BIOCHEMISTRY
including biophysical chemistry & molecular biology

ARTICLE

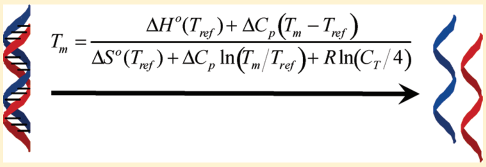
pubs.acs.org/biochemistry

Correcting for Heat Capacity and 5′-TA Type Terminal Nearest Neighbors Improves Prediction of DNA Melting Temperatures Using Nearest-Neighbor Thermodynamic Models

Curtis B. Hughesman,[†,‡] Robin F. B. Turner,[†,§] and Charles Haynes*[,†,‡]

[†]Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, BC, Canada V6T 1Z4
[‡]Department of Chemical and Biological Engineering, University of British Columbia, 2360 East Mall, Vancouver, BC, Canada V6T 1Z4
[§]Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver, BC, Canada V6T 1Z1

Supporting Information



ABSTRACT: Nearest-neighbor thermodynamic (NNT) models currently provide some of the most accurate predictions of melting thermodynamics, including melting temperature (T_m) values, for short DNA duplexes. Inherent to all existing NNT models is the assumption that $\Delta H°$ and $\Delta S°$ for the helix-to-coil transition are temperature invariant. Here we investigate the impact that this zero-ΔC_p assumption has on the accuracy of T_m predictions for 128 DNA duplexes. Previous and new melting thermodynamic data are analyzed to establish an estimate of ΔC_p^{bp}, the heat capacity change per base pair, of 42 ± 16 cal mol^{-1} K^{-1} bp^{-1}, as well as an optimal thermodynamic reference temperature (T_{ref}) of 53 ± 5 °C. These results were used to modify the unified NNT model to properly account for the temperature dependence of $\Delta H°$ and $\Delta S°$ and thereby extend the range over which T_m is accurately predicted. This new approach is shown to be especially useful for duplexes that melt at a T_m greater than 70 °C. Thermodynamic data collected by differential scanning calorimetry (DSC) for 16 duplexes designed to melt over a broad temperature range were used to verify the values of ΔC_p^{bp} and T_{ref} and to show that ΔC_p^{bp} is essentially constant above 37 °C. Additional DSC analysis of 12 duplex sequences containing all 10 nearest neighbors allowed for errors associated with different terminal nearest neighbors to be examined and showed that duplexes containing one or more terminal 5′-TA groups are significantly more stable than predicted by the unified NNT model. A correction to improve prediction of the hybridization thermodynamics of duplexes with terminal 5′-TA groups is provided.

The sequence-based prediction of the melting temperature (T_m) of short duplex DNA is critical to the proper design of oligonucleotides used as primers,[1] probes,[2,3] or antisense agents.[4] Derived from a statistical thermodynamic theory of the DNA helix-to-coil transition first proposed by Zimm[5] and later extended by Gray and Tinoco,[6] nearest-neighbor thermodynamic (NNT) models are widely used to predict melting thermodynamics of nucleic acid duplexes. For short Watson−Crick DNA duplexes, several versions of the NNT model,[7−9] each possessing a set of regressed nearest-neighbor (NN) parameters, have been reported over the past few decades and have served to improve the accuracy of predicted T_m values and associated changes in enthalpy ($\Delta H°$), entropy ($\Delta S°$), and Gibbs energy ($\Delta G°$). These models ignore any heat capacity change (ΔC_p) associated with hybridization of oligonucleotides, in part because reliable measurement of ΔC_p is difficult. However, the development of very sensitive differential scanning calorimetry (DSC) and isothermal titration calorimetry (ITC) instruments has led to an increasing number of publications reporting accurate non-zero values of ΔC_p for the melting of short DNA duplexes[10−15] and polymeric nucleic acids.[16,17] For such sequences, the heat capacity change per base pair (ΔC_p^{bp}) generally lies within the range of 30−60 cal mol^{-1} K^{-1} bp^{-1}.[18] The recognition of the non-zero value of ΔC_p^{bp} and the resulting temperature dependence of $\Delta H°$ and $\Delta S°$ have yielded a number of important insights into the hybridization thermodynamics of complementary oligonucleotides, including improved molecular interpretation of the well-known dependence of duplex RNA and DNA stabilities and conformations on nucleotide sequence.[17] It has also permitted the reconciliation of differences in NNT model parameters derived from either oligomeric or polymeric DNA.[19] However, attention has not been given to properly accounting for the temperature dependence of $\Delta H°$ and $\Delta S°$ within the NNT model to improve the prediction of T_m values and hybridization thermodynamics, particularly at higher temperatures.

A thorough analysis by Allawi and SantaLucia[20] permitted the conversion of existing versions of the NNT model, and their associated NN parameter sets, to a common thermodynamic reference state. This led to the welcomed establishment of a "unified" NNT model[21] that currently provides one of the most accurate methods for the prediction of melting thermodynamics

Received: October 14, 2010
Revised: February 9, 2011
Published: February 16, 2011

for short DNA duplexes.[22] The unified NNT model accurately estimates T_m and thermodynamic changes such as ΔG° at temperatures near 50 °C but becomes less accurate with an increasing departure of T_m from this temperature window.[23] This can prove to be problematic, as a number of new techniques common to molecular biology and clinical laboratories use high-resolution melting at elevated temperatures to identify subtle sequence differences, including single-nucleotide polymorphisms (SNPs) and somatic mutations.[24,25] To perform optimally, these assays require accurate T_m predictions for sequence-specific probes designed to melt at higher temperatures. It would therefore be beneficial to extend the range of temperatures over which NNT models accurately predict T_m.

In this paper, we first investigate the effect that the zero-ΔC_p assumption used in the development and application of current NNT models has on the accuracy of T_m predictions for higher-stability duplexes. We then measure ΔC_p^{bp} and remove the zero-ΔC_p assumption from the unified NNT model to create a method that extends the range over which T_m is accurately predicted. By properly accounting for the temperature dependence of ΔH° and ΔS°, we also identify a limitation in the unified NNT model that serves to increase T_m prediction errors. Specifically, duplex sequences with at least one terminal 5′-TA sequence were found to be more stable than those predicted by the model. We therefore designed a library of sequences with equal representation of all NNs, but with different terminal base pairs and/or terminal nearest neighbors, to elucidate the nature of the stability increase in duplexes with terminal 5′-TA sequences. A further correction to the unified NNT model for duplexes with one or two 5′-TA termini is proposed and shown to provide a significantly more accurate prediction of T_m for these duplexes.

MATERIALS AND METHODS

DNA Synthesis and Purification. Desalted and high-performance liquid chromatography-purified single-strand oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), with the purity of each product determined by capillary electrophoresis (CE) or calculated from the measured coupling efficiency to establish the concentration of the full-length product for each sample analyzed by DSC. The average purity of oligonucleotides following high-performance liquid chromatography processing was 97.9% (±1.8%), while the base coupling efficiency regressed from CE chromatograms was determined to be 99.3% for the desalted oligonucleotides.

Differential Scanning Calorimetry. Single-strand oligonucleotides were dissolved in 10 mM Na_2HPO_4 buffer (pH 7) containing 1 M NaCl and 1 mM Na_2EDTA. For the three duplexes for which different salt concentrations were tested, a NaCl concentration of 50 or 250 mM was used. The concentration of each single strand was determined using absorbance readings at 260 nm and 80 °C and extinction coefficients provided by the manufacturer. Duplex DNA sample solutions, each containing an equimolar concentration of the two complementary strands, were then prepared to the desired total molar concentration (C_T) of 50, 75, or 100 μM. Samples were degassed for 7 min under gentle stirring immediately prior to DSC analysis. All melting experiments were performed on a VP-DSC differential scanning calorimeter (Microcal, Inc., Northampton, MA). The nominal cell volume of the instrument was 0.52811 mL, and the scan rate used to collect data from 1 to 100 °C was 1 °C/min. For each experiment, buffer versus buffer excess thermal power baselines were measured immediately before the degassed duplex solution was loaded into the DSC sample cell and then averaged. The duplex-containing sample solution was then dynamically loaded on a thermal down scan, and several denaturation−renaturation scans were performed to verify the reversibility of the melt transition. Except for occasional instabilities in the first denaturation scan associated with the sample loading process, repeat denaturation scans for a given sample were superimposable, indicating that the helix-to-coil transition was fully reversible. The average of these denaturation scans was computed, and the average buffer versus buffer baseline was subtracted from it to produce an excess heat capacity (ΔC_p^{ex}) versus temperature curve for the helix-to-coil transition.

Regression of Melting Thermodynamics Data. Excess heat capacity data were normalized for concentration and then analyzed using two different methods to determine melting thermodynamics. The first method is model-independent and computes the calorimetric enthalpy change for the helix-to-coil transition (ΔH°_{cal}) as the area enclosed by the transition peak and the underlying pretransition-to-posttransition baseline; T_m^{cal} is the temperature at the area midpoint of the enclosed melting transition.

In the second method, melting transition data were fit to the nonlinear two-state transition model to determine ΔC_p, $\Delta H^\circ_{2\text{-st}}$, and $T_m^{2\text{-st}}$ described in the following equations:

$$C_p^{ex}(T) = b_{post}^{cal} + m_{post}^{cal}T - \Delta C_p(1 - \theta(T)) + \frac{[\Delta H^\circ + \Delta C_p(T - T_m)]^2}{RT^2}\left[\frac{1 - \theta(T)}{2/\theta(T) - 1}\right] \quad (1)$$

$$\theta(T) = \frac{-K(T) + \sqrt{[K(T)]^2 + 2K(T)C_T}}{C_T} \quad (2)$$

$$K(T) = \left(\frac{C_T}{4}\right)\exp\left[\frac{-\Delta H^\circ_{2\text{-st}}}{RT}\left(1 - \frac{T}{T_m^{2\text{-st}}}\right) - \frac{\Delta C_p}{RT}\left(T - T_m^{2\text{-st}} - T\ln\frac{T}{T_m^{2\text{-st}}}\right)\right] \quad (3)$$

where m_{post}^{cal} and b_{post}^{cal} are the slope and intercept, respectively, of the posttransition baseline, which was remarkably stable and easily fit to provide a reliable reference for regression of thermodynamic changes. θ(T) is the fraction of molecules in the single strand (random coil) state and $K(T)$ is the equilibrium constant for the dissociation (helix-to-coil) reaction. A typical melting transition measured by DSC and the best fit of the two-state transition model to those data are shown in Figure 1.

Error Analysis. Independent experiments were performed in triplicate for duplexes C1 and T1 to determine repeatability and to obtain an estimate of experimental error. Standard deviations for each thermodynamic parameter were determined from these repeat runs and are reported in Table 2. Experimental errors for the remaining duplexes, each studied by independent runs performed in duplicate, were estimated from errors for C1 and T1 to be ±33%, ±3%, ±3%, and ±0.5 °C for the thermodynamic parameters ΔC_p, ΔH°, ΔS°, and T_m, respectively.

RESULTS AND DISCUSSION

Our initial study focused on comparing T_m values predicted by available NNT models to published data for 128 cDNA duplexes,[26−28] with the aim of determining if the zero-ΔC_p assumption in

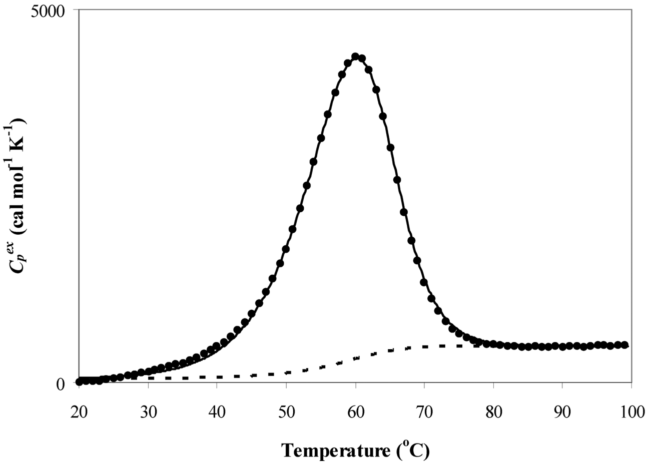


Figure 1. Differential scanning calorimetry data (●) and two-state model fit for the helix-to-coil transition of duplex sequence T1 (TTCATAGCCGT/ACGGCTATGAA). Sample preparation and properties are as reported in Materials and Methods.

current NNT models results in any prediction bias or systematic error. The duplexes included in the study have lengths ranging from 8 to 30 bases and were selected on the basis of a number of criteria, including data quality and demonstration of two-state melting behavior,[27,28] and the fact that they were not used for regression of the reported NNT model parameters.[7−9,20,21] Also, for each sequence, the true T_m, rather than the value of T_{max},[29] was available at 1 M Na$^+$, and therefore, no salt correction was required for comparison with model predictions. For each sequence, the experimental melting temperature (T_m^{exp}) was subtracted from the predicted melting temperature (T_m) to determine the model error (ΔT_m). Predictions from several available NNT models for short DNA duplexes were evaluated.[7−9] The NNT model employing the unified parameter set,[20,21] previously and hereafter termed the unified NNT model, was found to provide the lowest average ΔT_m and was therefore chosen as the reference (i.e., zero ΔC_p) model used for all subsequent NNT model predictions of melting thermodynamics reported in this paper.

Figure 2 shows ΔT_m values for the 128 duplexes within our study set. The T_m^{exp} ranged from 33.6 to 88.6 °C and therefore provided a good test of unified NNT model predictions at temperatures different from 50 °C. The average ΔT_m for the 128 sequences was 0.7 ± 1.8 °C, indicating that the unified NNT model typically provides good estimates of T_m. However, for sequences with a T_m^{exp} above 70 °C, the unified NNT model systematically overpredicts T_m such that $\Delta T_m = 1.6 \pm 1.8$ °C. We observed similar trends in ΔT_m error when analyzing other NNT models in the same manner, and these results suggest that model improvement could be realized by properly defining the temperature dependence of the sequence-specific $\Delta H°$ and $\Delta S°$ values computed by the model and used to predict T_m. Similarly, at lower temperatures, accounting for ΔC_p might serve to improve structural prediction algorithms for short DNA duplexes, which has recently been discussed.[18]

Introduction of ΔC_p into the Unified NNT Model Improves T_m Predictions. All previously published NNT parameter sets, including that used in the unified NNT model, have been determined with the assumption that ΔC_p is zero for DNA hybridization. An obvious benefit of this simplifying assumption is that it allows NNT parameters for $\Delta H°$, $\Delta S°$, and $\Delta G°$ to be regressed directly from experimental data without the need to correct for differences in T_m^{exp} within the data set. In principal, correcting a NNT model for a non-zero ΔC_p would require (re)collection of melting thermodynamics data, including the ΔC_p value, for a large set of duplex sequences to permit accurate regression of NNT parameters following correction of the data set to a common reference temperature (T_{ref}). The regressed NNT parameters are then specific to T_{ref}, and the T_m for a given sequence is predicted using eq 4

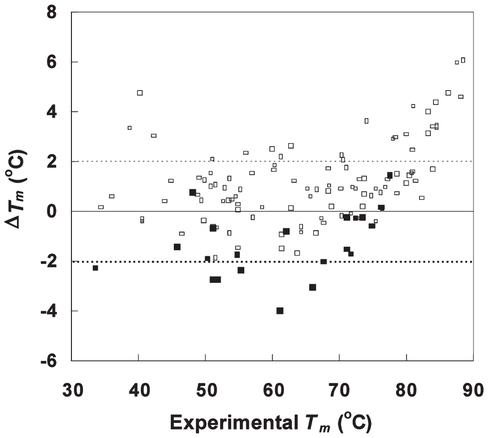


Figure 2. ΔT_m values for 128 test sequences with (■) and without (□) 5′-TA groups. For each duplex, the predicted T_m was determined by the unified NNT model.[20,21]

$$T_m = \frac{\Delta H°(T_{ref}) + \Delta C_p(T_m - T_{ref})}{\Delta S°(T_{ref}) + \Delta C_p \ln(T_m/T_{ref}) + R\ln(C_T/4)} \qquad (4)$$

where ΔC_p is assumed to be non-zero and temperature-independent. Although this approach arguably provides the most rigorous strategy for addressing the observed bias in the universal NNT model at high temperatures, its implementation is troublesome, as the large existing set of thermodynamic data for duplex DNA sequences that undergo the helix-to-coil transition according to two-state thermodynamics does not include accurate ΔC_p values for the sequences and therefore could not be utilized for parameter regression. What remains is a rather limited set of ΔC_p-inclusive melting thermodynamics data that is not sufficiently large to permit accurate regression of the required set of new NNT parameters needed to compute $\Delta H°(T_{ref})$ and $\Delta S°(T_{ref})$ in eq 4.

An alternative but less rigorous approach is to introduce a non-zero ΔC_p estimate into eq 4 such that it and the existing unified NNT model (and associated NN parameter set) can then be used to more accurately predict T_m values and helix-to-coil thermodynamics for duplexes that melt at temperatures different from T_{ref}. Existing thermodynamic data and the many advantages of the unified NNT model are thereby fully exploited.

As noted previously, SantaLucia[23] found that the unified NNT model most accurately predicts $\Delta G°$ at temperatures near 50 °C, which suggests that model estimates of $\Delta H°$ and $\Delta S°$ are also most reliable near this temperature. This result, which supports a T_{ref} of 50 °C in eq 4, is not surprising, as it is very close to the average T_m^{exp} (47 ± 11 °C) for the 108 duplexes originally used to regress the unified NNT parameter set.[20]

The value of ΔC_p^{bp} could likewise be drawn from existing literature. However, it is possible to estimate the values of ΔC_p^{bp} and T_{ref} using the T_m^{exp} data for the 128 duplex sequences within

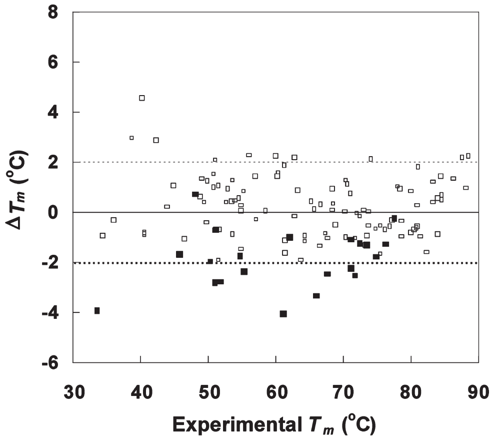


Figure 3. ΔT_m values for 128 test sequences with (■) and without (□) 5′-TA groups. For each duplex, the predicted T_m was determined by eq 4 with a ΔC_p^{bp} of 42 cal mol^{-1} K^{-1} bp^{-1} and a T_{ref} of 53 °C.

our test set. Regression of the required ΔC_p^{bp} and T_{ref} parameters is then achieved by minimizing the error of the residual (χ^2) between T_m predicted by eq 4 and T_m^{exp} as shown in eq 5

$$\chi^2 = \sum_i (T_m - T_m^{exp})^2 \qquad (5)$$

An average ΔC_p^{bp} of 42 ± 16 cal mol^{-1} K^{-1} bp^{-1} and a T_{ref} of 53 ± 5 °C were thereby obtained. Uncertainties in ΔC_p^{bp} and T_{ref} parameters were estimated from regression analysis and correspond to the upper and lower limits that result in a 10% increase in the value of χ^2. We note that the regressed T_{ref} is quite close to 50 °C. Moreover, the value of ΔC_p^{bp} falls well within the consensus range of previously reported experimental values[18] and is also in excellent agreement with the intrinsic ΔC_p^{bp} value of 48 ± 10 cal mol^{-1} K^{-1} bp^{-1} that has been reported on the basis of theoretical arguments.[12]

As demonstrated in Figure 3, the application of eq 4, where $\Delta H°(T_{ref})$ and $\Delta S°(T_{ref})$ are computed using the unified NNT model and ΔC_p is given by $n_{bp}\Delta C_p^{bp}$ (n_{bp} is the number of base pairs in the duplex), results in an improvement in both the overall accuracy and precision of T_m predictions, with the average ΔT_m now decreased from 0.7 ± 1.8 to −0.2 ± 1.4 °C for the 128 duplexes in the test set. This improvement can be attributed to better T_m predictions for duplexes melting above 70 °C, with the average ΔT_m for these duplexes improving to −0.2 ± 1.1 °C, as compared to 1.6 ± 1.8 °C when ΔC_p is assumed to be zero. There was no significant improvement in the average ΔT_m for duplexes with a T_m^{exp} between 40 and 70 °C, as errors were −0.1 ± 1.6 °C using eq 4 and 0.1 ± 1.6 °C using the existing unified NNT model.

Finally, we note that for a number of duplex sequences that melt at ≤40 °C, T_m is not predicted well by either method (zero or non-zero ΔC_p models). For many or all of these sequences, this is likely due to significant single-stranded ordering at lower temperatures that is not included in the unified NN thermodynamic parameters. This phenomenon has been carefully studied by others[11−13,30,31] and is not the focus of our investigation.

Regressed ΔC_p^{bp} and T_{ref} Values Are Supported by DSC Data. The accuracy of our method, embodied in eq 4, for predicting melting thermodynamics of short DNA duplexes depends on the validity of three key model assumptions. (1) A T_{ref} can be identified at which the unified NNT model provides accurate estimates of $\Delta H°(T_{ref})$ and $\Delta S°(T_{ref})$. (2) A value of ΔC_p^{bp} can be established and used to estimate the ΔC_p required in eq 4. (3) The said value of ΔC_p^{bp} is temperature-independent over the temperature range where estimates of T_m are typically required for primer and probe design.

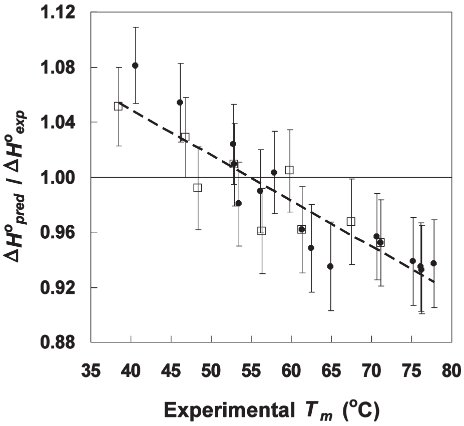


Figure 4. Relationship between predicted (unified NNT model) and experimental $\Delta H°$ values as a function of T_m for 16 duplexes with DSC data collected at 1 M NaCl (●). Duplexes with DSC data collected at various salt concentrations are also shown (□). The trend in the data (−−−) is also shown.

To address the validity of these assumptions, we collected DSC data on 16 duplexes at 1022 mM Na$^+$ with T_m values ranging from 40.6 to 77.8 °C. DSC data were also collected at two additional total salt concentrations, 272 and 72 mM Na$^+$, for a subset of these sequences consisting of three 12-mers with a GC content of 25, 50, or 75%. For each sample, the ratio of the predicted enthalpy ($\Delta H°_{pred}$) to the experimental enthalpy ($\Delta H°_{exp}$) was determined and plotted against the corresponding T_m^{exp} (Figure 4). The results show that $\Delta H°_{pred}$ agrees reasonably well with $\Delta H°_{exp}$ between 50 and 60 °C when experimental errors are taken into account. The trend line suggests a T_{ref} near 55 °C is optimal, which is in good agreement with our previously regressed value for T_{ref} of 53 ± 5 °C. It also supports our assumption that the unified model can provide reliable estimates of $\Delta H°(T_{ref})$ and $\Delta S°(T_{ref})$ at our chosen T_{ref} of 53 °C.

In accordance with previous experimental studies,[10,13,14,16,17,32] our data also show that $\Delta H°$ is temperature-dependent. DSC analysis of the 16 duplexes at 1022 mM Na$^+$ yielded an average ΔC_p^{bp} of 43 ± 9 cal mol^{-1} K^{-1} bp^{-1}, while an average ΔC_p^{bp} value of 47 ± 6 cal mol^{-1} K^{-1} bp^{-1} was obtained for the three duplexes studied at various salt concentrations. Both of these values are in good agreement with our average ΔC_p^{bp} value of 42 ± 16 cal mol^{-1} K^{-1} bp^{-1} regressed from a much larger data set. In addition, Figure 5 shows, at least for the 16 sequences examined here, that there is no statistically significant temperature dependence for ΔC_p^{bp} over the temperature range of 38.5−77.8 °C. We recognize that this assumption is likely not valid for all duplexes, especially those exhibiting a T_m lower than 40 °C or significant ordering (i.e., preorganization) of the single strands, including simple single-stranded base stacking or more sequence-specific structures such as hairpins or homodimers.[30] As noted previously, the contribution of significant single-stranded structure to ΔC_p and other thermodynamic values such as $\Delta H°$ and $\Delta S°$ for the helix-to-coil transition has been well documented and is not generally well predicted by current NNT models.[11−13,31] Although it has been suggested that the

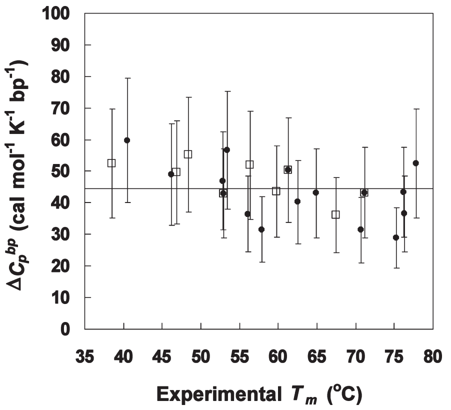


Figure 5. Measured ΔC_p^{bp} values as a function of T_{m} for 16 duplexes with DSC data collected at 1 M NaCl (●). Duplexes with DSC data collected at various salt concentrations are also shown (☐). The average ΔC_p^{bp} value of 45 cal mol^{-1} K^{-1} bp^{-1} (—) for the complete set of duplexes is also shown.

introduction of a temperature-dependent ΔC_p into the modeling of duplexes with a low $T_{\mathrm{m}}^{\mathrm{exp}}$ may improve predictions,[13] there are currently insufficient data available for this concept to be properly implemented.

Duplexes Terminating in a 5′-TA Group Have Statistically Significant ΔT_{m} Errors. The results in Figure 3 show that the application of eq 4 and its associated use of a non-zero ΔC_p significantly reduce the average error in predicted T_{m} values. However, a number of duplex sequences within the test set remain considerably more stable than predicted by eq 4, with some showing ΔT_{m} errors greater than −2 °C. A common trait for all of these duplexes is that the sequences contain at least one terminal 5′-TA group. To investigate this further, we conducted a complete analysis of ΔT_{m} errors as a function of duplex terminus chemistry, and the results are listed in Table 1. The 128 reference duplexes were categorized by their terminal base pairs, terminal base pair direction, and terminal nearest neighbors to determine if ΔT_{m} errors associated with a certain terminal sequence were significant. Of the 16 possible terminal nearest neighbors, only duplexes ending in a NN containing a 5′-TA group had an average ΔT_{m} prediction error (−1.8 ± 1.2 °C) greater than the standard deviation of ±1.4 °C for the entire set of 128 duplexes, indicating that the presence of a terminal 5′-TA sequence has a statistically significant stabilizing effect that is not included in current NNT models.

Table 1. Average ΔT_{m} (°C) Values and Errors Associated with Terminal Base Pairs and Terminal Nearest Neighbors

terminal base pair		terminal base pair direction[a]		terminal NN[a]	
A/T	−0.4 ± 1.5	EA/TE′	0.2 ± 1.4	EAA/TTE′	−0.8 ± 0.7
				EAC/GTE′	0.7 ± 1.2
				EAG/CTE′	0.5 ± 1.7
				EAT/ATE′	−0.7 ± 1.2
		ET/AE′	−0.7 ± 1.6	**ETA/TAE′**	**−1.8 ± 1.2**
				ETC/GAE′	0.3 ± 0.8
				ETG/CAE′	0.4 ± 1.2
				ETT/AAE′	0.0 ± 1.4
G/C	0.1 ± 1.4	EC/GE′	0.2 ± 1.3	ECA/TGE′	0.2 ± 1.1
				ECC/GGE′	0.2 ± 0.9
				ECG/CGE′	0.8 ± 1.2
				ECT/AGE′	−0.8 ± 1.3
		EG/CE′	0.0 ± 1.4	EGA/TCE′	−0.1 ± 1.7
				EGC/GCE′	0.3 ± 1.3
				EGG/CCE′	0.7 ± 1.0
				EGT/ACE′	−0.6 ± 1.2

[a] E denotes the 5′ end and E′ the 3′ end.

Recognizing that the library of published thermodynamic data used to generate the errors reported in Table 1 was not designed to isolate the contribution of the termini to duplex stability, we also created and then characterized by DSC a new library of short DNA sequences specifically designed to more cleanly segregate the contributions to duplex stability of internal versus terminal nearest neighbors. Table 2 lists the twelve 11-mer duplexes we designed to contain all 10 internal nearest-neighbor base pairs represented equally while varying the terminal base pairs and terminal nearest neighbors. Melting thermodynamic data we collected by DSC are also reported for this library and were used to determine the thermodynamic contributions of different terminal groups. For all sequences in the library, there is good agreement in the thermodynamic parameters determined by the two-state model and by direct integration of the calorimetric data. Agreement between these two methods of data analysis is a necessary but not sufficient condition for indicating that the melting transition obeys two-state thermodynamics.[33] The average ΔC_p for the twelve 11-mer duplexes was determined to be 378 ± 125 cal K^{-1} mol^{-1} or 34 ± 11 cal mol^{-1} K^{-1} bp^{-1}, which is consistent with our previously regressed value of ΔC_p^{bp}.

The sequences in Table 2 have been segregated such that those in a given cluster (e.g., C1−C5 or T1−T7) are predicted by the unified NNT model to have identical melting thermodynamics because they share the exact same composition of nearest neighbors and terminal base pairs. For each cluster, the average T_{m} agrees reasonably well with the average $T_{\mathrm{m}}^{\mathrm{exp}}$. Average sequence-dependent trends in the experimental data are also reasonably well predicted by the unified NNT model. For example, 11-mer sequences with terminal G-C base pairs on average hybridize with a greater enthalpic driving force ($\Delta\Delta H°$ = −5.4 ± 2.5 kcal mol^{-1}) and more entropy loss ($\Delta\Delta S°$ = −16.1 ± 6.9 cal mol^{-1} K^{-1}) compared to those sequences terminating in A-T base pairs but not containing a terminal 5′-TA sequence. Within experimental error, the unified NNT model captures this, predicting that $\Delta\Delta H°$ = −4.4 kcal mol^{-1} and $\Delta\Delta S°$ = −13.8 cal mol^{-1} K^{-1}.

Several insights can be drawn from comparison of model predictions with the acquired thermodynamic data, but among the most important for the purposes of this study is the confirmation that 5′-TA termini provide a duplex stabilizing effect that is not included in the unified NNT model[20,21] or in other NNT models.[7−9] In particular, 11-mer sequences in the T1−T7 cluster having a single 5′-TA terminus (T5−T7) exhibit a $T_{\mathrm{m}}^{\mathrm{exp}}$ that is on average 2.0 ± 0.5 °C higher than the average $T_{\mathrm{m}}^{\mathrm{exp}}$ for the four sequences devoid of a 5′-TA terminus. This increase in T_{m} is driven by an average $\Delta\Delta S°$ of 13.5 ± 6.4 cal mol^{-1} K^{-1} that serves to reduce the entropic penalty accompanying duplex formation. That entropy gain is only partially compensated by a weaker enthalpy of interaction ($\Delta\Delta H°$ = 4.0 ± 2.3 kcal mol^{-1}) between the complementary bases of the 5′-TA terminus.

It is interesting to note that although 5′-TA NN sequences stabilize the duplex, internal TA NN sequences are predicted by the unified NNT model[21] and other NNT models[7−9] to provide the smallest contribution to duplex stability of all 10 possible

Table 2. Measured Thermodynamic Data for the Helix-to-Coil Transition of 11-mer DNA Duplexes Used To Study Duplex End Effects[a]

			two-state model			calorimetric model		
name	sequence	ΔC_p (cal mol^{-1} K^{-1})	$\Delta H°_{2\text{-st}}$ (kcal/mol)	$\Delta S°_{2\text{-st}}$ (cal mol^{-1} K^{-1})	$T_m^{2\text{-st}}$ (°C)	$\Delta H°_{cal}$ (kcal/mol)	$\Delta S°_{cal}$ (cal mol^{-1} K^{-1})	T_m^{cal} (°C)
C1[b]	CTACGCATTCC	448 ± 147	82.5 ± 2.5	226.5 ± 7.3	59.4 ± 0.5	84.5 ± 4.6	232.3 ± 13.3	59.7 ± 0.5
C2	CTAACGGATGC	328	83.8	230.2	59.6	85.3	234.4	59.9
C3	CTATTGGCGAC	386	82.3	225.0	60.5	83.8	229.5	60.5
C4	CGTATTCAGGC	403	87.9	243.4	58.7	88.6	245.0	59.0
C5	CAATACGCCTC	403	78.8	215.7	58.8	77.7	212.2	59.2
average (C1−C5)		**394**	**83.1**	**228.2**	**59.4**	**84.0**	**230.7**	**59.6**
T1[b]	TTCATAGCCGT	356 ± 74	79.7 ± 1.3	218.1 ± 4.0	59.4 ± 0.1	80.4 ± 1.2	219.9 ± 3.6	59.6 ± 0.1
T2	TTCCGTAGCAT	280	74.6	202.9	59.2	74.9	203.6	59.4
T3	TGCGGATAAGT	308	79.7	217.9	59.6	79.7	217.7	59.7
T4	TCGGCTATTGT	367	78.2	214.5	58.2	77.6	212.4	58.3
average (T1−T4)		**328**	**78.1**	**213.3**	**59.1**	**78.1**	**213.4**	**59.3**
T5	TACTCCGCATT	478	73.7	199.4	60.5	74.0	199.9	60.9
T6	TAGACCGCAAT	320	74.4	200.8	61.3	74.1	199.8	61.4
T7	TATCGTTGCCT	463	74.1	199.8	61.4	74.0	199.4	61.8
average (T5−T7)		**421**	**74.1**	**200.0**	**61.1**	**74.0**	**199.7**	**61.4**
average (T1−T7)		**368**	**76.4**	**207.6**	**59.9**	**76.4**	**207.5**	**60.2**

[a] DSC data were collected at a C_T of 75 μM for all duplexes. [b] Errors for C1 and T1 determined from triplicate repeats. Estimated errors for other duplexes as stated in Materials and Methods.

unique NN sequences. This indicates that the pairing and stacking interactions of a TA NN sequence within a duplex are relatively poor. However, when a 5′-TA NN sequence is placed at a terminal position, that same weakness in base stacking and pairing energy provides for greater configurational freedom (entropy) at the 5′-TA terminus, resulting in a net stabilization of the duplex.

Correcting T_m Predictions for Duplexes Containing 5′-TA Type Termini. Our proposed model, embodied in eq 4, for predicting helix-to-coil transition thermodynamics and melting temperatures for short duplex DNA requires values for $\Delta H°(T_{ref})$ and $\Delta S°(T_{ref})$, which we compute using the unified NNT model. In that model, $\Delta H°$ is computed using 10 NN parameters (ΔH_i, where i represents the type of NN) and two additional "chain-initiation/end-effect" parameters (ΔH_j^{init}, where j represents the type of end, C-G or A-T) that account for the terminal base pairs as well as duplex initiation.[20,21,23] Likewise, a set of 10 NN parameters (ΔS_i), two end-effect parameters (ΔS_j^{init}), and one theoretically derived chain symmetry parameter (ΔS^{sym}) are used to compute $\Delta S°$. Given our findings that terminal 5′-TA groups are thermodynamically distinct from internal TA NN sequences, we therefore sought to determine enthalpic and entropic corrections, denoted $\Delta\Delta H_{ETA/TAE'}$ and $\Delta\Delta S_{ETA/TAE'}$, where E and E′ denote the 5′ and 3′ ends, respectively, that account for the excess duplex stabilizing effect of placing a TA type NN at a strand terminus. Values of $\Delta H°(T_{ref})$ and $\Delta S°(T_{ref})$ for duplexes with terminal A-T base pairs, either with or without terminal 5′-TA groups, were computed using the two-state and calorimetric values of $\Delta H°$ and $\Delta S°$ at T_m (Table 2) and the regressed ΔC_p^{bp} value of 42 cal K^{-1} mol^{-1} bp^{-1}, which gives a ΔC_p for each 11-mer duplex of 462 cal K^{-1} mol^{-1}. These results were then used to regress the $\Delta\Delta H_{ETA/TAE'}$ and $\Delta\Delta S_{ETA/TAE'}$ correction parameters for a 5′-TA type terminus by subtracting the average $\Delta H°(T_{ref})$ for duplexes with terminal 5′-TA groups from that for duplexes with terminal A-T base pairs but no terminal 5′-TA sequence. A similar calculation was completed with the entropy data, and the resulting values of $\Delta\Delta H_{ETA/TAE'}$ and $\Delta\Delta S_{ETA/TAE'}$ are listed in Table 3.

Table 3. Thermodynamic Parameters[a] for ETA/TAE′ NN at a T_{ref} of 53 °C

method	$\Delta\Delta H°_{ETA/TAE'}$ (kcal/mol)	$\Delta\Delta S°_{ETA/TAE'}$ (cal mol^{-1} K^{-1})
two-state	4.9	16.1
calorimetric	5.0	16.6
average	5.0	16.4

[a] Reported parameters are for duplex hybridization to be consistent with the unified NNT model.[21]

For duplex sequences possessing a 5′-TA group at one or both termini, $\Delta H°(T_{ref})$ and $\Delta S°(T_{ref})$ are then computed as

$$\Delta H°(T_{ref}) = \sum_{i=1}^{10} n_i \Delta H_i + \sum m_j \Delta H_j^{init} + l\Delta\Delta H_{ETA/TAE'} \quad (6)$$

and

$$\Delta S°(T_{ref}) = \sum_{i=1}^{10} n_i \Delta S_i + \sum m_j \Delta S_j^{init} + l\Delta\Delta S_{ETA/TAE'} + \Delta S^{sym} \quad (7)$$

where n_i is the number of NNs of type i, m_j is the number of terminal base pairs of type j, and l is the number of 5′-TA sequences.

Equations 4, 6, and 7 were used to recalculate T_m and ΔT_m for all duplexes with one or two 5′-TA sequences. As shown in Figure 6, a significant improvement in model predictions is observed for all such duplexes, with the average ΔT_m decreasing from −1.8 ± 1.2 to −0.6 ± 1.5 °C, an average error that is similar to those found for all other terminal NN sequences (Table 1).

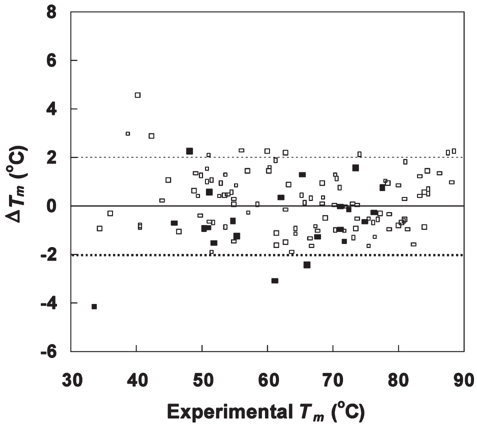


Figure 6. ΔT_m values for 128 test sequences with (■) and without (□) 5′-TA groups. For each duplex, the predicted T_m was determined using eq 4 with a ΔC_p of 42 cal mol^{-1} K^{-1} bp^{-1}, a T_{ref} of 53 °C, and values for $\Delta H°(T_{ref})$ and $\Delta S°(T_{ref})$ determined using eqs 6 and 7 and the 5′-TA correction parameters (Table 3).

CONCLUSIONS

Building on previous calorimetry studies that show ΔC_p exhibits positive, non-zero values,[10−19] we have used published and new melting thermodynamic data first to show how the simplifying assumption in current NNT models, that $\Delta H°$ and $\Delta S°$ are temperature-independent, biases predicted T_m values away from T_m^{exp} and then to regress an average value for ΔC_p^{bp} of 42 cal K^{-1} mol^{-1} bp^{-1} that can be used to more accurately compute the temperature dependence of $\Delta H°$ and $\Delta S°$. These results were combined with the unified NNT model to establish a method, embodied in eqs 4, 6, and 7, that more accurately predicts, across a broad temperature range, the sequence-dependent T_m and associated melting thermodynamics ($\Delta H°$, $\Delta S°$, and $\Delta G°$) of short complementary duplex DNA. The model exploits the ability of the unified NNT model to provide good estimates of the melting enthalpy and entropy at 53 °C, which was found in our work to be the optimal T_{ref} for model calculations. Key model assumptions were experimentally validated, including the finding that ΔC_p^{bp} is, to a first approximation, invariant over the temperature range within which the T_m values of probes and primers are typically designed.

Results from using the new method to reanalyze the melting thermodynamics for a large library of duplex DNA sequences suggested a significant stabilization of duplexes containing one or more 5′-TA termini. DSC data were therefore acquired and analyzed for duplexes designed to specifically isolate terminal NN effects. The results revealed that a terminal 5′-TA group provides excess stability to the duplex through an entropy increase that is likely related to duplex fraying and is only partially compensated by the expected enthalpy gain accompanying that fraying process. The data were also used to determine two parameters, $\Delta\Delta H°_{ETA/TAE'}$ and $\Delta\Delta S°_{ETA/TAE'}$, that can be incorporated into the model to improve T_m predictions for duplex sequences with a 5′-TA group.

ASSOCIATED CONTENT

Ⓢ Supporting Information. A table summarizing the DSC analysis for 16 duplexes used to generate Figures 3 and 4. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Telephone: (604) 822-5136. Fax: (604) 822-2114. E-mail: israels@chbe.ubc.ca.

Funding Sources

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and Canadian Institutes of Health Research (CIHR). C.H. receives salary support as the Canada Research Chair in Interfacial Biotechnology.

ABBREVIATIONS

NNT, nearest-neighbor thermodynamics; ΔC_p, heat capacity change; T_m, predicted melting temperature; T_{ref}, reference temperature; T_m^{exp}, experimental melting temperature; ΔT_m, difference between the predicted and experimental melting temperatures; DSC, differential scanning calorimetry; ITC, isothermal titration calorimetry.

REFERENCES

(1) Rychlik, W., Spencer, W. J., and Rhoads, R. E. (1990) Optimization of the annealing temperature for DNA amplification in vitro. *Nucleic Acids Res. 18*, 6409–6412.

(2) Rouillard, J. M., Zuker, M., and Gulari, E. (2003) OligoArray 2.0: Design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res. 31*, 3057–3062.

(3) Matveeva, O. V., Shabalina, S. A., Nemtsov, V. A., Tsodikov, A. D., Gesteland, R. F., and Atkins, J. F. (2003) Thermodynamic calculations and statistical correlations for oligo-probes design. *Nucleic Acids Res. 31*, 4211–4217.

(4) Matveeva, O. V., Mathews, D. H., Tsodikov, A. D., Shabalina, S. A., Gesteland, R. F., Atkins, J. F., and Freier, S. M. (2003) Thermodynamic criteria for high hit rate antisense oligonucleotide design. *Nucleic Acids Res. 31*, 4989–4994.

(5) Zimm, B. H. (1960) Theory of "Melting" of the Helical Form in Double Chains of the DNA Type. *J. Chem. Phys. 33*, 1349–1356.

(6) Gray, D. M., and Tinoco, I., Jr. (1970) A new approach to the study of sequence-dependent properties of polynucleotides. *Biopolymers 9*, 223–244.

(7) Breslauer, K. J., Frank, R., Blocker, H., and Marky, L. A. (1986) Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. U.S.A. 83*, 3746–3750.

(8) SantaLucia, J., Jr., Allawi, H. T., and Seneviratne, P. A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry 35*, 3555–3562.

(9) 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.

(10) Bruylants, G., Boccongelli, M., Snoussi, K., and Bartik, K. (2009) Comparison of the thermodynamics and base-pair dynamics of a full LNA:DNA duplex and of the isosequential DNA:DNA duplex. *Biochemistry 48*, 8473–8482.

(11) Holbrook, J. A., Capp, M. W., Saecker, R. M., and Record, M. T., Jr. (1999) Enthalpy and heat capacity changes for formation of an oligomeric DNA duplex: Interpretation in terms of coupled processes of formation and association of single-stranded helices. *Biochemistry 38*, 8409–8422.

(12) Jelesarov, I., Crane-Robinson, C., and Privalov, P. L. (1999) The energetics of HMG box interactions with DNA: Thermodynamic description of the target DNA duplexes. *J. Mol. Biol. 294*, 981–995.

(13) Mikulecky, P. J., and Feig, A. L. (2006) Heat capacity changes associated with DNA duplex formation: Salt- and sequence-dependent effects. *Biochemistry 45*, 604–616.

(14) Tikhomirova, A., Beletskaya, I. V., and Chalikian, T. V. (2006) Stability of DNA duplexes containing GG, CC, AA, and TT mismatches. *Biochemistry 45*, 10563–10571.

(15) Wu, P., Nakano, S., and Sugimoto, N. (2002) Temperature dependence of thermodynamic properties for DNA/DNA and RNA/DNA duplex formation. *Eur. J. Biochem. 269*, 2821–2830.

(16) Chalikian, T. V., Volker, J., Plum, G. E., and Breslauer, K. J. (1999) A more unified picture for the thermodynamics of nucleic acid duplex melting: A characterization by calorimetric and volumetric techniques. *Proc. Natl. Acad. Sci. U.S.A. 96*, 7853–7858.

(17) Tikhomirova, A., Taulier, N., and Chalikian, T. V. (2004) Energetics of nucleic acid stability: The effect of ΔC_p. *J. Am. Chem. Soc. 126*, 16387–16394.

(18) Mikulecky, P. J., and Feig, A. L. (2006) Heat capacity changes associated with nucleic acid folding. *Biopolymers 82*, 38–58.

(19) Rouzina, I., and Bloomfield, V. A. (1999) Heat capacity effects on the melting of DNA. 2. Analysis of nearest-neighbor base pair effects. *Biophys. J. 77*, 3252–3255.

(20) Allawi, H. T., and SantaLucia, J., Jr. (1997) Thermodynamics and NMR of internal G·T mismatches in DNA. *Biochemistry 36*, 10581–10594.

(21) SantaLucia, J., Jr. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U.S.A. 95*, 1460–1465.

(22) Chavali, S., Mahajan, A., Tabassum, R., Maiti, S., and Bharadwaj, D. (2005) Oligonucleotide properties determination and primer designing: A critical examination of predictions. *Bioinformatics 21*, 3918–3925.

(23) SantaLucia, J., Jr., and Hicks, D. (2004) The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct. 33*, 415–440.

(24) Erali, M., Voelkerding, K. V., and Wittwer, C. T. (2008) High resolution melting applications for clinical laboratory medicine. *Exp. Mol. Pathol. 85*, 50–58.

(25) Vossen, R. H., Aten, E., Roos, A., and den Dunnen, J. T. (2009) High-resolution melting analysis (HRMA): More than just sequence variant screening. *Hum. Mutat. 30*, 860–866.

(26) Levin, J. D., Fiala, D., Samala, M. F., Kahn, J. D., and Peterson, R. J. (2006) Position-dependent effects of locked nucleic acid (LNA) on DNA sequencing and PCR primers. *Nucleic Acids Res. 34*, e142.

(27) McTigue, P. M., Peterson, R. J., and Kahn, J. D. (2004) Sequence-dependent thermodynamic parameters for locked nucleic acid (LNA)-DNA duplex formation. *Biochemistry 43*, 5388–5405.

(28) Owczarzy, R., You, Y., Moreira, B. G., Manthey, J. A., Huang, L., Behlke, M. A., and Walder, J. A. (2004) Effects of sodium ions on DNA duplex oligomers: Improved predictions of melting temperatures. *Biochemistry 43*, 3537–3554.

(29) Owczarzy, R. (2005) Melting temperatures of nucleic acids: Discrepancies in analysis. *Biophys. Chem. 117*, 207–215.

(30) Ramprakash, J., Lang, B., and Schwarz, F. P. (2008) Thermodynamics of single strand DNA base stacking. *Biopolymers 89*, 969–979.

(31) Vesnaver, G., and Breslauer, K. J. (1991) The contribution of DNA single-stranded order to the thermodynamics of duplex formation. *Proc. Natl. Acad. Sci. U.S.A. 88*, 3569–3573.

(32) Rouzina, I., and Bloomfield, V. A. (1999) Heat capacity effects on the melting of DNA. 1. General aspects. *Biophys. J. 77*, 3242–3251.

(33) Zhou, Y., Hall, C. K., and Karplus, M. (1999) The calorimetric criterion for a two-state process revisited. *Protein Sci. 8*, 1064–1074.