicted $\delta(\Delta G)$ values determined with the Delcourt and Blake (*9*) parameters (Figure 3D) produced an R^2 value of 0.61 in the linear regression.

A different way of assessing the stacking parameter set determined from the TGGE data is to evaluate its ability to predict independent experimental data. SantaLucia et al. (*12*) determined ΔG°_{37} and other thermodynamic data for 59 DNA oligonucleotide duplexes in 1.0 M NaCl. We have compared free energy differences, $\delta(\Delta G)$, between pairs of the 59 DNA oligomers [Table 2 in ref (*12*)] with values predicted by our parameters. DNA oligomer duplexes were grouped by their base pair length from 4 to 16. One DNA of each size class

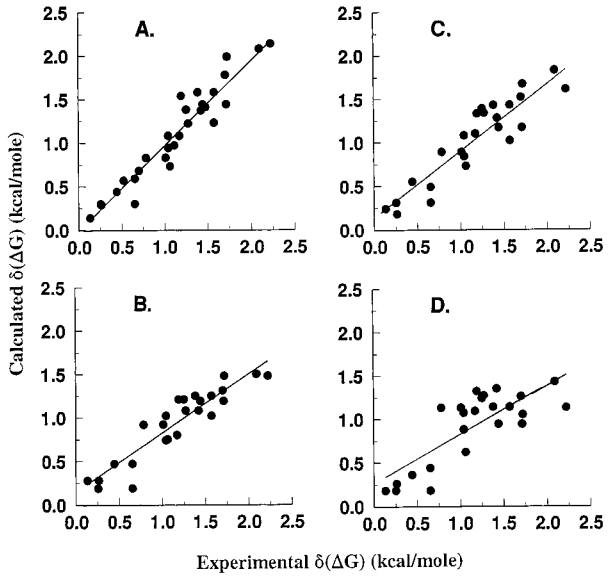

FIGURE 3: Linear regression plots of experimental TGGE data vs calculated free energy differences from Table 3. The calculated $\delta(\Delta G)$ values were from (A) this work, (B) Doktycz et al. (*10*), (C) Allawi and SantaLucia (*14*), and (D) Delcourt and Blake (*9*). Values of R^2 , the slope, and intercept for the four plots were as follows: (A) 0.91, 0.98, −0.013; (B) 0.85, 0.68, 0.15; (C) 0.83, 0.77, 0.14; and (D) 0.61, 0.57, 0.26.

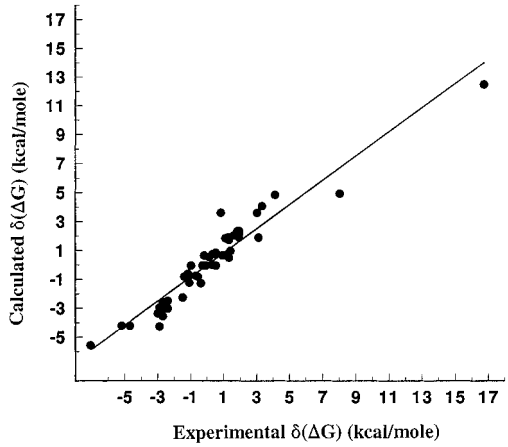

FIGURE 4: Linear regression plot of experimentally measured free energy differences determined from data given in Table 2 of SantaLucia et al. (*12*) with calculated values using the nearest-neighbor stacking parameters evaluated in this study (Table 2, column with all sites included). Values of R^2, slope, and intercept were 0.91, 0.84, and 0.014, respectively.

was randomly selected as a reference to calculate $\delta(\Delta G)$ values for the DNAs within the size class. Grouping the DNA oligomers by length and evaluating free energy differences between DNAs of the same length were done to minimize length-dependent influences on ΔG°_{37} values. A total of 10 reference DNAs and 49 DNAs produced the experimental $\delta(\Delta G)$ values. Figure 4 shows a plot of the predicted vs experimental $\delta(\Delta G)$ values. The correlation coefficient, $R^2 = 0.91$, is comparable to the correlation between the TGGE data and the TGGE-derived stacking parameter set (Figure 3A).

DISCUSSION

The analysis of DNA transition curves by TGGE provided a new approach for evaluating the DNA stacking free energy parameters. Although the method does not provide information on entropy and enthalpy changes, it allows one to analyze the effect of single base pair changes in the context of a long DNA. The influence of the strands−duplex nucleation parameter is eliminated.

The hierarchy of the stacking interactions determined under the low-salt TGGE conditions was in close agreement with the hierarchy obtained from an analysis of DNA oligomers in 1. M Na^+ (14). Additionally, close correspondence was observed between predicted and experimental free energy differences when stacking parameters evaluated in 0.115 and 1.0 M Na^+ were used to predict the low-salt TGGE data (Figure 3), and when the TGGE-derived parameters predicted results from DNA oligomer data in 1 M Na^+ (Figure 4). These comparisons indicate that the differences between DNA base pair stacking free energies are generally well conserved despite large changes in solvent ionic strength, and oligomer vs polymer contexts.

Results from Table 2 suggest that some of the deviations between calculated and experimental values of Figure 3A may be the result of local non-nearest-neighbor interactions. This table examined the constancy of the stacking energy parameter values when data from different sites were removed from the SVD analysis. The stacking parameter values and goodness of fit, Q, changed very little except when data from sites −43 or −34 were deleted. These two sites have tracts of A·T pairs, known to display unusual structural characteristics (*24*). This result may be a manifestation of interactions that are not explicitly accounted for by the nearest-neighbor model. Electrostatic interactions are expected to be longer range in the low ionic strength solvent employed, and may account for the observation. TGGE analysis of base pair substitutions at specific sites, and in other solvent conditions, may provide an approach to examine thermodynamic aspects of non-nearest-neighbor interactions in specific contexts.

REFERENCES

1. Yager, T. D., and von Hippel, P. H. (1991) *Biochemistry 30*, 1097−1118.
2. Petruska, J., Goodman, M. F., Boosalis, M. S., Sowers, L. S., Cheong, C., and Tinoco, I., Jr. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6252−6256.
3. Fodor, S. P. A., Rava, R. P., Huang, X. C., Pease, A. C., Holmes, C. P., and Adams, C. L. (1993) *Nature 364*, 555−556.
4. Southern, E. M. (1975) *J. Mol. Biol. 98*, 503−517.
5. Gotoh, O., and Tagashira, Y. (1981) *Biopolymers. 20*, 1033−1042.
6. Vologodskii, A. V., Amirikyan, B. R., Lyubchenko, Y. L., and Frank-Kamenetskii, M. D. (1984) *J. Biomol. Struct. Dyn. 2*, 131−148.
7. Wartell, R. M., and Benight, A. S. (1985) *Phys. Rep. 126*, 67−107.
8. Breslauer, K. J., Frank, R., Blöcker, H., and Marky, L. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3746−3750.
9. Delcourt, S. G., and Blake, R. D. (1991) *J. Biol. Chem. 266*, 15160−15169.
10. Doktycz, M. J., Goldstein, R. F., Paner, T. M., Gallo, F. J., and Benight, A. S. (1992) *Biopolymers 32*, 849−864.
11. Sugimoto, N., Nakano, S., Yoneyama, M., and Honda, K. (1996) *Nucleic Acids Res. 24*, 4501−4505.
12. SantaLucia, J., Jr., Allawi, H. T., and Seneviratne, P. A. (1996) *Biochemistry. 35*, 3555−3562.

13. Goldstein, R. F., and Benight, A. S. (1992) *Biopolymers 32*, 1679−1693.
14. Allawi, H. T., and SantaLucia, J., Jr. (1997) *Biochemistry 36*, 10581−10594.
15. SantaLucia, J., Jr. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 1460−1465.
16. Tatti, K. M., and Moran, C. P. (1985) *Nature 314*, 190−192.
17. Ke, S.-H., and Wartell, R. M. (1993) *Nucleic Acids Res. 21*, 5137−5143.
18. Ke, S.-H., and Wartell, R. M. (1995) *Biochemistry 34*, 4593−4600.
19. Clark, J. M. (1988) *Nucleic Acids Res. 16*, 9677−9686.
20. Wartell, R. M., Hosseini, S. H., and Moran, C. P., Jr. (1990) *Nucleic Acids Res. 18*, 2699−2705.
21. Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992) *Numerical Recipes in Fortran 77*, 2nd ed., Cambridge University Press, New York.
22. Lawson, C. L., and Hanson, R. J. (1974) *Solving Least Squares Problems*, Prentice Hall, Inc., Englewood Cliffs, NJ.
23. Gray, D. M. (1997) *Biopolymers 42*, 795−810.
24. Crothers, D. M., Haran, T. E., and Nadeau, J. G. (1990) *J. Biol. Chem. 265*, 7093−7096.

BI981093O