A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics

John SantaLucia, Jr.

*PNAS* 1998;95:1460-1465
doi:10.1073/pnas.95.4.1460

This information is current as of February 2007.
A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics

JOHN SANTA LUCIA, JR.*

Department of Chemistry, Wayne State University, Detroit, MI 48202

Communicated by Ignacio Tinoco, Jr., University of California, Berkeley, CA, December 24, 1997 (received for review October 8, 1997)

ABSTRACT A unified view of polymer, dumbbell, and oligonucleotide nearest-neighbor (NN) thermodynamics is presented. DNA NN $\Delta G^\circ$ parameters from seven laboratories are presented in the same format so that careful comparisons can be made. The seven studies used data from natural polymers, synthetic polymers, oligonucleotide dumbbells, and oligonucleotide duplexes to derive NN parameters; used different methods of data analysis; used different salt concentrations; and presented the NN thermodynamics in different formats. As a result of these differences, there has been much confusion regarding the NN thermodynamics of DNA polymers and oligomers. Herein I show that six of the studies are actually in remarkable agreement with one another and explanations are provided in cases where discrepancies remain. Further, a single set of parameters, derived from 108 oligonucleotide duplexes, adequately describes polymer and oligomer thermodynamics. Empirical salt dependencies are also derived for oligonucleotides and polymers.

The application of the nearest-neighbor (NN) model to nucleic acids was pioneered by Zimm (1) and by Tinoco and coworkers (2–6). Subsequently, several experimental and theoretical papers on DNA and RNA NN thermodynamics have appeared (7–22). There has been disagreement concerning a number of issues, particularly differences between DNA polymer and oligonucleotide NN thermodynamic trends and the salt dependence of nucleic acid denaturation. These differences have led to the notion that there is a "length dependency" to DNA thermodynamics (18). In this article, I show that there is a length dependence to salt effects but not for the NN parameters. Further, a single set of parameters, derived from 108 oligonucleotide duplexes, adequately describes polymer and oligomer thermodynamic behavior. The major sources of confusion in the literature are that the different studies use different oligonucleotide and polymer design, different methods for determining thermodynamics, different methods for analyzing data, different salt conditions, and different formats for presenting the NN parameters. In this article, the results from seven studies (7, 10, 12, 18, 20, 21) are presented in the same format so that direct comparisons can be made. These data are compared with the recently compiled "unified" oligonucleotide NN parameters based on a collection of 108 oligonucleotide duplexes from the literature (22). This work emphasizes the $\Delta G^\circ_{T}$ parameters because $\Delta G^\circ_{T}$ is more accurate than $\Delta H^\circ$ or $\Delta S^\circ$ due to compensating errors (22). Remarkably, there is consensus agreement among the parameters determined from six laboratories (7, 10, 17, 18, 20, 21).

Background

The NN Model. The NN model for nucleic acids assumes that the stability of a given base pair depends on the identity and orientation of neighboring base pairs. Throughout this paper, the 10 NN dimer duplexes are represented with a slash separating strands in antiparallel orientation (e.g., AC/TG means 5’-AC-3’ Watson–Crick base-paired with 3’-TG-5’). For oligonucleotide duplexes, additional parameters for the initiation of duplex formation are introduced. Importantly, all other sequence-independent effects are also combined into the initiation parameter including differences between terminal and internal NNs (23) and counterion condensation (24, 25).

To account for differences between duplexes with terminal A-T vs. terminal G-C pairs, two initiation parameters are introduced (22, 23): "initiation with terminal G-C" and "initiation with terminal A-T". An additional entropic penalty (26) for the maintenance of the C2 symmetry of self-complementary duplexes is also included. The total $\Delta G^\circ$ is given by:

$$\Delta G^\circ = \Sigma n_i \Delta G_i + \Delta G(\text{init w/term G-C}) + \Delta G(\text{init w/term A-T}) + \Delta G(\text{sym}),$$  \(1\)

where $\Delta G_i$ are the standard free-energy changes for the 10 possible Watson–Crick NNs (e.g., $\Delta G_e^T(1)$ = $\Delta G_e^{AA}$ (TA/TT), $\Delta G_e^T(2)$ = $\Delta G_e^{TA}$ (AT), . . . etc.), $n_i$ is the number of occurrences of each nearest neighbor, $i$, and $\Delta G(\text{sym})$ equals +0.43 kcal/mol (1 cal = 4.184 J) if the duplex is self-complementary and zero if it is non-self-complementary.

Application of the Unified NN Parameters. Fig. 1 illustrates the calculation of $\Delta G^\circ$ for the sequence CGTTGA-TCAACG using the unified NN parameters (22) in Table 1. The $\Delta H^\circ$ and $\Delta S^\circ$ parameters are analogously calculated from the parameters in Table 2. The $\Delta G^\circ$ can also be calculated from $\Delta H^\circ$ and $\Delta S^\circ$ parameters by using the equation:

$$\Delta G^\circ_T = \Delta H^\circ - T \Delta S^\circ.$$

If large temperature extrapolation from 37°C is required, then the difference between the heat capacities of the folded and denatured states, $\Delta C_p$, should be accounted for (27, 28). Previous data have indicated that $\Delta C_p$ is usually small for nucleic acids (29, 30). Due to enthalpy–entropy compensation, $\Delta G^\circ_T$ is relatively insensitive to $\Delta C_p$.

Prediction of the Melting Temperature $T_M$. $T_M$ is defined as the temperature at which half of the strands are in the double-helical state and half are in the "random-coil" state. For self-complementary oligonucleotide duplexes, the $T_M$ is calculated from the predicted $\Delta H^\circ$ and $\Delta S^\circ$ and the total oligonucleotide strand concentration $C_T$, by using the equation:

$$T_M = \Delta H^\circ / (\Delta S^\circ + R \ln C_T),$$

where $R$ is the gas constant (1.987 cal/Kmol). For non-self-complementary molecules, $C_T$ in Eq. 3 is replaced by $C_T/4$ if the strands are in equal concentration or by $(C_A + C_B)/2$ if the strands are at different concentrations, where $C_A$ and $C_B$ are the

Abbreviations: SVD, singular value decomposition; NN, nearest neighbor; e.u., entropy unit (cal/Kmol).

*To whom reprint requests should be addressed. e-mail: jsl@chem.wayne.edu.
ΔG°,\text{pred.} = ΔG°(CG/GC) + ΔG°(GT/CA) + ΔG°(TT/AA)
+ ΔG°(TG/AC)+ ΔG°(GA/CT) + ΔG°(init.)
= -2.17 - 1.44 - 1.00 - 1.45 - 1.30 + 0.98 + 1.03

ΔG°,\text{pred.} = -5.35 \text{ kcal/mol}
ΔG°,\text{obs.} = -5.20 \text{ kcal/mol}

FIG. 1. Application of the unified nearest neighbor parameters (Table 1) and Eq. 1 to predict ΔG°. Each arrow points to the middle of one of the NN dimers. The duplex CGTTGA:TCAACG is non-self-complementary and thus ΔG°(sym) is zero.

concentrations of the more concentrated and less concentrated strands, respectively. Synthetic polymers with simple repeat sequences usually melt in a single cooperative transition (approximately two-state) that is concentration-independent so that T_M = \Delta H^f/\Delta S^c. Natural polymers with heterogeneous sequences, on the other hand, usually melt with many stable intermediate states (non-two-state) and accurate prediction requires a statistical mechanical partition function approach (11, 31–33).

Methods

Formats of the NN Model. Three formats for presenting NN thermodynamic data are as follows: (i) the 10 NN base pair dimer-stacking energies approach (4, 12, 13, 20, 21), (ii) the linearly independent sequences approach (also known as the “polymer approach”) (3, 10, 19), and (iii) the “independent short sequences” (ISS) approach, which presents the oligonucleotide NN data in an irreducible representation (23). All three of these methods are valid and provide equivalent predictions within round-off error (J.S. and D. M. Gray, unpublished result). In this work, the NN parameters from the literature are cast in the dimer-stacking format because it is easier for nonexperts to apply, because all 10 NN parameters are required to fully describe oligonucleotide melting, and because the ISS model does not apply to polymers.

The Rank of the Stacking Matrix. An understanding of the effects of the rank of the stacking matrix is essential to reconcile the literature NN parameters derived from polymers with those derived from oligonucleotides. The stacking matrix S has dimensions M × N for a data set of M sequences and the columns N contain the number of occurrences in each DNA sequence of the 10 NN dimers plus initiation parameters for duplexes. The column rank of the stacking matrix determines the maximum number of parameters that can be uniquely determined (invariants) (19, 34). Because of constraints on the NN composition (23), there are only eight invariants for polymers that are usually expressed as linearly independent sequences. A convenient set of eight linearly independent sequences was provided by Volodogiiskii et al. (10):

\[
\begin{align*}
P_1 &= P_{AA/TT} \\
P_2 &= P_{GG/CC} \\
P_3 &= P_{AT/TA} + P_{TA/AT} \\
P_4 &= P_{GC/CG} + P_{CG/GC} \\
P_5 &= P_{AC/TA} + P_{TC/GA} \\
P_6 &= P_{AG/TC} + P_{GA/CT} \\
P_7 &= P_{AT/TA} + P_{TC/AG} + P_{CA/GT} \\
P_8 &= P_{AG/TC} + P_{GC/CG} + P_{CA/GT}
\end{align*}
\]

poly(A)-poly(T)
poly(G)-poly(C)
poly(AT)-poly(AT)
poly(GC)-poly(GC)
poly(AC)-poly(GT)
poly(AG)-poly(CT)
poly(ATC)-poly(GAT)
poly(AGC)-poly(GCT)

In this set, P is a measurable property such as ΔG°, ΔH^f, or ΔS^c. Linear combinations of these eight invariants can be used to completely describe the behavior of any DNA polymer within the limits of the NN model. For oligonucleotide dumbbells with fixed termini but different lengths, nine invariants can be determined (18, 19). To fully characterize the thermodynamics of oligonucleotide duplexes, parameters for all 10 NN dimers (plus initiation parameters) are required. Importantly, a set of sequences with a rank of 8 (or 9 for dumbbells) can be used to derive a set of the 10 NN dimer energies that are a linear least-squares fit of the data set, but the solution is not unique. To verify that the solution is not unique, one can add a constant, C, to one of the dimers [other than AA/TT or GG/CC, which are uniquely determined for both polymers and oligomers (18)] and then add or subtract C or zero from the other dimers subject to the constraints of the eight linearly independent sequences. An alternative solution is obtained that makes exactly equal predictions as the first solution but with different trends in the NNs. The method of singular value decomposition (SVD) (20, 22, 34) provides the solution with C equal to zero and represents the minimum sequence dependence of the 10 dimers consistent with the eight invariants. Other methods for obtaining a linear least-squares fit of the data in terms of 10 parameters, particularly iterative methods (e.g., Gauss elimination or Gauss–Jordan iteration with back substitution) provide solutions with a nonzero and arbitrary value of C; the 10 dimers from these methods have an artificially larger sequence dependence than that determined by SVD. The solution obtained from these other methods, however, is a linear least-squares fit and makes predictions that are equal to those of the parameter set obtained with SVD. These points are at the heart of reconciling the NN data sets of Gotoh and Tagashira (7) and of Delcourt and Blake (17) with the oligonucleotide NN parameters (see below).

Converting Polymer Stability Temperatures to Energies. Several studies (7, 10, 17, 18) present the NN stabilities in terms of Kelvin temperatures, T(i), where i are the NN dimers, instead of free energy changes, ΔG°(i). The dimer stacking ΔH°(i) can be calculated from T(i) with Eq. 4:

\[
\Delta H^i = T(i) \times \Delta S^i
\]

The polymer studies assume that the dimer propagation ΔS°(i) is −24.85 ± 1.74 cal/K/mol for all stacking dimers and is independent of the salt concentration (17, 18). The dimer ΔG° values are then calculated with Eq 2. For example, table 2 of Delcourt and Blake (17) lists T(TA/AT) = 56.31°C, which gives ΔH°(TA/AT) = (56.31 + 273.15) × −24.85 e.u. = −8,187 cal/mol (where e.u. is entropy unit). Using Eq. 2 gives ΔG° = −8,187 cal/mol − 310.15 × −24.85 e.u. = −0.48 kcal/mol. This conversion was performed for all of the NN parameters given in Delcourt and Blake (17), Volodogiiskii et al. (10), and Gotoh and Tagashira (7). To remove the C contribution, the literature NN parameters (7, 10, 17) were used to calculate the eight invariants given above (P_1 through P_8) and then SVD was used to produce a new set of NN parameters with C equal to zero (Table 1).

The studies of Volodogiiskii et al. (10) and Dokytzej et al. (18) separated the dimer temperature stabilities into “hydrogen bonding” contributions T_HB(j) that are dependent on percent G+C content and “dimer-stacking” contributions T_SS(j) that are perturbations that contain the NN sequence dependence (these terms also contain the other fundamental interactions such as electrostatics and conformational entropy). The experimental work of Frank-Kamenetski (35) indicated that hydrogen bonding contributions are salt-dependent and calculated as follows:

\[
T_{HB}(A\cdot T) = 555.55 + 7.95 \ln [NA^+] \\
T_{HB}(G\cdot C) = 391.55 + 4.89 \ln [NA^+]
\]

Biochemistry: SantaLucia
where $T_{H}(A-T)$ and $T_{H}(G-C)$ are the hypothetical melting temperatures of isolated $A-T$ and $G-C$ pairs without any stacking interactions with the initiation parameter set to zero (polymer behavior). Thus, $T_{H}(AA/TT)$, $T_{H}(AT/TA)$, and $T_{H}(TA/AT)$ are given by Eq. 5a; $T_{H}(GG/CC)$, $T_{H}(GC/GC)$, and $T_{H}(CG/GC)$ are given by Eq. 5b; and $T_{H}(CA/CT)$, $T_{H}(AC/TG)$, and $T_{H}(AG/TC)$ are given by the average of Eqs. 5a and 5b. Eqs. 5a and 5b were used to analyze the data of Vologodskii et al. (10). Doktycz et al. (18) performed their dumbbell study in 0.115 M Na$^+$ and found experimentally that $T_{H}(A-T)$ and $T_{H}(G-C)$ were 339.67 K and 383.67 K, respectively. Table 3 of Vologodskii et al. (10) lists $T_{S}$(i), and other works (17, 18) list the stacking perturbations as $\Delta G^{i}(i)$. Eq. 6 allows calculation of stability temperatures $T(i)$ from $T_{H}(i)$ and $T_{S}(i)$ or $\Delta G^{i}(i)$.

$$T(i) = T_{H}(i) + T_{S}(i) = T_{H}(i) + [\Delta G^{i}(i)/\Delta S^{i}(0)]. \tag{6}$$

Table 2. Unified oligonucleotide $\Delta H^{i}$ and $\Delta S^{i}$ NN parameters in 1 M NaCl (22)

<table>
<thead>
<tr>
<th>Sequence</th>
<th>$\Delta H^{i}$ (kcal/mol)</th>
<th>$\Delta S^{i}$ (cal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA/TT</td>
<td>-7.9</td>
<td>-22.2</td>
</tr>
<tr>
<td>AT/TA</td>
<td>-7.2</td>
<td>-21.3</td>
</tr>
<tr>
<td>GA/GT</td>
<td>-8.5</td>
<td>-22.7</td>
</tr>
<tr>
<td>GA/CT</td>
<td>-8.2</td>
<td>-22.2</td>
</tr>
<tr>
<td>GC/GC</td>
<td>-10.6</td>
<td>-27.2</td>
</tr>
<tr>
<td>GC/CC</td>
<td>-8.0</td>
<td>-19.9</td>
</tr>
</tbody>
</table>

For example, the study of Doktycz et al. (18) reported $\Delta G^{37}(AA/TT)$ as $-196$ cal/mol. Using Eqs. 6, 4, and 2 for the AA/TT stack gives:

$$T(AA/TT) = 339.67 K - 196.0 cal/mol/ -24.8 cal/K/mol$$

$$= 347.57 K,$$

$$\Delta H^{0}(AA/TT) = 347.57 K \times -24.8 cal/K/mol$$

$$= -8,620 cal/mol,$$

and

$$\Delta G_{37}^{0}(AA/TT) = -8,620 cal/mol - 310.15 K$$

$$\times -24.8 cal/K/mol = -0.93 kcal/mol.$$

All other parameters in Table 1 in the columns for Gotoh and Tagashira (7), Vologodskii et al. (10), Delcourt and Blake (17), and Doktycz et al. (18) were calculated similarly.

Results and Discussion

Table 1 presents the NN $\Delta G_{37}$ values for helix propagation and initiation from eight experimental studies. With all these data sets presented in a uniform format, a remarkable consensus is immediately evident. The qualitative trend observed in order of decreasing stability is GC/GC > GG/CC > GA/GT = GT/GA = CA/CT = CT/GA > AA/TT > AT/TA > TA/AT. To quantify the quality of the NN parameters, linear regression analysis was performed with the literature NN as the dependent variable ($\gamma$ axis) and the unified oligonucleotide NN parameters as the independent variable ($\chi$ axis) (Table 3). The slope of this plot indicates how close the range (i.e., the difference between the largest and the smallest NN parameters) of the literature NN agrees with the range observed in the unified NN parameters. The intercept indicates the quality of the initiation.
but the polymer NN data cannot be used to reliably predict that the oligomer NN can predict polymer behavior accurately, to predict polymer thermodynamics. A general result from this is that the oligomer NN has almost no effect on the eight invariants needed for the eight invariants of each polymer NN at three salt concentrations can be individually plotted vs. \( \ln[Na^+] \) (see below); this agrees with previously observed values for oligonucleotides (29, 30). The entropy correction is given by:

\[
\Delta S^o(\text{oligomer}, [Na^+]) = \Delta S^o(\text{unified oligomer, } 1 \text{ M NaCl}) + 0.368 \times N \times \ln[Na^+].
\]  [8]

If the \( \Delta H^o \) is assumed to be salt-concentration-independent (36, 30), the \( T_M \) values of the 26 oligonucleotides are predicted by using Eqs. 8 and 3 with an average deviation of 2.2°C. The salt corrections given in Eqs. 7 and 8 can be viewed as either length-dependent corrections to the initiation parameter or as corrections to the propagation parameters because there are N NNS in an oligonucleotide duplex.

Salt Dependence of Polymers. Helix formation in polymers does not formally involve an initiation parameter so the salt dependence is by default incorporated into the NN propagation terms (35). The observation that polymer NN parameters and the unified oligonucleotide NN are highly correlated (Table 3) suggests a relationship could be determined that would allow prediction of polymer behavior from the unified NN parameters with an appropriate salt correction. The \( \Delta G^o_{37} \) differences of the three polymer NN data sets (Table 1) and the unified NN (1 M NaCl) data set were plotted vs. \( \ln[Na^+] \) of the polymer data. From a least squares fit of this plot (30 data points) and the assumption that the salt correction is sequence independent, the following empirical equation was derived:

\[
\Delta G^o_{37}(\text{polymer NN, } [Na^+]) = \Delta G^o_{37}(\text{unified NN, } 1 \text{ M NaCl}) - 0.175 \ln[Na^+] - 0.20.
\]  [9]

The standard deviations in the slope (−0.175 kcal/mol) and intercept (−0.20 kcal/mol) are 0.034 kcal/mol and 0.11 kcal/mol, respectively. Note that this correction is given in kcal/mol of base pairs. Alternatively, the \( \Delta G^o_{37} \) of each polymer NN at three salt concentrations can be individually plotted vs. \( \ln[Na^+] \) to test the sequence dependence of salt effects. Unfortunately, the \( \Delta G^o_{37}(i) \) vs. \( \ln[Na^+] \) plots for the CT/GA, CG/GC, and TA/AT neighbors show correlation coefficients \( R^2 \) that are less than 0.9. Nonetheless, for the seven NN that show a linear salt dependence of \( \Delta G^o_{37} \left( R^2 > 0.95 \right) \), it does appear that A+T-rich NNs show a larger salt dependence than the G+C-rich NNs, consistent with earlier observations (see Eq. 5) (35). Blake (37) has provided a tentative salt dependence of the 10 dimers.

Eq. 9 gives \( \Delta G^o_{37}/\ln[Na^+] = -0.175 \text{ kcal/mol, which corresponds to a } \partial T_M/\partial \ln[Na^+] \text{ of } 16.2°C \text{ (when a sequence-independent } \Delta S^o \text{ of } -24.85 \text{ e.u. is assumed), which agrees well with the widely used value for polymers of } 16.6°C \text{ (38). Fig. 2 plots the NN stabilities observed in the three polymer studies (Table 1) (7, 10, 17) versus those predicted with Eq. 9. The slope and intercept are close to 1 and 0, respectively, and the correlation coefficient } R^2 \text{ is 0.96 (see Fig. 2). The standard deviation between the experimental NN and that predicted with the unified NN by using the linear least-squares slope and intercept shown. This reflects the agreement in the NN trend. See text for comparisons among the polymer studies.}

**Table 3. Linear regression parameters for plots of literature \( \Delta G^o_{37} \) NN parameters vs. the unified NN parameters**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Gottoh (ref. 7)</th>
<th>Vologodskii (ref. 10)</th>
<th>Breslauer (ref. 12)</th>
<th>Benight (ref. 18)</th>
<th>SantaLucia (ref. 20)</th>
<th>Sugimoto (ref. 21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept, kcal/mol</td>
<td>0.94</td>
<td>0.81</td>
<td>1.41</td>
<td>0.82</td>
<td>0.87</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>0.41</td>
<td>-0.17</td>
<td>0.19</td>
<td>-0.04</td>
<td>-0.13</td>
<td>0.02</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.97</td>
<td>0.98</td>
<td>0.82</td>
<td>0.95</td>
<td>0.81</td>
<td>0.97</td>
</tr>
<tr>
<td>SD, kcal/mol</td>
<td>0.09</td>
<td>0.06</td>
<td>0.36</td>
<td>0.11</td>
<td>0.22</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Fit to the equation \( \Delta G^o_{37}(\text{unified NN}) = \Delta G^o_{37}(\text{literature NN}) \times \text{slope} + \text{intercept} \) (see text). See text for interpretation of the slope, intercept, and \( R^2 \) (correlation coefficient). SD between the experimental NN and that predicted with the unified NN by using the linear least-squares slope and intercept shown. This reflects the agreement in the NN trend. See text for comparisons among the polymer studies.
The most reliable estimate for polymers is mental actually predict the stability of polymers. A plot of the experi-
mental data is 0.12 kcal/mol. The curves were fit with the Poland partition function algorithm (32) using the
Fixman–Freire approximation for the loop functions (33) and modified to incorporate heterogeneous stacking (7). Vologodskii
et al. (10) critically evaluated the work of Gotoh and Tagashira (7) and concluded that the low salt concentration was responsible for
the observed hysteresis and suggested that this indicated non-
equilibrium conditions. This work, however, shows that with the
proper salt extrapolation, the parameters of Gotoh and Tagashira
(7) are in remarkable agreement with other polymer parameters
(10, 17) and with the unified oligonucleotide NN parameters (Fig.
2). This suggests any nonequilibrium effects in Gotoh and Ta-
gashira’s study must have been relatively small.

Analysis of Vologodskii et al. (10). Vologodskii et al. (10) derived NN parameters by using the linearly independent se-
quences approach from eight natural DNA polymer restriction
fragments dissolved in 0.195 M NaCl. The high salt concentration
used ensured equilibrium conditions throughout the melting
curve. Vologodskii’s study used a partition function approach similar to that in the study of Gotoh and Tagashira (7). Volo-
godskii et al. (10) presented their NN parameters as both eight
linearly independent sequences and nonuniqueneighbors parame-
ters by using the assumptions that AT/TA = TA/AT and GC/C
= CG/GC. Oligonucleotide experiments reveal that these assumptions are approximately correct for DNA but not for
RNA (13). The Vologodskii $\Delta G_{37}$ NN parameters show remark-
able agreement with the unified NN parameters (Tables 1 and 3).
The results presented herein verify that the experimental design and
analysis methods used in ref. 10 are fundamentally sound.

Analysis of Breslauer et al. (12). Breslauer et al. (12) derived NN thermodynamic parameters by using differential scanning
 calorimetry and UV melting analysis of 19 oligonucleotide du-
plexes (dissolved in 1 M NaCl) and nine synthetic DNA polymers
(dissolved in low salt with results extrapolated to 1 M NaCl). It is
not possible to rederive the reported parameters, however, be-
cause much of the primary thermodynamic data have not been
published. This work demonstrated good insight in that the
authors reasoned that polymer and oligomer NN trends should be
similar. However, the assumption that the initiation $\Delta G_{37}$ is 5.2
kcal/mol is most likely what led to the incorrect NN determined
(Tables 1 and 3). Breslauer’s NN predict the $\Delta G_{37}$, $\Delta F$, $\Delta S$, and
$T_M$ of the unified data set with average deviations of 16.7%,
10.1%, 10.6%, and 6.0°C, respectively. Predictions are particularly
poor for oligonucleotides shorter than 8 bp. For example, the $T_M$
of the sequence CACAGCTGTC (41) is incorrectly predicted by
31°C. Other groups have also been unable to reconcile the
Breslauer parameters with experiments (18, 20, 21, 37, 42).

Analysis of Delcourt and Blake (17). Delcourt and Blake
(17) studied 41 restriction fragments of natural polymers
dissolved in 0.075 M NaCl and expressed their NN parameters
in terms of 10 nonuniqueneighbors that make good predictions of
polymers but do not represent the real trends in NN
stability. The results presented herein verify that the experi-
mental design and analysis methods used in Delcourt and
Blake (17) are fundamentally sound.
Analysis of Doktycz et al. (18). Benight and coworkers (18) recognized that there was consensus agreement among the polymer studies and their own dumbbell studies but could not reconcile the oligomer literature parameters (12). On the basis of this problem, the authors proposed that there must be a “length dependency” to DNA NN thermodynamics. Herein I show that the NN parameters themselves are not “length-dependent” but that the salt dependence is length-dependent in ways that are still not fully understood.

Benight and coworkers (18) used UV melting analysis of 17 oligonucleotide dumbbells with 14–18 bp to determine nine linearly independent sequences that follow a NN model. The experimental design for this study precluded measurement of an initiation parameter for duplex formation and a 10th NN parameter. The analysis was performed under four salt conditions, including 25, 55, 85, and 115 mM Na⁺. These data suggested that the NN model breaks down at salt concentrations of less than 85 mM but works well at 115 mM Na⁺. It is possible that the neglect of the length dependence of salt effects (Eq. 7) is what led to the apparent breakdown of the NN model at low salt concentrations. Doktycz et al. (18) assumed that the ΔS° for helix propagation was sequence- and salt-independent (~24.8 e.u. per bp) (17), which is incorrect for oligonucleotides. With the exception of the CG/GC neighbor, the dumbbell ΔG°NN parameters in 0.115 M Na⁺ show good agreement with the unified NN parameters (Table 1).

Analysis of SantaLucia et al. (20). SantaLucia et al. (20) derived NN parameters from thermodynamics determined by a van’t Hoff analysis of UV melting data for 23 oligonucleotides combined with calorimetric or UV melting results from the literature for 21 other sequences. To minimize “fraying artifacts,” all sequences included in the linear regression analysis to determine NN parameters had terminal G:C pairs (12). The SantaLucia parameters are within experimental error of the unified parameters in Table 1.

Data were available in the literature for eight sequences with terminal T:A pairs. These data were included in the first fit of the NN parameters and the stacking matrix was not rank deficient but nonphysical results were obtained for six of the nearest neighbors (AT/TA, TA/AT, CA/CT, AC/TA, GA/CT, and AG/TC) (20). For example, the ΔH° parameters for AT/TA and TA/AT neighbors were found to be -10.80 and -10.80 kcal/mol, respectively, which is unlikely (J.S., unpublished results). We now know that four of the sequences with terminal T:A pairs exhibited non-two-state behavior (20, 21). Upon removal of the sequences with terminal T:A pairs, more reasonable results were obtained, but the rank of the stacking matrix was reduced to 10 (nine linearly independent sequences plus one initiation parameter) (23).

Analysis of Sugimoto et al. (21). Sugimoto et al. (21) derived NN parameters from thermodynamics determined by a van’t Hoff analysis of UV melting data for 50 oligonucleotides combined with data for 15 sequences from other laboratories obtained by both calorimetry and UV melting. Except for the initiation ΔG°NN, the CG/GC ΔG°NN, and the GG/CC ΔH°, the Sugimoto parameters (21) are in good agreement with the unified NN parameters (Tables 1 and 3). With the proper linear regression analysis, Sugimoto’s data set provides NN and initiation parameters that are in excellent agreement with the unified parameters (22). Important results of this work are that separate parameters for terminal T:A base pairs and for initiation at A:T are not required.

Conclusion
A unified set of NN parameters is now available for making accurate predictions of DNA oligonucleotide, dumbbell, and polymer thermodynamics. The agreement among the various polymer and oligomer studies provides a great deal of confidence in their reliability.

I thank Douglas H. Turner and Hatim Allawi for stimulating conversations and for critical reading of the manuscript. I also thank Wayne State University and Hitachi Chemical Research for financial support.