Richard Owczarzy[1]
Peter M. Vallone[1]
Frank J. Gallo[1]
Teodoro M. Paner[2]
Michael J. Lane[2,3]
Albert S. Benight[1,2]*
[1] *Department of Chemistry, 845 West Taylor St., Room 4500, University of Illinois, Chicago, IL 60607*

[2] *Tm Technologies, Inc., 82 Cummings Park, Woburn, MA 01801*

[3] *Departments of Medicine and Microbiology/ Immunology, State University of New York— Health Science Center at Syracuse, Syracuse, NY 13207*

Received 6 November 1997; accepted 18 January 1998

Predicting Sequence-Dependent Melting Stability of Short Duplex DNA Oligomers

Abstract: *Many important applications of DNA sequence-dependent hybridization reactions have recently emerged. This has sparked a renewed interest in analytical calculations of sequence-dependent melting stability of duplex DNA. In particular, for many applications it is often desirable to accurately predict the transition temperature, or* t_m*, of short duplex DNA oligomers (~ 20 base pairs or less) from their sequence and concentration. The thermodynamic analytical method underlying these predictive calculations is based on the nearest-neighbor model. At least 11 sets of nearest-neighbor sequence-dependent thermodynamic parameters for DNA have been published. These sets are compared. Use of the nearest-neighbor sets in predicting* t_m *from the DNA sequence is demonstrated, and the ability of the nearest-neighbor parameters to provide accurate predictions of experimental* t_m*'s of short duplex DNA oligomers is assessed.* © 1998 John Wiley & Sons, Inc. Biopoly 44: 217–239, 1997

Keywords: *short duplex DNA; sequence-dependent hybridization reactions; sequence-dependent melting stability; oligomers*

Correspondence to: *Albert S. Benight
Contract grant sponsor: NIH
Contract grant number: GM39471

CCC 0006-3525/98/030217-23

INTRODUCTION

With the primary aim of predicting DNA stability from sequence alone, over the past 15 years a number of melting studies have been conducted to evaluate sequence-dependent thermodynamic stability of DNA in terms of nearest-neighbor (n-n) base pair interactions.[1–13] The duplex DNA samples studied have been long DNA restriction fragments ~ 150–4400 base pairs in length,[1–5] long repeating copolymers,[6,7,9] DNA dumbbells,[8] and very short (6–16 base pairs) linear synthetic oligomers.[9–12] Published studies based on the n-n model have reported thermodynamic values for calculating DNA stability from sequence.[1–13] A direct comparison of n-n parameters from different laboratories can be difficult because published sets were evaluated from measurements of DNA samples in different molecular environments and solvent ionic strengths.[8,14] For instance, the n-n set derived from melting curves of DNA dumbbells[8] was evaluated in 115 m*M* Na^+. The n-n sets reported by Wartell and Benight,[1] Gotoh and Tagashira,[2] Vologodskii et al.,[3] McCampbell et al.,[4] and Delcourt and Blake[5] originate from melting studies of DNA restriction fragments in ionic strengths ranging from 19.5 to 200 m*M* Na^+. The n-n sets of Breslauer et al.,[9] SantaLucia et al.,[10] Allawi and SantaLucia,[11] and Sugimoto et al.[12] were evaluated from analysis of optical melting curves of a variety of short synthetic DNA duplexes in 1*M* Na^+. Breslauer et al.[9] also included optical and calorimetric melting data on long synthetic repeating sequence DNA polymers in their database. The n-n set of Ornstein and Fresco[6] was developed by fitting experimental data of repeating DNA polymers with empirical potential functions. Aida[7] estimated values for n-n stacking parameters from in vacuo ab initio molecular orbital calculations. There are barriers to obtaining a meaningful comparison of the reported n-n sets. First, the n-n sets were determined and presented in different statistical thermodynamic formalisms. In addition, the n-n sets were evaluated from analysis of melting experiments of different types of DNA samples in different solvent ionic environments. Attempts to clarify these issues and enable direct comparisons of the different n-n sets on an equal footing have been published,[8,14] but further efforts are needed.

DNA sequence-dependent hybridization is a primary component of many research and commercial reactions. Perhaps most notedly, the advent and wide use both in research and commercial environments of the polymerase chain reaction,[15] and the development and implementation of DNA chip technology,[16] have ignited newly heightened interests in accurate predictions of sequence-dependent melting stability of short duplex DNA oligomers. For many practical applications hybridization reactions result in formation of relatively short duplex regions. Stability of the hybridized products is determined by the sequence dependence of the transition temperatures t_m's of the duplex complexes relative to their unhybridized component strands. In order to optimize probe and primer sequence design,[17] it is of particular necessity to be able to accurately predict the t_m's of DNA oligomer duplexes under a given set of conditions (ionic strength and strand concentration) from their base pair sequence.

The study of melting short duplex DNA oligomers as a means to evaluate sequence-dependent thermodynamic stability of DNA was pioneered by Breslauer and Marky in the early 1980s.[9,18] In 1986 they reported the first set of n-n sequence-dependent stability parameters for DNA evaluated from melting curves of short DNA oligomers.[9] Recently, several new and putatively improved n-n sets, evaluated from melting experiments of short DNA duplex oligomers in 1.0*M* Na^+, have appeared.[10–12]

The studies described here were undertaken to address the following question. How well do n-n sequence-dependent stability parameters, evaluated by melting analysis of DNA dumbbells, polymers, and short duplex DNA oligomers, actually predict experimental t_m's of short duplex DNA oligomers? In order to compare the relative stabilities of short duplex DNA oligomers, we calculated the stability of all 10-mer sequences using 11 published sets of n-n parameters. We tested how well the n-n sets evaluated from melting studies of DNA dumbbells published over five years ago,[8] those evaluated from melting studies of short duplex DNA oligomers published over 10 years ago,[9] and those new and putatively improved sets recently published[10–12] compared in overall accuracy of predicting the melting stability of short duplex DNA oligomers from their sequences. We readily acknowledge that this test will naturally favor those n-n sets derived from studies that include the largest number of duplexes with lengths around 10 base pairs.

ANALYTICAL METHODS

Nearest-Neighbor Sequence Dependence

In the n-n model, sequence-dependent stability is considered in terms of n-n (two) base pair doublets.

In duplex DNA there are 10 such unique doublets. These are (5′-3′) AA = TT; AG = CT; AC = GT; GA = TC; GG = CC; TG = CA, CG, GC, AT, and TA. Whether thermodynamic quantities (enthalpies and entropies) for melting each of the 10 unique doublets can be uniquely determined from melting studies of an appropriately chosen set of molecules depends on assumptions about the ends. In principle, there are four possible types of sequence-dependent end interactions. As previously described,[19] representing an end as E, the possible n-n end interactions are (5′-3′) EA = TE, ET = AE, EG = CE, and EC = GE. If these end interactions are assumed to be zero, then for an appropriately chosen set of molecules, in which all 10 n-n sequences are adequately represented, 10 linearly independent equations exist that can be constructed. In order to solve for the required 10 unknowns (the n-n interactions), one must have 10 linearly independent equations. Because 10 linearly independent equations for the 10 unknowns exist, a unique solution for each of the 10 n-n base pair stacking interactions can be obtained (subject to the initial assumption regarding the ends). If the ends are not assumed to be zero, but instead assumed to be the same, with a single sequence-independent constant value, then 11 linearly independent equations can be written, and a unique solution can be obtained for the 10 n-n sequence interactions plus the end interaction. However, if the 4 additional n-n sequence-dependent end interactions are assumed not to be equivalent, or zero, there are 14 unknowns (the 10 n-n sequence interactions and the 4 end interactions) that should be evaluated. Unfortunately, for this situation there are only 12 linearly independent equations that can be written.[19] Consequently, a unique solution for all 14 unknowns cannot be obtained. Instead, 12 linearly independent combinations of the n-n sequence interactions, and the ends, can be solved for. Evaluations of these combinations can provide direct insight into the magnitudes of different n-n sequence combinations and sequence-dependent end interactions. The linearly independent linear combinations can be used directly to calculate sequence-dependent stability.[8,14] Details of the underlying analytical methods for treatment of the ends and evaluation of the n-n combinations have been published.[8,19–21] In published melting studies of short DNA oligomers the sequence-specific end interactions have been assumed to be zero or constant,[8–14] with some studies attempting to minimize the range of end effects by ''closing'' all oligomeric duplexes with G-C type base pairs.[9] In a later section the influence of the sequence dependent n-n end interactions on melting of duplex DNA oligomers is evaluated explicitly.

The Singlet and Doublet Formats

In the n-n approximation, sequence-dependent interactions in duplex DNA are considered to arise from two basic sources. These are Watson–Crick base pairing (H-bonding) and stacking interactions between n-n base pairs (stacking). Within the n-n model two types of computational methods have been employed to evaluate n-n parameters and predict DNA sequence-dependent thermodynamic stability. These are referred to as the singlet and doublet formats. In the singlet format, contributions from H-bonding and stacking interactions are considered separately.[1,4,8,14] In the doublet format the entire n-n interaction (H bonding and stacking) is considered in a single parameter.[2,3,5–7,9–12] Detailed descriptions of these formats and how they are equated have been provided.[8,14] Calculated results from the singlet and doublet formats are numerically equivalent so long as the appropriate correction factor for the end base pairs is employed.[14] In the comparative studies described here, the calculated stability of DNA oligomers was determined using 11 different n-n sets from the published literature. For 8 of these sets,[1–8] sequence-dependent stability was calculated using the singlet format and the parameters previously reported for these sets in the singlet format.[8] In the following, use of these parameters to calculate duplex thermodynamic stability is demonstrated.

Calculation of Duplex Stability in the Singlet Format

Although there are 10 unique n-n base pair doublets in duplex DNA, for circular polymers or semi-infinite polymers with essentially no ends there are only eight linearly independent combinations of n-n sequence-dependent free energies,[19–22] $\Delta G_{\text{n-n}}$. For dumbbells with the same type of end (same sequence and loop size), there are 9 linearly independent values, eight that correspond to those of the polymers and 1 that accounts for the explicit type of end.[8] The numbers and free energies for the eight linearly independent combinations that do not involve the ends are given in Table I.

In the singlet format, the average singlet energy (H bonding) of a base pair is given by $\Delta G_{\text{H bond}}$ values that can take on two values, depending on whether the base pair is of the A·T or G·C type.[1,8] The n-n sequence dependence is included as the

Table I Numbers and Free Energies for Eight Linearly Independent Linear Combinations of the Nearest-Neighbor Base Pair Sequences in DNA[a]

Numbers	Free Energies
$N_1 = N_{AA/TT}$	$\delta G_1 = \delta G_{AA/TT}$
$N_2 = N_{CC/GG}$	$\delta G_2 = \delta G_{CC/GG}$
$N_3 = N_{AT/AT} + N_{TA/TA}$	$\delta G_3 = (\delta G_{AT/AT} + \delta G_{TA/TA})/2$
$N_4 = N_{CG/CG} + N_{GC/GC}$	$\delta G_4 = (\delta G_{CG/CG} + \delta G_{GC/GC})/2$
$N_5 = N_{AC/GT} + N_{CA/TG}$	$\delta G_5 = (\delta G_{AC/GT} + \delta G_{CA/TG})/2$
$N_6 = N_{AG/CT} + N_{GA/TC}$	$\delta G_6 = (\delta G_{AG/CT} + \delta G_{GA/TC})/2$
$N_7 = N_{AT/AT} - N_{TA/TA} + N_{CG/CG} - N_{GC/GC} + 2(N_{GA/TC} - N_{AG/CT})$	$\delta G_7 = (\delta G_{AT/AT} - \delta G_{TA/TA} + \delta G_{CG/CG} - \delta G_{GC/GC})/12 + (\delta G_{GA/TC} - \delta G_{AG/CT})/6$
$N_8 = N_{AT/AT} - N_{TA/TA} - N_{CG/CG} + N_{GC/GC} + 2(N_{CA/TG} - N_{AC/GT})$	$\delta G_8 = (\delta G_{AT/AT} - \delta G_{TA/TA} - \delta G_{CG/CG} + \delta G_{GC/GC})/12 + (\delta G_{CA/TG} - \delta G_{AC/GT})/6$

[a] Numbers of the linear combinations shown on the left correspond to the linear combinations of the n-n sequence-dependent free energies shown on the right. These numbers and combinations were taken from Tables I and II of Goldstein and Benight.[19]

deviation from average stacking energy $\delta G_{n\text{-}n}$ for each type of n-n sequence.[1,8] These n-n sequence-dependent values are assumed to be entirely enthalpic and are summarized for eight n-n sets in Table II. The n-n sets are those determined from melting analysis of DNA dumbbells (Doktycz et al.,[8] column 1); melting studies of restriction fragments and polymers (Blake and Delcourt,[5] column 2; Wartell and Benight,[1] column 3; Vologodskii et al.,[3] column 4; Gotoh and Tagashira,[2] column 5; McCampbell et al.,[4] column 6) and theoretical calculations (Orstein and Fresco,[6] column 7; Aida,[7] column 8).

The values in Table II can be utilized to calculate the n-n sequence-dependent stability of any duplex DNA oligomer. In the example calculations that follow, base pairs on the ends are assumed to behave just as any other base pair and any n-n dependent end interactions are assumed to be zero. In this case the n-n sequence-dependent transition enthalpy of the duplex is written in terms of a hydrogen-bonding component $\Delta H_{H\ bond}$, whose magnitudes range with the %GC of the duplex, and a n-n interaction component, $\Delta H_{n\text{-}n}$. The duplex transition enthalpy is then determined according to

$$\Delta H_{duplex} = \Delta H_{H\ bond} + \Delta H_{n\text{-}n} = \Delta S_{bp}[N_{AT}T_{AT} + N_{GC}T_{GC}] + \sum_s N_s(\delta G_s) \quad (1)$$

Where N_{AT} and N_{GC} are the numbers of A·T or G·C type base pairs in the duplex sequence. The average melting temperatures of an A·T or G·C

Table II Values of the Nearest-Neighbor Sequence-Dependent Free-Energies for Eight Nearest-Neighbor Sets[a]

δG_i	1	2	3	4	5	6	7	8
δG_1	−190	−35	−54	−86	−99	−80	270	410
δG_2	146	155	178	184	300	−100	−30	1980
δG_3	−28	6	78	18	113	90	440	340
δG_4	−240	−153	−60	−154	−151	−75	−385	−2040
δG_5	−113	−49	−114	−46	−72	−30	−155	−365
δG_6	−6	44	97	108	57	0	−20	895
δG_7	−25	12	43	−23	−27	−18	199	−27
δG_8	−39	24	−94	−4	−6	18	−44	−117

[a] For eight n-n sets the sequence dependence is included as the deviation from average stacking energy $\delta G_{n\text{-}n}$, for each type of n-n sequence.[1,8] These n-n sequence-dependent values are assumed to be entirely enthalpic and are summarized for eight n-n sets. The n-n sets are as follows: column 1, Doktycz et al.[8]; column 2, Blake and Delcourt[5]; column 3, Wartell and Benight[1]; column 4, Vologodskii et al.[3]; column 5, Gotoh and Tagashira[2]; column 6, McCampbell et al.[4]; column 7, Orstein and Fresco[6]; column 8, Aida.[7] Tabulated values correspond to the linear combinations given on the right-hand side of Table I.

base pair are given by T_{AT} or T_{GC}. These are readily calculated as a function [Na^+] using the following equations derived from Frank-Kamenetskii's relationships.[23]

$$T_{AT} = 355.55 + 7.95 \ln[Na^+] \quad (2a)$$

$$T_{GC} = 391.55 + 4.89 \ln[Na^+] \quad (2b)$$

The summed term on the right in Eq. (1) includes the explicit n-n sequence dependence. N_s is the number of times linear combination s ($s = 1-8$) occurs in the duplex sequence, and δG_s is the linear combination of deviations in the free energy from average stacking for sequence combination s.

The entropy change of base pair melting ΔS_{bp} is assumed to be independent of sequence and [Na^+] over the range from 0.02 to 1.0*M* Na^+, with an average value of $\Delta S_{bp} = (-24.85\ cal \cdot mol^{-1} \cdot K^{-1})$.[5] This assumption of a constant ΔS_{bp} value is not fully supported by some published data on oligomeric duplexes,[9,18] but nevertheless serves the purposes of the analyses presented here. With this assumption, the total transition entropy of the duplex is simply

$$\Delta S_{duplex} = \Delta S_{bp}[N_{AT} + N_{GC}] \quad (3)$$

The duplex melting transition free energy $\Delta G_{duplex}(T) = \Delta H_{duplex} - T\Delta S_{duplex}$. With this relationship and Eqs. (1) and (3),

$$\Delta G_{duplex}(T) = \Delta S_{bp}[N_{AT}(T_{AT} - T) + N_{GC}(T_{GC} - T)] + \sum_s N_s(\delta G_s) \quad (4)$$

For example, consider melting the 10-mer duplex sequence 5′-A-T-T-A-T-G-G-G-G-C-3′ in 115 m*M* Na^+. We wish to calculate the free energy of melting ΔG_{duplex} at 25°C. From Eqs. (2a) and (2b) in 0.115*M* Na^+, $T_{AT} = 338.36$ K (65.21°C) and $T_{GC} = 380.97$ (107.82°C). There are five A·T and five G·C base pairs, $N_{AT} = N_{GC} = 5$. The first term on the right of Eq. (4) includes the H-bond free energies and is evaluated as

$$\begin{aligned}\Delta S_{bp}[N_{AT}(T_{AT} - T) + N_{GC}(T_{GC} - T)] &= -24.85[5(338.36 - 298.15) \\ &\quad + 5(380.97 - 298.15)] \\ &= -15{,}286 \text{ cal/mol}\end{aligned}$$

To determine the n-n sequence-dependent contributions to the stability, we must assess the numbers of the 8 linear combinations, as given in Table I, that are present in the given duplex sequence. For the 10-mer sequence given, the numbers of these combinations are as follows: $N_1 = N_{AA/TT} = 1$; $N_2 = N_{CC/GG} = 3$; $N_3 = N_{AT/AT} + N_{TA/TA} = 2 + 1 = 3$; $N_4 = N_{CG/CG} + N_{GC/GC} = 0 + 1 = 1$; $N_5 = N_{AC/GT} + N_{CA/TG} = 0 + 1 = 1$; $N_6 = N_{AG/CT} + N_{GA/TC} = 0 + 0 = 0$; $N_7 = N_{AT/AT} - N_{TA/TA} + N_{CG/CG} - N_{GC/GC} + 2(N_{GA/TC} - N_{AG/CT}) = 2 - 1 + 0 - 1 + 2(0 - 0) = 0$; $N_8 = N_{AT/AT} - N_{TA/TA} - N_{CG/CG} + N_{GC/GC} + 2(N_{CA/TG} - N_{AC/GT}) = 2 - 1 - 0 + 1 + 2(1 - 0) = 4$. With these resident combinations and their corresponding values in Table I, the second term in Eq. (4) is

$$\begin{aligned}\sum_s N_s(\delta G_s) &= 1(\delta G_1) + 3(\delta G_2) + 3(\delta G_3) + 1(\delta G_4) \\ &\quad + 1(\delta G_5) + 0(\delta G_6) + 0(\delta G_7) + 4(\delta G_8) \\ &= -190 + 3(146) + 3(-28) - 240 - 113 \\ &\quad + 4(-39) = -345 \text{ cal/mol}\end{aligned}$$

Thus, from the n-n set in column 1 of Table II, the predicted value of $\Delta G_{duplex}(25°C) = -15{,}286 - 345 = -15{,}631$ cal/mol. The calculated free energy for this sequence can be determined in an analogous manner using the other n-n sets in columns 2–8 of Table II.

Calculation of Duplex Stability in the Doublet Format

In this format, the contributions of H bonding and n-n stacking are combined in a single parameter for each type of n-n base pair doublet. Values of the enthalpies and entropies for the 10 n-n doublets were evaluated by Breslauer et al.,[9] SantaLucia et al.,[10] Allawi and SantaLucia,[11] and Sugimoto et al.,[12] and reported in the doublet format. As stated above, in their evaluations these authors assumed that n-n interactions with the ends were either zero or a sequence independent constant. The n-n set reported by Breslauer et al.[9] was determined from melting studies of short DNA oligomers and repeating DNA copolymers. The sets reported by SantaLucia et al.,[10] Allawi and SantaLucia,[11] and Sugimoto et al.[12] were determined from melting analysis of short DNA oligomers. The reported n-n doublet enthalpies ΔH_{ij} and entropies ΔS_{ij}, where ij represents one of the 10 unique n-n doublet sequences, are listed in Table III for each n-n set. Utilizing these values the calculated duplex melting enthalpy is just the sum of the individual enthalpies of the constituent n-n doublets of the duplex. That is,

Table III Values of the Nearest-Neighbor Sequence Dependent Thermodynamic Parameters[a]

	A		B		C		D	
Sequence (*ij*)	ΔH_{ij}	ΔS_{ij}	ΔH_{ij}	ΔS_{ij}	ΔH_{ij}	ΔS_{ij}	ΔH_{ij}	ΔS_{ij}
AA/TT	−9.1	−24.0	−8.4	−23.6	−7.9	−22.2	−8.0	−21.9
AG/CT	−7.8	−20.8	−6.1	−16.1	−7.8	−21.0	−6.6	−16.4
AT/AT	−8.6	−23.9	−6.5	−18.8	−7.2	−20.4	−5.6	−15.2
AC/GT	−6.5	−17.3	−8.6	−23.0	−8.4	−22.4	−9.4	−25.5
GA/TC	−5.6	−13.5	−7.7	−20.3	−8.2	−22.2	−8.8	−23.5
GG/CC	−11.0	−26.6	−6.7	−15.6	−8.0	−19.9	−10.9	−28.4
GC/GC	−11.1	−26.7	−11.1	−28.4	−9.8	−24.4	−10.5	−26.4
TA/TA	−6.0	−16.9	−6.3	−18.5	−7.2	−21.3	−6.6	−18.4
TG/CA	−5.8	−12.9	−7.4	−19.3	−8.5	−22.7	−8.2	−21.0
CG/CG	−11.9	−27.8	−10.1	−25.5	−10.6	−27.2	−11.8	−29.0
Initiation if at least one G·C	0.0	−16.8	0.0	−5.9	—	—	0.6	−9.0
Initiation if only A·T	0.0	−20.1	0.0	−9.0	—	—	0.6	−9.0
Symmetry correction	0.0	−1.3	0.0	−1.4	0.0	−1.4	0.0	−1.4
5′ T·A correction	—	—	0.4	0.0	—	—	—	—
A·T base pair on the end	—	—	—	—	2.3	4.1	—	—
G·C base pair on the end	—	—	—	—	0.1	−2.8	—	—

[a] The reported enthalpies ΔH_{ij} and entropies ΔS_{ij}, for each type of unique n-n base pair doublet in duplex DNA (*ij*) are summarized as reported for the four n-n sets as follows. Column A: Breslauer et al.[9]; column B: SantaLucia et al.[10]; column C: Allawi and SantaLucia[11]; column D: Sugimoto et al.[12] In addition to values for the unique n-n sequence doublets, the symmetry corrections and nucleation parameters reported by these authors are entered in the lower half of the table.

$$\Delta H_{\text{duplex}} = \sum_{ij} N_{ij}\Delta H_{ij} \tag{5}$$

where N_{ij} is the number of times the particular n-n doublet *ij* (*i*, *j* = A, T, G, C) appears in the duplex. The duplex melting entropy is determined in an analogous manner:

$$\Delta S_{\text{duplex}} = \sum_{ij} N_{ij}\Delta S_{ij} \tag{6}$$

Finally, the duplex melting transition free energy is given by

$$\Delta G_{\text{duplex}}(T) = \sum_{ij} N_{ij}(\Delta H_{ij} - T\Delta S_{ij}) \tag{7}$$

For example, again consider the 10-mer sequence 5′-A-T-T-A-T-G-G-G-G-C-3′ and calculate the duplex melting free energy at 25°C. The numbers of different types of n-n doublets that occur in the sequence are $N_{\text{AA/TT}} = 1$, $N_{\text{AG/CT}} = 0$, $N_{\text{AT/AT}} = 2$, $N_{\text{AC/GT}} = 0$, $N_{\text{GA/TC}} = 0$, $N_{\text{GG/CC}} = 3$, $N_{\text{GC/GC}} = 1$, $N_{\text{TA/TA}} = 1$, $N_{\text{TG/CA}} = 1$, and $N_{\text{CG/CG}} = 0$. With these values, and the n-n parameters reported by Breslauer et al.[9] (column A of Table III) in Eq. (7), $\Delta G_{\text{duplex}}(25°\text{C})$ is found accordingly:

$$\begin{aligned}
\sum_{ij} N_{ij}(\Delta H_{ij} - T\Delta S_{ij}) &= 1(\Delta H_{\text{AA/TT}} - T\Delta S_{\text{AA/TT}}) \\
&\quad + 2(\Delta H_{\text{AT/AT}} - T\Delta S_{\text{AT/AT}}) \\
&\quad + 3(\Delta H_{\text{GG/CC}} - T\Delta S_{\text{GG/CC}}) \\
&\quad + 1(\Delta H_{\text{GC/GC}} - T\Delta S_{\text{GC/GC}}) \\
&\quad + 1(\Delta H_{\text{TA/TA}} - T\Delta S_{\text{TA/TA}}) \\
&\quad + 1(\Delta H_{\text{TG/CA}} - T\Delta S_{\text{TG/CA}}) \\
&= (-9100 - 298.15(-24.0)) \\
&\quad + 2(-8600 - 298.15(-23.9)) \\
&\quad + 3(-11{,}000 - 298.15(-26.6)) \\
&\quad + (-11{,}100 - 298.15(-26.7)) \\
&\quad + (-6000 - 298.15(-16.9)) \\
&\quad + (-5800 - 298.15(-12.9)) \\
&= -1944 - 2948 - 9208 - 3139 - 961 \\
&\quad - 1954 = -20{,}154 \text{ cal/mol}
\end{aligned}$$

Compared to the value predicted using the singlet format and the n-n parameters evaluated from DNA

dumbbells,[8] the Breslauer et al. n-n set[9] predicts the specific sequence considered to be more stable by nearly −5 kcal/mol, in the direction of enhanced stability that one would expect for the higher salt concentration used in the studies of Breslauer et al.[9] Clearly, some of this discrepancy can be understood in terms of the higher $[Na^+]$ (1.0*M* Na^+) where the Breslauer et al.[9] n-n set was evaluated, and where DNA is more stable and would be expected to have a lower free energy. An additional source of the discrepancy could be the significant weighting of polymeric duplex results in obtaining the Breslauer et al.[9] n-n set. Additional details of the predictive accuracies of these n-n sets and others are presented later.

It should be mentioned at this point that in order to calculate the total stability of short DNA oligomers that can then be compared directly to experimentally measured values, it is necessary to consider in addition to the n-n sequence-dependent duplex free energy given above, the free energy of helix nucleation. The nucleation free energy accounts for energetically unfavorable interactions between complementary single strands that must be overcome to allow duplex formation. Slightly different values for nucleation parameters have been reported and employed by authors of the different n-n sets. These are given in the bottom rows of Table III. As described below, in the relative statistical comparison of calculated stabilities of 10-mers that was performed, the nucleation free energies were not included in the calculated values, $\Delta G(25°C) = \Delta G_{duplex}(25°C)$. In addition, the entropy of symmetry correction for self-complementary duplexes ($-1.4\ cal \cdot mol^{-1} \cdot K^{-1}$) is smaller than errors often found for other parameters and was omitted from the comparative free-energy calculations.

Method of the Statistical Comparisons

Predictions from 11 different n-n sets (the eight sets in Table II) and the sets of Breslauer et al.,[9] SantaLucia et al.,[10] and Sugimoto et al.[12] (in Table III) were first compared using a relative statistical approach. The n-n set of Allawi and SantaLucia[11] was not employed in this comparison because it was derived from the data reported by SantaLucia et al.[10] and Sugimoto et al.[12] Using each n-n set, the free energy $\Delta G(25°C)$ for each possible 10 base pair duplex sequence was calculated. Different n-n sets were then compared through comparison of the calculated $\Delta G(25°C)$ values for all 10-mers. There are 4^{10} different 10 base single-strand sequences. A small fraction of these sequences is self-complementary in that two strands of the same sequence can form a duplex. Sequences were considered self-complementary when the first five bases from the 5′ end were complementary to the first five bases from the 3′ end. The number of such self-complementary oligomers for 10 base pair sequences is 4^5. The remaining $4^{10}-4^5$ nonself-complementary single strands were assumed to combine with their complement strand to form $(4^{10}-4^5)/2$ duplexes, where each duplex is comprised of two *different* single strands. Consequently, the total number of unique 10 base pair DNA duplex sequences is $(4^{10}-4^5)/2 + 4^5 = 524{,}800$. Free energies of these 10-mer sequences were calculated using each n-n set. Comparison of results for all sequences obtained with two different n-n sets identified particular sequences whose calculated stabilities were different for the two n-n sets.

The following were steps in the comparison process: (1) The $\Delta G(25°C)$ values for the 524,800 possible unique 10 base pair duplex sequences were calculated using each of the 11 n-n sets. (2) For each n-n set, differences in $\Delta G(25°C)$, for every possible pair of sequences were determined. For the 524,800 possible 10 base pair sequences, there are $(524{,}800)(524{,}799)/2 = 1.377 \times 10^{11}$ such pairs of sequences. (3) Within each n-n set, these pairwise differences in $\Delta G(25°C)$ were compared. The difference between sequence *i* and sequence *j* was defined as $\Delta\Delta G(i\text{-}j) = \Delta G(i) - \Delta G(j)$. If ϵ is the relative error in the free energy calculated using n-n set *X*, then by definition, if $\Delta\Delta G(i\text{-}j) < 0$ and $|\Delta\Delta G(i\text{-}j)| > |(\epsilon)(\Delta G(i) + \Delta G(j))/2|$ sequence *i* is *more stable* than sequence *j*. Likewise, if $|\Delta\Delta G(i\text{-}j)| \leq |(\epsilon)(\Delta G(i) + \Delta G(j))/2|$ sequence *i* has *stability equal* to sequence *j*. This means the calculated stability of sequences *i* and *j* is the same within the error. If $\Delta\Delta G(i\text{-}j) > |(\epsilon)(\Delta G(i) + \Delta G(j))/2|$ sequence *i* is *less stable* than sequence *j*. For each n-n set *X* ($X = 1-11$) $\Delta\Delta G(i\text{-}j)$ values for every possible sequence pair were tabulated. (4) Then for every possible combination of two n-n sets, calculated $\Delta\Delta G(i\text{-}j)$ values for every possible sequence pair were compared. Sequence pairs for which the two n-n sets predicted the opposite order of stability were denoted *discordant pairs.* For each pair of n-n sets compared the number of discordant pairs were determined and tabulated. We chose to compare $\Delta\Delta G(i\text{-}j)$ values because within a given n-n set, these values are probably less sensitive to the different reference duplex and single-strand states that could occur in the different molecular and sodium ion environments where the different

n-n sets were determined. Thus, within a given n-n set relative differences were considered to be more suitable for direct comparisons with other n-n sets.

Two-State Melting Theory

Historically, melting curves of short duplex DNA oligomers have been analyzed assuming an all-or-none or two-state melting transition.[24] Following is a description of the two-state model. For convenience, instead of melting, consider the reversible equilibrium annealing reaction of two DNA single strands, S_1 and S_2, to form a duplex D with equilibrium constant K_D.

$$S_1 + S_2 \overset{K_D}{\rightleftarrows} D \tag{8}$$

K_D can be expressed in terms of the ratios of the statistical weights of the internal and external degrees of freedom of the duplex and single strands, viz.[25,26]

$$K_D = [Z_{\text{int}}(D)Z_{\text{ext}}(D)]/ [Z_{\text{int}}(S_1)Z_{\text{int}}(S_2)Z_{\text{ext}}(S_1)Z_{\text{ext}}(S_2)] \tag{9}$$

The statistical weight ratios for the internal degrees of freedom account for the thermodynamic differences between the duplex $Z_{\text{int}}(D)$ and single strands $Z_{\text{int}}(S_1)$ and $Z_{\text{int}}(S_2)$ for the concentration-independent part of duplex formation. Sequence-dependent hydrogen bonding and stacking interactions in the duplex comprise this term. This statistical weight ratio is represented as

$$Z_{\text{int}}(D)/[Z_{\text{int}}(S_1)Z_{\text{int}}(S_2)] = K_{\text{duplex}} \tag{10}$$

Where K_{duplex} is the equilibrium constant for changes in the internal degrees of freedom required for establishment of the forces (H bonding and stacking) in duplex formation.

$$\Delta H_{\text{duplex}} - T\Delta S_{\text{duplex}} = -RT \ln K_{\text{duplex}} \tag{11}$$

The quantities ΔH_{duplex} and ΔS_{duplex} are the duplex melting transition enthalpy and entropy, respectively, for the sequence. These thermodynamic quantities are calculated directly from the base pair sequence using any of the available n-n stability values[1–12] and are assumed to be temperature independent.

The statistical weight ratio for the external degrees of freedom contains the concentration dependence and associated factors of the external degrees of freedom involved in duplex formation.[26,27] The definition is made,

$$Z_{\text{ext}}(D)/[Z_{\text{ext}}(S_1)Z_{\text{ext}}(S_2)] \equiv \beta \tag{12}$$

where β is the nucleation parameter. From Eqs. (10) and (12),

$$K_D = K_{\text{duplex}}\beta \tag{13}$$

The total concentration of strands C_T is given by $C_T = [S_1] + [S_2] + 2[D]$. At any given temperature, the net fraction of intact base pairs θ_{net} can be expressed in terms of the external and internal degrees of freedom of the duplex and single strands,[25,26] i.e.,

$$\theta_{\text{net}} = \theta_{\text{ext}}\theta_{\text{int}} \tag{14}$$

Where θ_{ext} is the fraction of strands with at least one intact base pair and θ_{int} is the fraction of intact base pairs on duplex strands having at least one base pair. If the melting transition is two state, and the fully intact duplex and completely dissociated single strands are the only states that are populated by every strand throughout the entire melting transition, then $\theta_{\text{int}} = 1$ and $\theta_{\text{net}} = \theta_{\text{ext}}$. Depending on whether S_1 and S_2 are self-complementary, two slightly different expressions for θ_{ext} in terms of K_D and C_T can be derived.

For the case where the strands are different ($S_1 \neq S_2$),

$$\theta_{\text{ext}} = [1 + C_TK_D - (1 + 2C_TK_D)^{0.5}]/C_TK_D \tag{15a}$$

When the two strands of the duplex are the same ($S_1 = S_2$), C_T is replaced by $4C_T$,

$$\theta_{\text{ext}} = [1 + 4C_TK_D - (1 + 8C_TK_D)^{0.5}]/4C_TK_D \tag{15b}$$

The transition temperature $T = T_m$ is defined as the temperature where $\theta_{\text{ext}} = 0.5$. Consequently, at $T = T_m$, $C_TK_D = \alpha$. Where $\alpha = 4$ when ($S_1 \neq S_2$), and $\alpha = 1$ when ($S_1 = S_2$).

With the above relationships, at $T = T_m$,

$$K_{\text{duplex}}(T = T_m) = K_D/\beta = (\alpha/C_T)\beta \tag{16}$$

Since $-RT_m \ln K_{\text{duplex}} = \Delta G_{\text{duplex}} = \Delta H_{\text{duplex}} - T_m\Delta S_{\text{duplex}}$,

$$\Delta H_{\text{duplex}} - T_m \Delta S_{\text{duplex}} = RT_m \ln(C_T/\alpha) + RT_m \ln \beta \quad (17)$$

The last term on the right of Eq. (17) is the nucleation free energy, $\Delta G_{\text{nuc}} = -RT_m \ln \beta = \Delta H_{\text{nuc}} - T_m \Delta S_{\text{nuc}}$. With these definitions, an expression for $1/T_m$ vs $\ln(C_T/\alpha)$ is obtained, viz.,

$$1/T_m = (R/(\Delta H_{\text{duplex}} + \Delta H_{\text{nuc}}))\ln(C_T/\alpha) + (\Delta S_{\text{duplex}} + \Delta S_{\text{nuc}})/(\Delta H_{\text{duplex}} + \Delta H_{\text{nuc}}) \quad (18a)$$

This expression is usually written in the more compact familiar form as

$$1/T_m = (R/\Delta H_T)\ln(C_T/\alpha) + \Delta S_T/\Delta H_T \quad (18b)$$

Where, obviously, $\Delta H_T = (\Delta H_{\text{duplex}} + \Delta H_{\text{nuc}})$ and $\Delta S_T = (\Delta S_{\text{duplex}} + \Delta S_{\text{nuc}})$. Equation (18b) is the familiar van't Hoff expression usually employed to evaluate ΔH_T and ΔS_T from plots of $1/T_m$ as a function of $\ln(C_T/\alpha)$. Rearranging Eq. (18a) yields a general expression for T_m in terms of the respective thermodynamic quantities and total strand concentration,

$$T_m = (\Delta H_{\text{duplex}} + \Delta H_{\text{nuc}})/(R \ln(C_T/\alpha) + \Delta S_{\text{duplex}} + \Delta S_{\text{nuc}}) \quad (19)$$

The expression in Eq. (19) was employed to calculate T_m's of duplex sequences as a function of strand concentration. These were compared directly to experimental measurements. Calculations required input values for ΔH_{duplex}, ΔH_{nuc}, ΔS_{duplex}, ΔS_{nuc}, and C_T. As demonstrated earlier, it is straightforward to employ the n-n stability parameters to calculate ΔH_{duplex} and ΔS_{duplex} from the base pair sequence. But the nucleation parameters have been reported in several different forms. If ΔH_{nuc} is assumed to be zero, then ΔS_{nuc} is determined from the nucleation free energy, i.e., $\Delta S_{\text{nuc}} = -\Delta G_{\text{nuc}}/T = -R \ln \beta$. Recently, nonzero values of both ΔS_{nuc} and ΔH_{nuc} have been reported.[10–12] In our analysis several different forms of ΔS_{nuc} and $\Delta H_{\text{nuc}} \neq 0$ were assumed and tested for their ability to improve average predictions of T_m's of short duplex DNA oligomers.

EXPERIMENTAL METHODS

Melting Data of Short Duplex DNA Oligomers

The database used in the analysis was comprised of melting curve data collected from the published literature and acquired in our laboratory. Results of melting analysis of 131 unique duplex DNA oligomers ranging in length from 4 to 16 base pairs collected in 1.0*M* Na^+ were recently reported by Allawi and SantaLucia.[11] Although they reported results for 131 DNAs, only 108 of these were used in their evaluation of n-n parameters. Close examination of their list of sequences (supplied as supplementary information) reveals one of the 108 sequences, the 8-mer, 5′-CGATATCG-3′, was included twice. The additional 23 molecules for which melting data were provided were reported to exhibit marginal two-state or non-two-state melting behavior. Three of these sequences, 5′-GAAGCTTC-3′, 5′-GGAATTCC-3′, and 5′-CGCGAATTCGCG-3′, were reported to display both two-state and non-two-state behavior, and their melting data were reported twice. In addition, the data reported in their paper for the sequence 5′-CAACCAACCAAC-3′ is not that given in the cited literature reference.[28] Consequently, Allawi and SantaLucia[11] used 107 unique duplex molecules in their n-n parameter evaluations. In our analysis we utilized the melting data for the 119 molecules reported by Allawi and SantaLucia[11] to melt in a two-state or marginally two-state manner. From this list, redundant data for the three duplicate sequences given above were averaged, and data for the sequences 5′-CGCGAATTCGCG-3′ and 5′-CAACCAACCAAC-3′ were removed from consideration. This left melting data for a total of 114 unique duplex sequences supplied by Allawi and SantaLucia[11] that were used in our analysis. Additional melting data for another 136 eight base pair DNAs, collected in 1.0*M* NaCl, were published by Doktycz and co-workers.[29] Although they claimed to report data for 140 molecules, for 4 of the molecules, 5′-GCATGGAC-3′, 5′-GCCTGGAC-3′, 5′-GCGTGGAC-3′, and 5′-GCTTGGAC-3′, data were reported twice. Thus, the total number of unique sequences actually reported was 136. In addition, melting data for the duplex sequence 5′-ACAAGCTTGCATGCCT-3′ was acquired from the published literature.[30] When taken together, our set of collected results from the published literature contained melting data measured in 1.0*M* Na^+ for 251 different duplex DNA sequences ranging in length from 4 to 16 base pairs. This data set was used to test the predictive accuracy of different published n-n parameters. We also employed this database to evaluate the influence of different forms of the nucleation free energy on overall accuracy of predicted stabilities of short duplex DNA oligomers.

A second set of melting data obtained for another 76 DNA duplexes ranging in length from 6 to 24 base pairs was measured in our laboratory and collected from additional literature sources.[10,14,28,31–49] Sequences of these molecules are shown in the Appendix Table A1. Melting data for these molecules were collected in 1.0*M* and/or 0.115*M* Na^+. That is, not all sequences were melted in both ionic strength environments. Of the 76 molecules comprising this second set, 48 were prepared, melted, and analyzed in our laboratory. For 16 of these molecules,

ranging in length from 10 to 24 base pairs, and varying %GC from 10 to 65%, melting data were collected in both Na^+ environments. These data were used to determine an empirical correction factor for differences in t_m observed in 1.0 and 0.115*M* Na^+ that also depends on the %GC of the sequence.

DNA Synthesis and Purification

For the 48 short DNA duplex oligomers that were prepared in our laboratory, the required single strands were synthesized on an Applied Biosystems 380B synthesizer using the standard β-cyanoethyl phosphoramidite method.[50] Strands were cleaved from solid support by incubation overnight in concentrated ammonium hydroxide at 58°C. DNA solutions were vacuum dried, rehydrated in double distilled water (ddH_2O), and exhaustively dialyzed vs ddH_2O. Strands were purified on 15% polyacrylamide gels at 60°C. Gels were briefly shadowed with uv light, while the major band was sliced out. DNA was electroluted out of gel slices in TBE buffer (50 m*M* Tris HCl, 50 m*M* boric acid, 1 m*M* Na_2EDTA, pH 8.1). Gel debris was removed by filtration through a 0.2 μm filter. Purified DNA samples were dialyzed exhaustively vs ddH_2O, then melting buffer (100 m*M* or 1*M* NaCl, 5 m*M* Na_2HPO_4, 5 m*M* NaH_2PO_4, 1 m*M* Na_2EDTA, pH 6.8).

UV Melting Experiments

Forty of the duplexes for which melting data were acquired in our laboratory were prepared and melted for other studies. Results for some of these studies have been published.[14] Methods for collecting and analyzing melting data for these molecules are essentially those described.[14] For the remaining eight 10-mers a series of melting experiments was performed expressly for the present study. These molecules were melted in 115 m*M* and 1.0*M* Na^+ at different strand concentrations. Details of these experiments are explicitly described next.

Absorbance vs temperature curves were collected using a single-beam Hewlett-Packard 8452 spectrophotometer, kept in a cold room maintained at 10°C. This cooler surrounding ambient temperature limited condensation on cuvette surfaces at low temperatures. Cuvettes with path lengths of 0.1 and 1.0 cm were used. Sample temperature was monitored as the sample holder temperature with an HP 89090A Peltier Temperature Controller. An external probe inserted directly in the sample cuvette gave reproducible temperature readings that were within 0.1°C of the sample holder. Strand concentrations were determined from absorbance readings at 260 nm, 25°C, using calculated extinction coefficients determined by the nearest-neighbor method.[51] Mixing curves of the strands required to make each 10-mer duplex confirmed duplex formation occurred at a 1:1 molar ratio mixture of the strands. Prior to melting experiments, 10-mer samples were dialyzed vs 1.0*M* NaCl or 100 m*M* NaCl in melting buffer.

Samples were prepared for melting experiments by dilution with the appropriate melting buffer and degassing by bubbling with a fine stream of helium, then tightly sealed with Teflon stoppers. Samples were heated at a rate of 35°C/h over the temperature range from 7.2 to 85.1°C and cooled at the same rate while the absorbance at 268 nm was measured every 0.1°C. Duplex oligomers were melted over the 120-fold range in DNA concentration from 1 to 120 μM. At least three heating and cooling absorbance vs temperature curves were collected for every 10-mer sample at each concentration. Raw absorbance vs temperature curves were smoothed,[52] normalized to upper and lower sloping baselines, and converted to θ_B vs T curves, where θ_B is the fraction of melted duplexes. At the melting temperature t_m, the fraction of melted duplexes $\theta_B = 0.5$. For each duplex oligomer, t_m's at a given concentration were reproducible within 0.4°C. For all samples, heating and cooling curves overlapped, indicating the melting transitions were at equilibrium and reversible. Evaluations of the transition thermodynamic parameters ΔH_T and ΔS_T were made from linear fits of plots of $1/T_m$ vs $\ln(C_T/\alpha)$. The variance of $1/T_m$ values were estimated from errors in fitting the slopes of these plots,

$$\sigma^2(1/T_m) = 1/(N-2)\sum_{i=1}^{N}[y(i) - a - b\cdot x(i)]^2 \quad (20)$$

where $y(i) = (1/T_m)(i)$ and $x(i) = (\ln(C_T/\alpha))(i)$ and a and b are the fitted y intercept and slope, respectively. From this parameter the variance and covariance of the slope and y intercept were evaluated.[53,54] These were in turn used to estimate errors on the graphically determined thermodynamic parameters[55,56] ΔH and ΔS. This analysis assumes that plots of $1/T_m$ vs $\ln(C_T/\alpha)$ are absolutely linear, and that the enthalpy and entropy are temperature independent, and in addition, that the change in heat capacity at constant pressure for the melting transition $\Delta C_p = 0$. As Chaires has shown,[57] slight curvature in van't Hoff plots due to $\Delta C_p \neq 0$ can lead to large errors in the graphically evaluated thermodynamic parameters.

RESULTS AND DISCUSSION

Energy Distributions of 10-mers Predicted From 11 Nearest-Neighbor Sets

Calculated free-energy distributions determined at 25°C, ΔG(25°C), for all possible 10-mer duplex sequences are shown in Figure 1. These energy distributions were calculated using the 11 sets of published n-n stability parameters given in Tables II and III. Distributions shown in Figure 1 were constructed by dividing the range between minimum and maximum ΔG(25°C) values into $(524{,}800)^{0.5} = 724$ intervals. The fraction of sequences with calculated ΔG(25°C) that fall within each interval are plotted vs calculated ΔG(25°C).

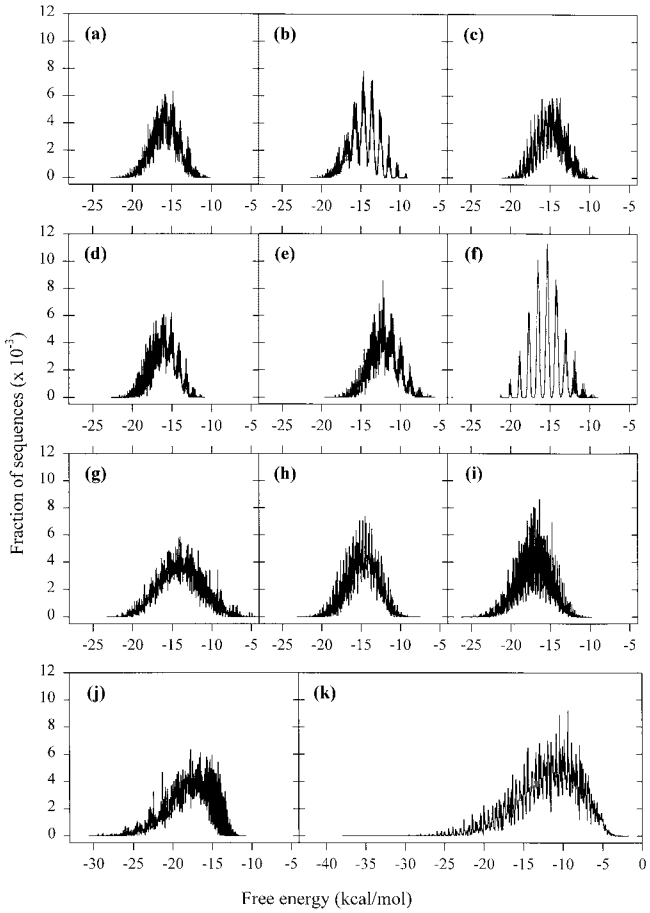

FIGURE 1 Calculated free-energy distributions for all 10 base pair duplex DNAs. Fraction of sequences is plotted vs the calculated free energy at 25°C, ΔG(25°C) for all nonself-complementary sequences 10 base pairs long. As described in the text, these calculations considered only the sequence-dependent base pair energetics and did not include the nucleation free energy. Distributions shown were obtained from the following n-n sets: (a) Doktycz et al.,[8] (b) Delcourt and Blake,[5] (c) Wartell and Benight,[1] (d) Vologodskii et al.,[3] (e) Gotoh and Tagashira,[2] (f) McCampbell et al.,[4] (g) Orstein and Fresco,[6] (h) SantaLucia et al.,[10] (i) Sugimoto et al.,[12] (j) Breslauer et al.,[9] and (k) Aida.[11]

Comparison of the distributions in Figure 1 reveals a number of interesting features of the different n-n sets. For example, the n-n sets reported by Delcourt and Blake[5] (Figure 1b) and McCampbell et al.[4] (Figure 1f) display polymodal distributions with nearly discrete peaks. These peaks correspond to values of ΔG(25°C) for the 10-mers obtained considering only the hydrogen bonding component of base pair stability. Spikes occur at each increment of increasing $\%G \cdot C$ base pairs in the sequences. That distinct peaks are observed for these n-n sets reveals that the majority of base pair stability arises from the average hydrogen-bonding free energy (determined by $\%G \cdot C$), and deviations from the hydrogen bonding free energy due to n-n sequence-dependent stacking interactions are relatively small. Distributions from other n-n sets such as those reported by Breslauer et al.[9] (Figure 1j) and the theoretical set of Aida[7] (Figure 1k) are nonsymmetric and skewed to higher ΔG(25°C), revealing for these n-n sets that n-n interactions tend to destabilize a substantial number of the sequences and/or the significant impact of melting data on polymeric duplexes used in the evaluation of the n-n set reported by Breslauer et al.[9]

Comparisons of Discordant Pairs

There are 137.7×10^9 different pairs of 10 base pair duplexes. In the n-n model a small number of

these pairs cannot be discordant because they are comprised of precisely the same number of n-n base pair doublets, and therefore have the same predicted stability. Obviously, any pair of such sequences can never be discordant. The following sequences (5′-3′) are a few examples: ACCGTCAAGA, AAGGACGTCA, AGGACGTTGA, AGTGAACGGA, ACCGTCTCAA, GAAGACACGG, GAGTTGACGG, GTGTTCGAGG, GGTTCGACAG, GGTGAACTCG. Neglecting the end interactions, the number of different types of n-n doublets that occur in each sequence are as follows: $N_{AT/AT} = 0$, $N_{TA/TA} = 0$, $N_{AA/TT} = 1$, $N_{AC/GT} = 2$, $N_{CA/TG} = 1$, $N_{TC/GA} = 2$, $N_{CT/AG} = 1$, $N_{CG/CG} = 1$, $N_{GC/GC} = 0$, $N_{GG/CC} = 1$. Therefore, the predicted ΔG(25°C) for these sequences using any given n-n set is the same. Because they all have the same predicted ΔG(25°C), no pair of them can ever be discordant. Of 137.7 $\times 10^9$ possible pairs of 10-mer sequences, only a small fraction of these (62,255,048 pairs) have exactly the same number and types of n-n doublets.

The number of discordant pairs [assuming an error in $\Delta\Delta G$(25°C) of 5%], for all pairwise combinations of the 11 n-n sets in Tables II and III, are given in Table IV. This comparison reveals the sets reported by Doktycz et al.,[8] Vologodskii et al.,[3] and Gotoh and Tagashira[2] are identical within the assumed error. Relatively low numbers of discordant pairs are also found between these sets and those reported by Delcourt and Blake,[5] Wartell and Benight,[1] and McCampbell et al.[4] Relatively larger numbers of discordant pairs are found for comparisons of the experimentally evaluated n-n sets[1–5,8–12] with the theoretically calculated n-n sets of Ornstein and Fresco[6] and Aida.[7] Also, relatively larger numbers of discordant pairs are found between the n-n sets reported by Breslauer et al.,[9] SantaLucia et al.,[10] and Sugimoto et al.[12] For example, at the assumed 5% relative error (ϵ), comparison between the n-n set of Doktycz et al.[8] and that of Breslauer et al.[9] reveals 3.7 $\times 10^9$ discordant pairs. There are 2.2 $\times 10^8$ discordant pairs between the sets of Doktycz et al.[8] and SantaLucia et al.[10] In absolute terms this is a significant number of sequences. However in relative terms, of the 1.377 $\times 10^{11}$ possible pairs, 3.7 $\times 10^9$ discordant pairs is only 2.7% of the total, and 2.2 $\times 10^8$ discordant pairs is only 0.2% of the total possible, yielding pairwise concordance of 97.3% between Doktycz et al.[8] and Breslauer et al.,[9] and 99.8% between Doktycz et al.[8] and SantaLucia et al.[10] Considering the differences in the number of data entries, the duplex length distribution, and the salt concentrations used to construct these n-n sets, the levels of concordance are impressive. Re-

Table IV Discordant Pairs for Pairwise Comparisons of 11 Nearest-Neighbor Sets[a]

Set	Doktycz et al.	Delcourt and Blake	Wartell and Benight	Vologodskii et al.	Gotoh and Tagashira	McCampbell et al.	Ornstein and Fresco	Aida	Breslauer et al.	SantaLucia et al.	Sugimoto et al.
Doktycz et al.		1.7*e* + 05	6.9*e* + 05	0.0*e* + 00	0.0*e* + 00	5.8*e* + 06	8.3*e* + 08	1.4*e* + 10	3.7*e* + 09	2.2*e* + 08	1.1*e* + 08
Delcourt and Blake	1.7*e* + 05		4.5*e* + 07	0.0*e* + 00	1.1*e* + 04	6.1*e* + 04	4.0*e* + 08	1.9*e* + 10	6.0*e* + 09	9.6*e* + 08	3.1*e* + 08
Wartell and Benight	6.9*e* + 05	4.5*e* + 07		7.1*e* + 06	1.3*e* + 08	1.1*e* + 08	2.3*e* + 08	1.7*e* + 10	6.0*e* + 09	1.4*e* + 09	1.3*e* + 09
Vologodskii et al.	0.0*e* + 00	0.0*e* + 00	7.1*e* + 06		0.0*e* + 00	7.5*e* + 05	7.2*e* + 08	1.3*e* + 10	3.6*e* + 09	1.8*e* + 08	5.2*e* + 07
Gotoh and Tagashira	0.0*e* + 00	1.1*e* + 04	1.3*e* + 08	0.0*e* + 00		3.7*e* + 07	1.6*e* + 09	1.9*e* + 10	7.1*e* + 09	7.2*e* + 08	3.7*e* + 08
McCampbell et al.	5.8*e* + 06	6.1*e* + 04	1.1*e* + 08	7.5*e* + 05	3.7*e* + 07		8.4*e* + 08	2.5*e* + 10	5.9*e* + 09	1.4*e* + 09	4.3*e* + 08
Ornstein and Fresco	8.3*e* + 08	4.0*e* + 08	2.3*e* + 08	7.2*e* + 08	1.6*e* + 09	8.4*e* + 08		2.5*e* + 10	1.2*e* + 10	4.3*e* + 09	3.3*e* + 09
Aida	1.4*e* + 10	1.9*e* + 10	1.7*e* + 10	1.3*e* + 10	1.9*e* + 10	2.5*e* + 10	2.5*e* + 10		2.6*e* + 10	1.6*e* + 10	1.8*e* + 10
Breslauer et al.	3.7*e* + 09	6.0*e* + 09	6.0*e* + 09	3.6*e* + 09	7.1*e* + 09	5.9*e* + 09	1.2*e* + 10	2.6*e* + 10		2.2*e* + 09	1.8*e* + 09
SantaLucia et al.	2.2*e* + 08	9.6*e* + 08	1.4*e* + 09	1.8*e* + 08	7.2*e* + 08	1.4*e* + 09	4.3*e* + 09	1.6*e* + 10	2.2*e* + 09		1.5*e* + 05
Sugimoto et al.	1.1*e* + 08	3.1*e* + 08	1.3*e* + 09	5.2*e* + 07	3.7*e* + 08	4.3*e* + 08	3.3*e* + 09	1.8*e* + 10	1.8*e* + 09	1.5*e* + 05	

[a] The number of discordant pairs [assuming an error in $\Delta\Delta G$(25°C) of 5%] for all pairwise combinations of the 11 n-n sets given in Tables II and III determined as described in the text. References for the n-n sets are given in the Figure 1 caption.

Table V Sequences and Experimentally Evaluated Melting Thermodynamics for Eight 10 Base Pair DNAs[a]

Duplex	Top Strand
1	5′-GTAGTAGTAG-3′
2	5′-TGAAAAAAAA-3′
3	5′-TTAATAGGGG-3′
4	5′-ACAGTGACAC-3′
5	5′-GACGTGTGAC-3′
6	5′-ATTATGGGGC-3′
7	5′-AAAAAAAGTT-3′
8	5′-CATATATATG-3′

	115 n*M* Na^+				1*M* Na^+			
Duplex	ΔH	ΔS	ΔG_{25}	$\Delta\Delta G_{25}$	ΔH	ΔS	ΔG_{25}	$\Delta\Delta G_{25}$
1	−86.6 ± 5	−258 ± 18	−9.7 ± 0.1	−2.9 ± 0.1	−73.9 ± 3	−211 ± 8	−11.0 ± 0.1	−1.6 ± 0.1
2	−72.4 ± 4	−220 ± 14	−6.8 ± 0.1		−62.2 ± 2	−177 ± 8	−9.4 ± 0.1	
3	−67.6 ± 3	−196 ± 10	−9.2 ± 0.1	2.6 ± 0.4	−64.9 ± 2	−182 ± 5	−10.6 ± 0.1	3.3 ± 0.6
4	−70.2 ± 7	−196 ± 23	−11.8 ± 0.4		−79.5 ± 8	−220 ± 24	−13.9 ± 0.6	
5	−82.9 ± 2	−234 ± 6	−13.1 ± 0.1	−1.3 ± 0.6	−85.6 ± 4	−237 ± 13	−14.9 ± 0.3	−3.1 ± 0.6
6	−89.3 ± 15	−260 ± 48	−11.8 ± 0.6		−62.5 ± 8	−170 ± 25	−11.8 ± 0.5	
7	−68.7 ± 7	−205 ± 24	−7.6 ± 0.1	−1.1 ± 0.2	−63.3 ± 2	−180 ± 8	−9.6 ± 0.1	−0.8 ± 0.3
8	−81.6 ± 16	−252 ± 55	−6.5 ± 0.2		−83.9 ± 13	−252 ± 42	−8.8 ± 0.3	

[a] For the eight 10 base pair duplexes with the sequences shown at the top, values of the transition enthalpy ΔH and entropy ΔS were evaluated from plots of $1/T_m$ vs $ln(C_T/\alpha)$ in both 115 n*M* and 1.0*M* Na^+ environments. The free-energy values at 25°C, ΔG_{25}, determined from ΔH and ΔS in the two solvent environments, are given for the eight duplexes. Free-energy differences $\Delta\Delta G_{25}$, for duplex pairs 1 & 2, 3 & 4, 5 & 6, and 7 & 8 are also tabulated.

sults of this global relative comparison indicate there is quite good agreement between the n-n sets reported by SantaLucia et al.[10] and Sugimoto et al.[12]

Comparison of Predictions and Experiments for Four Pairs of 10-mer Duplexes

For the eight 10 base pair duplexes with the sequences shown in Table V, plots of $1/T_m$ vs $ln(C_T/\alpha)$ were linear in both 115 m*M* and 1.0*M* Na^+ environments (not shown). Assuming a two-state melting transition, the fitted slopes and intercepts of these linear plots provided evaluations of the transition enthalpy ΔH_D and entropy ΔS_D. Values for these parameters, and the free energy at 25°C, ΔG(25°C) (ΔG_{25} in Table V), in the two solvent environments examined are summarized for each of the 8 duplexes in Table V. The standard deviations on the parameters given in Table V were estimated from error in the fits according to Eq. (20). Note, in Table V the duplexes are grouped 1&2, 3&4, 5& 6, and 7&8. These pairs of duplexes were selected because for several of the n-n sets their predicted $\Delta\Delta G_{25}$ values were some of the most discordant found. Melting experiments were performed on these molecules to determine which n-n sets actually predict the observed order of stability for the four duplex pairs. The experimentally determined difference values, $\Delta\Delta G(25°C) = \Delta\Delta G_{25}$, for each pair, in both Na^+ environments are also given in Table V.

Predictions of the 11 n-n sets and the experimental results given in Table V are compared in Table VI. In the first column, the n-n set is listed and the Na^+ environment where the set was evaluated is given in parentheses. The experimental values from Table V are given in the first row. For each pair of duplexes, the calculated difference at 25°C, $\Delta\Delta G(i\text{-}j) = \Delta G(i) - \Delta G(j)$ was determined using the n-n sets. Comparisons in Table VI reveal the n-n parameters reported in Table II seem to predict the correct order of stability. That is, the predicted $\Delta\Delta G(i\text{-}j)$ has the same sign as experimentally observed for $\Delta\Delta G$(1-2), $\Delta\Delta G$(3-4), and $\Delta\Delta G$(5-6). For the last pair, $\Delta\Delta G$(7-8), seven of the eight n-n sets in Table II (the theoretical set of Aida[7] being the exception) predict the two duplexes to have nearly equivalent stability. The n-n sets summarized in Table III predict the correct order of

Table VI Comparison of Predictions with Experiments for Melting of Eight 10 Base Pair DNAs[a]

	$\Delta\Delta G$(1-2) (kcal/mol)		$\Delta\Delta G$(3-4) (kcal/mol)		$\Delta\Delta G$(5-6) (kcal/mol)		$\Delta\Delta G$(7-8) (kcal/mol)	
	0.115*M* Na^+	1.0*M* Na^+	0.115*M* Na^+	1.0*M* Na^+	0.115*M* Na^+	1.0*M* Na^+	0.115*M* Na^+	1.0*M* Na^+
Experiment	−2.9 ± 0.1	−1.6 ± 0.1	2.6 ± 0.4	3.3 ± 0.6	−1.3 ± 0.6	−3.1 ± 0.6	−1.1 ± 0.2	−0.8 ± 0.3
n-n set								
Doktycz et al. (115 m*M*)	−1.5	—	1.9	—	−1.6	—	0.4	—
Delcourt and Blake (75 m*M*)	−3.4	—	1.8	—	−1.9	—	0.7	—
Wartell and Benight (100 m*M*)	−2.1	—	2.2	—	−1.1	—	0.9	—
Vologodskii et al. (200 m*M*)	−2.0	—	1.7	—	−1.6	—	0.5	—
Gotoh and Tagashira (19.5 m*M*)	−2.4	—	2.5	—	−2.6	—	0.0	—
McCampbell et al. (102 m*M*)	−2.4	—	1.2	—	−1.4	—	−0.2	—
Ornstein and Fresco (50 m*M*)	−6.0	—	3.4	—	−2.1	—	−0.2	—
Aida (50 m*M*)	−2.6	—	10.6	—	−7.7	—	3.8	—
Breslauer et al. (1.0*M*)	—	5.4	—	−3.7	—	4.1	—	−3.9
SantaLucia et al. (1.0*M*)	—	1.4	—	2.2	—	−1.7	—	−3.4
Sugimoto et al. (1.0*M*)	—	0.2	—	1.3	—	−1.2	—	−2.3

[a] Predictions of the 11 n-n sets[1−10,12] are compared with the experimental results given in Table V for eight 10 base pairs DNAs melted in 115 m*M* and 1.0*M* Na^+. The n-n set used for the predictions are listed in the first column. The Na^+ environments where the n-n sets were evaluated are given in parentheses. References for these n-n sets are given in the Figure 1 caption. Differences in free energies at 25°C, $\Delta\Delta G(i\text{-}j)$, for the four sequence pairs, are given in each column. Experimental results (from Table V) are given in the first row. In the following rows, calculated values from each n-n set are given for each duplex pair.

stability for $\Delta\Delta G$(7-8). For $\Delta\Delta G$(1-2), the n-n sets summarized in Table III predict the opposite order of stability to what is experimentally observed. For the most part the magnitudes of the predicted and measured free-energy differences, although close in some cases [for $\Delta\Delta G$(5-6)], are not within the standard deviations summarized in Table V.

At first sight the comparison in Table VI would suggest for the few molecules examined that none of the n-n sets are able to accurately predict all of the observed $\Delta\Delta G(i\text{-}j)$ values within the experimental standard deviation. Initially, this may appear somewhat disheartening. However, we should recall that the sequences examined were selected deliberately because they were found to be the most discordant in the pairwise comparisons of predictions from different n-n sets. Examination of the sequences in Table V reveals they are either repetitive or have strings of A·T base pairs. Perhaps the discordance in these sequences is due the anomalous melting behavior that occurs in them.[58–62] If so, such sequence-dependent features are not currently considered in the n-n model. Also, it is likely that these molecules may not melt in a purely two-state manner, in which case the disagreement underscores the deficiencies and potential problems associated with the two-state approach to melting analysis. Further investigations of these types of sequences will be required to resolve these potential issues. For the purposes of the present study, the real question is, How well do the n-n sets actually predict (on average) the experimental melting temperature, t_m? Results that address this question follow below.

The Salt-Dependent Correction

In order to compare calculated melting stabilities from n-n sets evaluated in 115 m*M* and in 1.0*M*, it was necessary to have a correction factor to scale calculated t_m values from one salt to the other. From the melting data of 16 DNA oligomers with varying %GC, measured in 1.0 and 115 m*M* Na^+ environments, such a correction factor for adjusting the calculated t_m's was determined. Results of the melting analysis are shown in Figure 2. The difference in t_m measured in 115 m*M* and 1.0*M* Na^+, at the same DNA concentration Δt_s, for 16 different oligomers ranging in length from 10 to 24 base pairs is plotted vs the fraction of G·C type base pairs, f_{GC}. Oligomers of different lengths are indicated by the different symbols as defined in the Figure 2 caption. The sloping solid line was fit through the data points with the equation

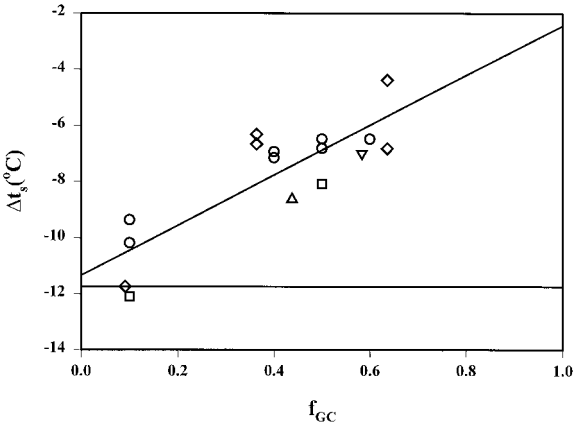

FIGURE 2 Correction factor for scaling t_m between different sodium ion environments. The difference in t_m measured in 115 m*M* and 1.0*M* Na^+, at the same DNA concentration Δt_s for 16 different oligomers ranging in length from 10 to 24 base pairs, is plotted vs the fraction of G·C type base pairs f_{GC}. Oligomers of different lengths are indicated by the different symbols. The base pair lengths and corresponding symbols are as follows. Circles (10), triangles (16), squares (20), diamonds (22), and inverted triangles (24). The sloping line was fit through the data as given in Eq. (21a) of the text. For comparison, the salt correction reported by SantaLucia et al.[10] over a similar Na^+ range is shown by the horizontal line because the reported correction did not include the dependence on f_{GC}.

$$\Delta t_s = t_m(0.115) - t_m(1.0) = (8.91)f_{GC} - 11.34 \quad (21a)$$

It is likely this correction will also depend on duplex length, but the data set used to evaluate it did not reveal a simple length dependence (Figure 2). The salt correction in Eq. (21a) differs from the one reported recently by SantaLucia et al.[10] over a similar Na^+ range. Their reported correction did not include the dependence on f_{GC}. As shown by the solid horizontal line in Figure 2, their equation predicts $\Delta t_s = -11.74$°C. Equation (21a) is very similar to the salt-dependent correction of t_m for polymers reported by Frank-Kamenetskii[23] over 25 years ago. Using his expression for t_m, a comparable expression for Δt_s between 1.0*M* and 115 m*M* Na^+ is obtained;

$$\Delta t_s = (6.62)f_{GC} - 17.19 \quad (21b)$$

Considering the differences in the DNA samples used to arrive at Eqs. (21a) and (21b), their similarity is remarkable. The difference in Eqs. (21a) and

(21b) is $\Delta\Delta t_s = 2.29 f_{GC} + 5.85$, revealing that the differences in t_m between 115 mM and 1.0M Na^+ provided by either equation ranges from 5.85 to 8.15°C. Apparently, this reveals that the salt- and sequence-dependent stability of oligomers and polymers, although qualitatively similar, is quantitatively different.

Breslauer has reported that for oligomeric duplexes shorter than about 14 base pairs, the change in t_m with salt concentration, $\partial T_m/\partial \log[Na^+]$, depends primarily on the duplex length n up to $[Na^+]$ of about 0.4M.[63,64] In fact, Breslauer suggests that one can crudely approximate $\partial T_m/\partial \log[Na^+]$ by assuming the derivative has a value of $(n + 2)$, with base composition and sequence identity exerting only secondary effects. The somewhat stronger dependence of t_m on $[Na^+]$ predicted by this approach reflects the lower salt concentration range over which these studies were conducted where the $\partial T_m/\partial \log[Na^+]$ dependence has not yet begun to plateau, as it does between 115 mM and 1.0M Na^+.

Calculated Stability of DNA Oligomers

Results of the statistical comparison summarized in Table IV suggests that there is only minor disagreement between the calculated melting stability of all possible 10 base pair sequences using the different n-n sets. However, the comparison in Table VI indicates that for certain sequences the different n-n sets cannot predict experimental results with quantitative agreement. Although these comparisons are interesting and provide a new vantage from which to compare the n-n sequence dependent interactions, they provide little practical utility. What remains is to determine which n-n sets can in general provide the most quantitatively accurate predictions of experimentally measured t_m's of duplex DNA oligomers given their sequence and concentration. For the comparisons that follow we restricted our attention to a subset of five n-n sets. These are the set of Doktycz et al.[8] evaluated from analysis of melting curves of short DNA dumbbells (column 1 of Figure 2) and the four sets evaluated from melting curves of short duplex DNA oligomers. These sets were reported by Breslauer et al.,[9] actually determined from optical and calorimetric melting analysis of short DNA oligomers and long synthetic repeating DNA polymers, SantaLucia et al.,[10] Allawi and SantaLucia,[11] and Sugimoto et al.[12] Values for these n-n sets are given in Table III. Comparisons of the n-n sets occurred in several stages. First, the five n-n sets were employed to predict the reported experimentally measured t_m's as a function of concentration for 251 DNA oligomers ranging in length from 4 to 16 base pairs melted in 1.0M NaCl. Of course, this small range of duplex lengths will necessarily bias against predictions from n-n sets (such as that of Breslauer et al.[9]) that also involve significant polymer melting data. For the calculations that were performed, the n-n parameters determined from dumbbells, evaluated in 115 mM Na^+, were scaled to 1.0M Na^+ using the correction in Eq. (21a). The other n-n sets were determined in 1.0M Na^+ and used directly. For every duplex sequence a two-state melting process was assumed and t_m was calculated according to Eq. (19) using the n-n values in Tables II and III.

The required sequence-dependent melting enthalpies and entropies ΔH_{duplex} and ΔS_{duplex} were calculated directly from the base pair sequence using the parameters in Tables II and III and Eqs. (1)–(7). Thermodynamic parameters of nucleation ΔH_{nuc} and ΔS_{nuc} were varied and evaluated as described below. The constant α was determined by symmetry of the sequence; $\alpha = 1$ for self-complementary sequences and $\alpha = 4$ for nonself-complementary sequences. At an arbitrary total strand concentration $C_T = 4\ \mu M$, the t_m of each of the 251 DNA oligomers in the data set was predicted, $t_m(P)$, and compared directly with the published experimentally measured values, $t_m(M)$, that were either reported at that concentration or extrapolated to 4 μM using reported ΔH_D and ΔS_D values. The corresponding absolute value of the difference between predicted $t_m(P)$ and measured $t_m(M)$, $|\Delta t_m|$, was found for each DNA sequence. Then the average $|\Delta t_m|$ over the entire 251 member set, $|\Delta t_m|_{\text{ave}}$, was determined. This procedure was repeated using each of the five n-n sets, and assuming four different combinations of ΔH_{nuc} and ΔS_{nuc}. Results are displayed in Table VII. In the first case (F1), predictions of $t_m(P)$ for the 251 DNAs were made using the published suggested values for ΔH_{nuc} and ΔS_{nuc}. Values taken from the literature for the five n-n sets examined were employed as follows: $\Delta G_{\text{nuc}}(37°C) = RT\ln \beta = \Delta H_{\text{nuc}} - \Delta S_{\text{nuc}}(310.15)$. The n-n sets of Breslauer et al.[9] and SantaLucia et al.,[10] assigned values of $\Delta G_{\text{nuc}}(37°C) = 6.23$ and 2.79 kcal/mol, respectively, to duplexes comprised exclusively of $A \cdot T$ type base pairs. Breslauer et al.[9] and SantaLucia et al.[10] also reported values of $\Delta G_{\text{nuc}}(37°C) = 5.21$ and 1.83 kcal/mol, respectively, for duplexes containing at least one $G \cdot C$ type base pair. As seen at the bottom of Table III, these sets assume for the most part that $\Delta G_{\text{nuc}}(37°C)$ is entirely entropic, and $\Delta H_{\text{nuc}} = 0$. An exception is for the set of SantaLucia et al.[10] that assumes a slight enthalpic penalty, ΔH_{nuc}

Table VII Predictions for 251 DNA Duplex Molecules Ranging in Length From 4 to 16 Base Pairs[a]

	$\lvert\Delta t_m\rvert_{ave} \pm \sigma$			
Nearest Neighbor Set	F1	F2	F3	F4
Doktycz et al.	19.7 ± 4.4	2.8 ± 2.3	2.3 ± 1.9	1.6 ± 1.5
Breslauer et al.	6.4 ± 5.2	4.9 ± 4.1	4.2 ± 3.9	4.4 ± 3.8
SantaLucia et al.	2.0 ± 1.7	2.0 ± 1.7	1.8 ± 1.6	1.8 ± 1.6
Allawi and SantaLucia	1.5 ± 1.3	1.5 ± 1.3	1.5 ± 1.2	1.5 ± 1.2
Sugimoto et al.	2.0 ± 1.6	2.0 ± 1.6	1.8 ± 1.5	1.8 ± 1.4

[a] The five n-n sets indicated were employed to predict reported experimentally measured t_m's as a function of strand concentration of 251 DNA oligomers ranging in length from 4 to 16 base pairs melted in 1.0*M* NaCl. The absolute value of the difference between predicted t_m(P) and measured t_m(M), $|\Delta t_m|$, was found for each DNA sequence, and the average $|\Delta t_m|$ over the entire 251 member set, $|\Delta t_m|_{ave}$, and standard deviation σ were determined. Four different combinations of ΔH_{nuc} and ΔS_{nuc} were assumed. For the results in the first column (F1) predictions of t_m(P) were made using the suggested published values for ΔH_{nuc} and ΔS_{nuc}, as described in the text and listed in Table III. Results summarized in the second column (F2) correspond to the values of ΔH_{nuc} used in the F1 calculations with ΔS_{nuc} varied as a single adjustable constant to fit the data. Results in the third column (F3) correspond to the initial values of ΔH_{nuc} used in the F1 calculation and assuming the three parameter linear functional form, $\Delta S_{nuc} = Rln(S_1 + S_2(f_{GC}) + S_3(N))$, where S_1, S_2, and S_3 are linear coefficients, f_{GC} is the fraction of G·C type base pairs, and N is the oligomer length. Results in the fourth column (F4) were obtained setting ΔS_{nuc} values to those used in the F1 calculation and assuming the three parameter functional form, $\Delta H_{nuc} = H_1 + H_2(f_{GC}) + H_3(N)$.

= 0.4 kcal/mol, for duplexes that contain a 5′-*T*·*A*-3′ base pair. For the n-n set of Sugimoto et al.,[12] a constant value of $\Delta G_{nuc}(37°C) = 3.39$ kcal/mol was assumed, comprised of an enthalpic component, $\Delta H_{nuc} = 0.6$ kcal/mol, and entropic component, $-\Delta S_{nuc}(310.15) = 2.79$ kcal/mol. For the n-n set of Allawi and SantaLucia,[11] $\Delta G_{nuc}(37°C)$ depends on the identity of the base pair that terminates the duplex. For a terminal $A \cdot T$ base pair, $\Delta G_{nuc}(37°C) = 1.03$ kcal/mol. For a terminal $G \cdot C$ base pair $\Delta G_{nuc}(37°C) = 0.98$ kcal/mol. These nucleation free energies are comprised of different enthalpic and entropic contributions depending on the terminal base pair as summarized at the bottom of Table III. For the above mentioned n-n sets,[9–12] a symmetry correction, $\Delta G_{sym}(37°C) = 0.43$ kcal/mol was also included in the calculations for self-complementary sequences. For the n-n set determined from melting analysis of DNA dumbbells,[8] in the first case (F1 of Table VII) it was assumed that $\Delta H_{nuc} = 0$, $\Delta S_{nuc} = 0$ and $\Delta G_{sym}(37°C) = 0$.

Resulting $|\Delta t_m|_{ave}$ values are summarized in the first column (F1) of Table VII. The set of Doktycz et al.[8] provides poor predictions of the data, with the set of Breslauer et al.[9] not being quite as good as the other n-n sets, which yield good predictions of the data. We should mention of course that the different sets of molecules used to evaluate these n-n sets are a subset of the 251 molecules predicted. We next investigated whether improved predictions could be obtained using the Breslauer et al.[9] and Doktycz et al.[8] n-n sets by making adjustments in ΔS_{nuc}. Results presented in the second column of Table VII (F2) were obtained using the initial values of ΔH_{nuc}, and varying ΔS_{nuc} as a single adjustable constant to fit all of the data. As can be seen, this improves the predictions of these n-n sets. Predictions of the other n-n sets are unchanged. In the third column (F3) results are presented that were obtained with the initial values of ΔH_{nuc}, and assuming the three parameter linear functional form $\Delta S_{nuc} = R\ ln(S_1 + S_2(f_{GC}) + S_3(N))$, where S_1, S_2, and S_3 are linear coefficients, f_{GC} is the fraction of $G \cdot C$ type base pairs, and N is the oligomer length. Evidently when this assumption is invoked improvements over using a single parameter are not seen. Finally, in the fourth column (F4), ΔS_{nuc} values were set to their suggested constant literature values, and the three parameter functional form $\Delta H_{nuc} = (H_1 + H_2(f_{GC}) + H_3(N))$, was assumed. The linear coefficients were varied to produce the best global fit to all the data. Table VII indicates the n-n set of Doktycz et al.[8] evaluated from melting curves of DNA dumbbells, gets slightly better if a nonzero ΔH_{nuc} is assumed. Predictions of the other n-n sets do not improve significantly when a sequence-dependent, nonzero ΔH_{nuc} is assumed. This observation is consistent with the molecular structure of the DNA samples. Because of the loops on both ends the energetic cost of nucleating duplexes in dumbbells is minimal. Therefore, it is perhaps not surprising that when used to predict melting behavior of linear duplex oligomers, the n-n set evaluated from dumbbells requires a significant correction to ac-

count for nucleation. The evaluated nucleation enthalpy that should be used in conjunction with the dumbbell n-n parameters is

$$\Delta H_{\text{nuc}} = 7945.5 - 3413.0(f_{GC}) - 201(N) \text{ cal/mol} \quad (22)$$

Thus, as f_{GC} and N increases, the enthalpic cost of duplex nucleation decreases.

Results summarized in Table VII indicate that four of the five n-n sets examined are able to provide predictions of the melting temperatures of 251 duplex DNA sequences ranging in length from 4 to 16 base pairs to within $\sim 1.8 \pm 1.6$°C, with the set of Breslauer et al.[9] predicting t_m within 4.4 ± 3.8°C. There are several potential sources for this observation. It should be noted that the latter n-n set emerged after the initial pioneering work of Breslauer and Marky on calorimetric melting of DNA oligomers.[18] This n-n set is at least six years older than the other n-n sets. Due to the less developed protocols for preparing and purifying synthetic DNAs at that time, fewer sequence variants could be studied. Thus, the n-n set benefits less from averaging of effects. In addition, the n-n set was evaluated directly by calorimetric measurements. In the calorimetric measurement, the primary experimental observable is the transition enthalpy ΔH. The remaining thermodynamic parameters (ΔS and ΔG) are determined in a manner that is uncoupled to the t_m values. As a result, in contrast to most analyses of optical melting data, the thermodynamic parameters derived calorimetrically are not computationally coupled in a manner that favors optimal t_m predictions. Instead, they provide the most direct model-independent determination of ΔH, ΔS, and ΔG. Therefore, their actual accuracy may be best assessed by predictions of calorimetrically determined transition enthalpies rather than t_m values.

Finally, given the significant contributions from polymer melting data to the n-n set of Breslauer et al.,[9] the average experimental t_m values of the molecules they studied were considerably higher than those of the test oligomeric duplexes with lengths between 4 and 16 base pairs. Consequently, significant temperature extrapolation was required for application of the Breslauer et al.[9] n-n set to the duplexes of the test set, thereby placing a large burden on the assumptions that $\Delta C_p = 0$ and the value of $\partial T_m / \partial \log[\text{Na}^+]$.

How Well Do the n-n Sets Predict the Melting Behavior of 76 Independent Molecules?

We wished to see how well the five n-n sets could predict the measured stability of the 76 independently measured molecules shown in the Appendix Table A1. The predicted $|\Delta t_m|_{\text{ave}}$ values for the molecules are shown in Table VIII. Parameters derived from fitting the melting data of the 251 molecules with the results shown in Table VII were used to calculate the t_m values of the 76 molecules from their sequences. Results given in columns F1–F4 of Table VIII were obtained using the parameter values evaluated from the fits summarized in columns F1–F4 of Table VII. These results reveal that the n-n sets reported by Doktycz et al.,[8] Allawi and SantaLucia,[11] and Sugimoto et al.[12] can accurately predict the average melting temperature of the independent set within $\sim 2.4 \pm 1.8$°C. Considering the discrepancies in the calculated and measured thermodynamics for the eight 10-mers, summarized in Table VI, which are part of the independent set, this predictive accuracy is very good. Probably for the reasons given above, the n-n set reported by Breslauer et al.[9] predicts a somewhat higher average deviation of 6.3 ± 4.7°C.

It should be emphasized that the reported $|\Delta t_m|_{\text{ave}}$ are averages over the entire set of molecules ranging in length from 5 to 24 base pairs. All were assumed to melt in a two-state manner, but this was not verified to be so in every case. In fact, several of the longer fragments 22 base pairs in length have been found by comparisons with direct calorimetric measurements to deviate from two-state melting behavior (data not shown). And Δt_m values for these molecules are outside the average range given above. Thus, including these molecules necessarily acts to increase the spread in predicted average $|\Delta t_m|_{\text{ave}}$ values above those found in Table VII. Thus, the values summarized in Table VII, which were determined from analysis of molecules that display two-state melting transitions, probably provide a more accurate indication of the range of $|\Delta t_m|_{\text{ave}}$.

The Influence of Ends

Utilizing results from melting analysis of 114 DNA oligomers measured in 1.0M Na^+ reported by Allawi and SantaLucia[11] we investigated whether n-n sequence-dependent interactions with the ends contribute significantly to the stability of short duplex DNAs. Evaluation of the 10 unique n-n sequence-

Table VIII Predictions for 76 Independent Duplex DNA Molecules Ranging in Length from 5 to 24 Base Pairs[a]

	$\lvert\Delta t_m\rvert_{ave} \pm \sigma$			
Nearest Neighbor Set	F1	F2	F3	F4
Doktycz et al.	9.9 ± 5.9	3.4 ± 2.4	3.0 ± 1.9	2.4 ± 1.8
Breslauer et al.	8.3 ± 5.7	10.3 ± 5.7	8.1 ± 4.8	6.3 ± 4.7
SantaLucia et al.	3.4 ± 2.3	3.5 ± 2.4	4.6 ± 3.1	4.1 ± 2.9
Allawi and SantaLucia	2.3 ± 1.7	2.3 ± 1.7	3.1 ± 2.1	2.5 ± 1.8
Sugimoto et al.	4.6 ± 2.6	4.5 ± 2.6	2.4 ± 1.8	2.3 ± 1.8

[a] Parameters derived from fitting melting data of 251 molecules with the results shown in Table VII, were used to calculate the t_m values of 76 independent molecules. Results shown in columns F1–F4 were obtained using the parameter values evaluated from the fits that provided the results given in columns F1–F4 of Table VII.

dependent interactions in DNA depends on how end interactions are treated. If end interactions are assumed to be zero, then there are 10 possible unique n-n sequence-dependent interactions that can be evaluated. However, if sequence-specific ends are considered there are 14 unknowns, i.e., the 10 n-n sequence interactions and the four possible n-n interactions with the ends. However when n-n explicit ends are considered, only 12 linearly independent equations can be written and thus only 12 linear combinations of the n-n interactions can be solved for. Of these 12 linear combinations, two include contributions from only the n-n sequence-dependent end interactions as shown in Table II of Goldstein and Benight.[19]

Results of the evaluation of the 12 linearly independent combinations, assuming first that the end interactions are zero and fitting the remaining 10 combinations, then fitting for all 12 linear combinations were obtained. The reported transition enthalpies, entropies, and free-energies at 37°C were fit in the analysis. Results of this analysis are shown in Table IX. Results on the left (without ends) were obtained assuming the n-n sequence-dependent interactions with ends are zero. On the right (explicit ends included) are results obtained when the end interactions are explicitly fit from the data. Interestingly, the first eight linear combinations, θ_1 through θ_8, which consider only the n-n sequence-dependent interactions of the base pairs in the duplex (not the ends), are in reasonable agreement (within 25% for any thermodynamic parameter) whether the ends are fit or assumed to zero. Further on the right side of Table IX, the linear combinations corresponding explicitly to the ends (θ_9 and θ_{10})[19] indicate although their fitted ΔH and ΔS values are different, their ΔG(37°C) are essentially identical. The same result is obtained if a single end interaction (referred to by some[11,12] as a nucleation parameter) is assumed. In this case there are 11 unknowns to be solved for from 11 linearly independent equations that can be constructed. However, the fact that a single 11th parameter can be fit within the error is because the free energies of the end interactions are the same. Recently, Gray has reported a detailed study on the analytical treatment of sequence-dependent ends and n-n parameters.[20,21]

SUMMARY AND CONCLUSIONS

Eleven sets[1–10,12] of published n-n stability parameters were compared through differences in their predicted stabilities of all possible nonself-complementary 10 base pair sequences. Five of the published stability parameter sets were employed to calculate the melting stability of over 300 short (6–24 base pairs) oligomer sequences melted in solvents containing 0.115 and 1.0*M* Na^+. From a subset of the collected data a linear expression for extrapolating t_m between these ionic strength environments was derived. This salt correction found for oligomers is in qualitative agreement with a similar correction reported for the ionic strength dependence of the melting stability of long duplex DNA polymers.

The analyses reveal that the n-n set evaluated from melting analysis of DNA dumbbells[8] requires a substantial unfavorable correction for the enthalpy of nucleation ΔH_{nuc}, that depends on the %*GC* and length of the duplex. When this correction is employed, the dumbbell n-n set is able to provide accurate predictions of t_m's of short DNA oligomers. This predictive accuracy is comparable to that obtained using several recently reported new and im-

Table IX Effects of Nearest-Neighbor Sequence-Dependent Interactions with the Ends[a]

Linear Combination of Subunits	Without Ends			Explicit Ends Included		
	ΔH (cal/mol)	ΔS cal · mol^{-1} · K^{-1}	ΔG_{37} (cal/mol)	ΔH (cal/mol)	ΔS cal · mol^{-1} · K^{-1}	ΔG_{37} (cal/mol)
θ_1	−7908	−22.97	−791.5	−7750	−21.73	−1013.3
θ_2	−8441	−22.39	−1507.3	−7952	−19.77	−1829.4
θ_3	−6303	−18.63	−543.8	−6373	−18.39	−681.0
θ_4	−10385	−27.90	−1740.7	−9866	−24.84	−2165.8
θ_5	−8084	−22.40	−1142.7	−7797	−20.59	−1413.0
θ_6	−7735	−21.60	−1027.8	−7526	−20.06	−1291.4
θ_7	−164	−0.45	−22.1	−78	−0.17	−23.0
θ_8	202	0.78	−36.1	165	0.59	−15.9
θ_9	0	0.00	0.0	169	−2.79	993.1
θ_{10}	0	0.00	0.0	−1374	−7.70	1002.0
θ_{11}	362	1.11	19.0	44	0.13	2.7
θ_{12}	−49	0.12	−91.7	−6	0.01	−10.2

[a] The 12 linearly independent combinations of n-n sequence-dependent interactions were evaluated in two ways. First, assuming end interactions were zero and fitting the remaining 10 linear combinations, then fitting for all 12 linear combinations. Results on the left (without ends included) were obtained for the former case. Results on the right (explicit ends) were obtained for the later case. Note, the first eight linear combinations, θ_1 through θ_8, which correspond to sequence-dependent interactions of the base pairs (not the ends) are in reasonable agreement (within 25% for any thermodynamic parameter), whether the ends are fit or assumed to zero. Further, the linear combinations corresponding explicitly to the ends (θ_9 and θ_{10})[19] indicate, although their fitted ΔH and ΔS values are different, their ΔG (37°C) are essentially the same.

proved n-n sets[10–12] evaluated from melting experiments of short DNAs directly.

To summarize, given the concentration and sequence, the t_m's of duplex DNA oligomers having 20 or less duplex base pairs, in solvent ionic strength ranging from 115 m*M* to 1.0*M* Na^+, can be calculated (on average) to within 2°C, provided the oligomers melt in a two-state manner. However, as we found for certain 10-mer sequences, agreement between the calculated sequence-dependent melting thermodynamics and parameters evaluated directly from two-state model analysis of melting data cannot always be obtained (using any n-n set). Whatever the source of these observations, i.e., peculiar melting behavior of the particular sequences examined and/or deviations of their melting transitions from a strictly two-state process, they serve to underscore shortcomings of the two-state model and raise the following concern. For any sequence, unless two-state melting is independently verified, uncertainties due to potential deviations from two-state behavior will always surround predictions of melting stability based on the two-state model.

Table A1 Melting Data for the 76 Independent Duplex DNA Molecules[a]

Sequence 5′-3′	*N*	*Ct* (μM)	t_m[b]	t_m[c]	Ref.
CCCGGG	6	400.0	—	43.0	31
TCATGA	6	100.0	—	21.8	13
TGATCA	6	100.0	—	20.9	13
CCCAGGG	7	100.0	—	48.0	31
CGGCGCCG	8	46.0	—	66.9	33
GGGCGCCC	8	15.0	—	61.0	34
GGGTACCC	8	15.0	—	42.0	34
TGAATTCT[f]	8	99.8	21.2	—	—
GGATGGGAG	9	4.4	—	43.6	35
TATCCCTAT	9	1.9	—	30.1	36
AAAAAAAGTT[e,f]	10	46.6	29.2	—	—

Table A1 **(***Continued from the previous page.*)

Sequence 5′-3′	*N*	*Ct* (μM)	t_m[b]	t_m[c]	Ref.
ACAGTGACAC[e,f]	10	4.3	42.3	—	—
ATTATGGGGC[e,f]	10	1.8	35.8	—	—
CATATATATG[e,f]	10	5.5	21.2	—	—
CGTATTATGC	10	10000.0	—	64.0	37
GAAAATTTTC	10	230.0	—	49.9	38
GACGTGTGAC[e,f]	10	35.6	48.8	—	—
GCCGGATCGC	10	1.8	—	58.0	39
GTAGTAGTAG[e,f]	10	3.6	29.6	—	—
GTTTTAAAAC	10	230.0	—	45.6	38
TGAAAAAAAA[e,f]	10	4.2	20.5	—	—
TTAATAGGGG[e,f]	10	3.8	28.9	—	—
AAAAAGCTTTTT[f]	12	5.0	33.7	—	14
ACACCCAATTCT	12	20.0	—	57.0	40
ATATAGCTATAT[e,f]	12	5.0	30.4	—	14
ATATATGGATAC[f]	12	2.5	28.6	—	—
CCGGCCGCGCGC	12	5.0	—	76.8	33
CGCAAATTTGCG	12	40.0	—	68.5	41
CTTTCTCTCCCT	12	5.0	—	53.3	42
GTTGGTTGGTTG	12	8.3	—	61.5	28
TGGGGCTGCATG	12	6.0	—	60.0	43
CGCATGAGTACGC[f]	13	435.0	—	74.0	44, 45
CGCATGGGTACGC	13	6.0	—	63.6	46
GCGTACGCATGCG[f]	13	5.0	—	63.5	47
AAAAGAATTCTTTT[f]	14	2.4	35.5	—	—
ATATGAATTCATAT[f]	14	2.3	37.0	—	—
GTATACCGGTATAC	14	100.0	—	61.9	10
TAATTAATTAATTA	14	0.7	—	38.3	48
TGACTTAGCTGCAT	14	6.0	—	57.0	43
AAAAAAAGCTTTTTTT[e,f]	16	5.0	44.0	—	14
AAATATAGCTATATTT[e,f]	16	5.0	40.4	—	14
AACGTGAATTCTGGCA[e,f]	16	7.4	58.3	—	—
AATATGAATTCTAATA[f]	16	6.2	41.4	—	—
ATATATAGCTATATAT[e,f]	16	5.0	39.1	—	14
CATATTGGCCAATATG	16	100.0	—	65.3	10
GTATAACCGGTTATAC	16	100.0	—	65.9	10
AAGCTATATATATCGAT[d,f]	17	4.9	50.2	—	—
AATATATAGCTATCGAT[d,f]	17	3.9	49.7	—	—
AATATATGATCATCGAT[d,f]	17	3.3	50.0	—	—
AGATCATATATATCGAT[d,f]	17	4.4	49.2	—	—
AAAAAAAAAGCTTTTTTTTT[e,f]	20	5.0	51.0	—	14
AAATAAATAGCTATTTATTT[e,f]	20	6.5	47.1	—	—
AACGCGTGAATTCTGGCCAA[e,f]	20	6.0	66.6	—	—
AATATATGAATTCTAATTAA[e,f]	20	5.7	44.7	—	—
ATAAATAAAGCTTTATTTAT[f]	20	6.9	46.2	—	—
ATATATATAGCTATATATAT[e,f]	20	5.0	44.8	—	14
ATATATATAGCTTTTTTTTT[f]	20	6.2	44.3	—	—
GCTAAAAAGAGAGAGAGATCG	21	3.2	—	69.0	49
AAAAAAAAGAATTCTTTTTTTT[e,f]	22	6.3	49.9	—	—
AAATATAAGAATTCTTATATTT[e,f]	22	6.7	48.9	—	—
AACGCCGGTAGAGTGCGCGCAA[e,f]	22	5.9	79.1	—	—
AACGCGCGTAGAGTAATTATAA[e,f]	22	5.6	65.0	—	—
AACGCGCGTAGAGTGGCCGCAA[e,f]	22	5.9	77.3	—	—
AATATATATAGAGTAATTATAA[e,f]	22	4.8	45.1	—	—

Table A1 (*Continued from the previous page.*)

Sequence 5′-3′	*N*	*Ct* (μM)	t_m[b]	t_m[c]	Ref.
AATATATATAGAGTGCGCGCAA[e,f]	22	3.3	64.1	—	—
AATATATATAGAGTGGCCGCAA[e,f]	22	5.5	65.2	—	—
ACTGATTAGGATCCATATGTCA[f]	22	4.8	58.3	—	—
ACTGGCGAGGATCCAGCGGTCA[f]	22	4.1	71.8	—	—
ATAAAAATGAATTCATTTTTAT[e,f]	22	7.1	52.8	—	—
ATAAATTAGGATCCATATAAAT[f]	22	5.8	47.7	—	—
ATACGCGAGGATCCAGCGCAAT[f]	22	3.9	71.1	—	—
ATATATATGAATTCATATATAT[e,f]	22	5.9	48.9	—	—
GCGGATTAGGATCCATATCGCG[f]	22	3.9	65.4	—	—
GCGGGCGAGGATCCAGCGCGCG[f]	22	5.3	77.2	—	—
AACGCGCGTGAATTCTGGCCGCAA[e,f]	24	5.5	75.8	—	—
AATATATATGAATTCTAATTATAA[f]	24	4.4	46.2	—	—

[a] Experimental data for 76 duplex DNA molecules are summarized. The 5′-3′ sequences are listed in the first column. Subsequent columns correspond to the number of base pairs (*N*), the strand concentration [$C_t(\mu M)$] where the transition temperature t_m of each duplex was determined, and the reference for the sequence.

[b] Melted in 115 m*M* Na^+.

[c] Melted in 1.0*M* Na^+.

[d] Duplexes contain one base 5′ dangling ends.

[e] At least one experiment was performed in both 0.115 m*M* Na^+ and 1.0*M* Na^+.

[f] Duplex was melted at two or more concentrations.

We gratefully acknowledge helpful comments and suggestions of the Dr. Ken Breslauer. Portions of the work were supported by NIH grant GM39471.

REFERENCES

1. Wartell, R. M. & Benight, A. S. (1985) *Phys. Rep.* **126,** 67–107.
2. Gotoh, O. & Tagashira, Y. (1981) *Biopolymers* **20,** 1033–1042.
3. Vologodskii, A. V., Amirikyan, B. R., Lyubchenko, Y. L. & Frank-Kamenetskii, M. D. (1984) *J. Biomol. Strut. Dynam.* **2,** 131–148.
4. McCampbell, C. R., Wartell, R. M. & Plaskon, R. R. (1989) *Biopolymers* **28,** 1745–1758.
5. Delcourt, S. G. & Blake, R. D. (1991) *J. Biol. Chem.* **266,** 15160–15169.
6. Ornstein, R. L. & Fresco, J. R. (1983) *Biopolymers* **22,** 1979–2000.
7. Aida, M. (1988) *J. Theor. Biol.* **130,** 327–335.
8. Doktycz, M. J., Goldstein, R. F., Paner, T. M., Gallo, F. J. & Benight, A. S. (1992) *Biopolymers* **32,** 849–864.
9. Breslauer, K. J., Frank, R., Blöcker, H. & Marky, L. A. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 3746–3750.
10. SantaLucia, J., Jr., Allawi, H., & Seneviratne, P. (1996) *Biochemistry* **35,** 3555–3562.
11. Allawi, H. T. & SantaLucia, J., Jr. (1997) *Biochemistry* **36,** 10581–10594.
12. Sugimoto, N., Nakano, S.-I., Yoneyama, M. & Honda, K.-I. (1996) *Nucleic Acids Res.* **24,** 4501–4505.
13. Sugimoto, N., Honda, K.-I., & Sasaki, M. (1994) *Nucleos. Nucleot.* **13,** 1311–1317.
14. Benight, A. S., Gallo, F. J., Paner, T. M., Bishop, K. D., Faldasz, B. D. & Lane, M. J. (1995) *Adv. Biophys. Chem* **5,** 1–55.
15. Mullis, K. B. (1990) *Ann. Biol. Clin.* **48,** 579–582.
16. O'Donnell-Maloney, M. J. & Little, D. P. (1996) *Genet. Anal.* **13,** 151–157.
17. Rychlik, W., Spencer, W. J. & Rhoads, R. E. (1990) *Nucleic Acids Res.* **18,** 6409–6412.
18. Marky, L. A. & Breslauer, K. J. (1982) *Biopolymers* **21,** 2185–2194.
19. Goldstein, R. F. & Benight, A. S. (1992) *Biopolymers* **32,** 1679–1693.
20. Gray, D. M. (1997). *Biopolymers* **42,** 783–793.
21. Gray, D. M. (1997). *Biopolymers* **42,** 795–810.
22. Gray, D. M. & Tinoco, I. (1970) *Biopolymers* **9,** 223–244.
23. Frank-Kamenetskii, M. D. (1971) *Biopolymers* **10,** 2623–2624.
24. Marky, L. A. & Breslauer, K. J. (1987) *Biopolymers* **26,** 1601–1620.
25. Poland, D. & Scheraga, H. A. (1970) in *Theory of Helix-Coil Transitions in Biopolymers: Statistical Mechanical Theory of Order-Disorder Transitions in Biological Macromolecules,* Academic Press, New York.

26. Benight, A. S., Wartell, R. M. & Howell, D. K. (1981) *Nature* **289,** 203–205.
27. Benight, A. S. & Wartell, R. M. (1983) *Biopolymers* **22,** 1409–1425.
28. Ratmeyer, L., Vinayak, R., Zhong, Y. Y., Zon, G. & Wilson, W. D. (1994) *Biochemistry* **33,** 5298–5304.
29. Doktycz, M. J., Morris, M. D., Dormady, S. J., Beattie, K. L. & Jacobson, K. B. (1995) *J. Biol. Chem.* **270,** 8439–8445.
30. Sheppard, T. L. & Breslow, R. C. (1996) *J. Am. Chem. Soc.* **118,** 9810–9811.
31. Arnold, F. H., Wolk, S., Cruz, P. & Tinoco, I. Jr. (1987) *Biochemistry* **26,** 4068–4075.
32. Lesnik, E. A. and Frier, S. M. (1995) *Biochemistry* **34,** 10807–10815.
33. Jung, K.-E. & Switzer, C. (1994) *J. Am. Chem. Soc.* **116,** 6059–6061.
34. Rabinovich, D., Haran, T., Eisenstein, M. & Shakked, Z. (1988) *J. Mol. Biol.* **200,** 151–161.
35. Bolli, M., Trafelet, H. U. & Leumann, C. (1996) *Nucleic Acids Res.* **24,** 4660–4667.
36. Aurup, H., Tuschl, T., Benseler, F., Ludwig, J. & Eckstein, F. (1994) *Nucleic Acids Res.* **22,** 20–24.
37. Taylor, J.-S., Garrett, D. S., Brockie, I. R., Svoboda, D. L. & Telser, J. (1990) *Biochemistry* **29,** 8858–8866.
38. Park, Y.-W. & Breslauer, K. J. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 1551–1555.
39. vanHemelryck, B., Guittet, E., Chottard, G., Girault, J.-P., Herman, F., Huynh-Dinh, T., Lallemand, J.-Y., Igolen, J. & Chottard, J.-C. (1986) *Biochem. Biophys. Res. Commun.* **138,** 758–763.
40. Morvan, F., Porumb, H., Degols, G., Lefebvre, I., Pompon, A., Sproat, B. S., Rayner, B., Malvy, C., Lebleu, B. & Imbach, J. L. (1993) *J. Med. Chem.* **36,** 280–287.
41. Leonard, G. A., Thomson, J., Watson, W. P. & Brown, T. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 9573–9576.
42. Hashimoto, H., Nelson, M. G. & Switzer, C. (1993) *J. Am. Chem. Soc.* **115,** 7128–7134.
43. Quartin, R. S. & Wetmur, J. G. (1989) *Biochemistry* **28,** 1040–1047.
44. Vesnaver, G., Chang, C.-N., Eisenberg, M., Grollman, A. P. & Breslauer, K. J. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 3614–3618.
45. Vesnaver, G. & Breslauer, K. J. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 3569–3573.
46. Plum, G. E., Grollman, A. P., Johnson, F. & Breslauer, K. J. (1992) *Biochemistry* **31,** 12096–12102.
47. Plum, G. E., Grollman, A. P., Johnson, F. & Bresaluer, K. J. (1995) *Biochemistry* **34,** 16148–16160.
48. Kibler-Herzog, L., Zon, G., Uznanski, B., Whittier, G. & Wilson, W. D. (1991) *Nucleic Acids Res.* **19,** 2979–2986.
49. Hildbrand, S. & Leumann, C. (1996) *Angew. Chem. Int. Ed. Engl.* **35,** 1968–1970.
50. Caruthers, M. H. (1985) *Science* **230,** 281–285.
51. Cantor, C. R. & Warshaw, M. M. (1970) *Biopolymers* **9,** 1059–1077.
52. Kaiser, J. F. & Reed, W. A. (1977) *Rev. Sci. Instrum.* **48,** 1447–1457.
53. Meyer, S. L. (1975) in *Data Analysis for Scientists and Engineers,* John Wiley & Sons, New York.
54. Bevington, P. R. & Robinson, D. K. (1992) in *Data Reduction and Error Analysis for the Physical Sciences,* McGraw-Hill, New York.
55. Magnus, P. & Guengerich, F. P. (1994) *Biochemistry* **33,** 8662–8672.
56. SantaLucia, J. Jr., Kierzek, R. & Turner, D. H. (1991) *J. Am. Chem. Soc.* **113,** 4313–4322.
57. Chaires, J. B. (1997) *Biophys. Chem.* **64,** 15–23.
58. Herrera, J. E. & Chaires, J. B. (1989) *Biochemistry* **28,** 1993–2000.
59. Chan, S. S., Breslauer, K. J., Hogan, M. E., Kessler, D. J., Austin, R. H., Ojemann, J., Passner, J. M. & Wiles, N. C. (1990) *Biochemistry* **29,** 6161–6171.
60. Park, Y.-W. & Breslauer, K. J. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 1551–1555.
61. Breslauer, K. J. (1991) *Curr. Opin. Struct. Biol.* **1,** 416–422.
62. Chan, S. S., Breslauer, K. J., Austin, R. H. & Hogan, M. E. (1993) *Biochemistry* **32,** 11776–11784.
63. Breslauer, K. J. (1987) *Methods Enzymol.* **259,** 221–241.
64. Breslauer, K. J. (1994) *Methods Mol. Biol.* **26,** 347–372.