Heat Capacity Effects on the Melting of DNA. 2. Analysis of Nearest-Neighbor Base Pair Effects

Ioulia Rouzina and Victor A. Bloomfield
Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, St. Paul, Minnesota 55108 USA

ABSTRACT The stability of a DNA double helix of any particular sequence is conventionally estimated as the average of the stabilities of the 10 different nearest-neighbor (NN) base pair doublets that it contains. Therefore, much effort has been devoted to the experimental characterization and tabulation of the enthalpy, entropy, and free energy of melting for each of the NN doublets. Although data from different research groups generally agree for the NN free energies and melting temperatures, there are major disagreements for the enthalpies and entropies. The largest differences are between the parameters obtained on oligomeric relative to polymeric DNA. This disagreement interferes with the practical application of NN thermodynamic parameters. It also raises doubts regarding several fundamental assumptions about DNA melting, such as the absence of longer range interactions, the length dependence of DNA melting parameters per base pair, the applicability of polyelectrolyte theory to the description of salt effects on oligomers, and the purely enthalpic difference between NN doublets. Here we show that if one takes into account the significant heat capacity increase associated with DNA melting, all of the above assumptions are self-consistently reconciled with experiment.

INTRODUCTION

It has been known for some time that the main effect of DNA composition on double helix stability is due to the large enthalpic difference (~1000 cal/mol) between the GC and AT base pairs, coming from the extra hydrogen bond in the former. This major energetic difference, however, cannot account for the fine details of DNA differential melting profiles. The next obvious step is to take into account energetic differences between nearest-neighbor (NN) doublets of base pairs, which should arise from different stacking interactions. For DNA there are 10 possible different NN doublets, so that 10 sets of ΔH_{ij} and ΔS_{ij} parameters should fully characterize the stability of a particular DNA molecule. The additivity principle, when applied to the NN problem, means that each thermodynamic quantity of a DNA molecule is a linear combination of the corresponding NN quantities, weighted by the frequencies f_{ij} of the corresponding NN doublets:

$$\Delta H = \sum f_{ij}\Delta H_{ij}, \quad \Delta S = \sum f_{ij}\Delta S_{ij}, \quad \Delta G = \sum f_{ij}\Delta G_{ij}. \qquad (1)$$

Given the fundamental and practical importance of understanding DNA stability, experimental characterization of ΔH_{ij} and ΔS_{ij} has been attempted by many groups over the last 20 years. Measurements of ΔH and ΔS were made by both calorimetry and van't Hoff analysis for a large number of DNA sequences, and the results were fitted to a set of 10 NN parameters as in Eq. 1. Such studies were performed on high polymeric DNA molecules (Blake and Delcourt, 1998; Gotoh and Tagashira, 1981; Vologodskii et al., 1984), oligomers (Doktycz et al., 1992; Santalucia, 1998; Sugimoto et al., 1996), and combinations of the two (Breslauer et al., 1986). As recently summarized by Santalucia (1998), there is a good consensus on ΔG_{ij} values between most of the studies after conversion to the same temperature, 37°C = 310 K, and ionic strength. This is not true, however, for the individual ΔH_{ij} and ΔS_{ij} components of ΔG_{ij}, as seen in Fig. 1.

The fitted ΔH_{ij} and ΔS_{ij} parameters vary substantially from one study to another, the main difference being between their oligomeric and polymeric values. For example, the polymeric transition entropy appears to be independent of the NN identity and equals $\Delta S = 24.85 \pm 1.74$ cal/mol K in $I = 0.075$ M NaCl (Delcourt and Blake, 1991), while the oligomeric ΔS_{ij} varies between 19.9 and 27.2 cal/mol K, with an average of 22.4 cal/mol K (Santalucia, 1998). Another important feature is the differential stabilization effect of salt on various base pair doublets, which also appears to depend on DNA length.

This raises doubts regarding many conventional assumptions about DNA melting, such as 1) additivity of NN thermodynamic parameters, i.e., absence of longer range interactions; 2) length independence of DNA melting parameters per base pair; 3) applicability of polyelectrolyte theory in describing salt effects; 4) purely enthalpic difference between NN doublets; and 5) zero heat capacity changes associated with DNA melting. This last assumption is always implicitly made in the data analysis but is rarely discussed. Here we show that if one does not assume constancy of ΔC_p in the melting transition, then the other assumptions can be reconciled with experiment.

RESULTS

The key point is that there is always, in some sense, a long-range interaction between the base pairs. The transi-

Received for publication 21 may 1999 and in final form 17 August 1999.
Address reprint requests to Dr. Victor A. Bloomfield, Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, 1479 Gortner Avenue, St. Paul, MN 55108. Tel.: 612-625-2268; Fax: 612-625-5780; E-mail: victor.a.bloomfield-1@tc.umn.edu.

0006-3495/99/12/3252/04 $2.00

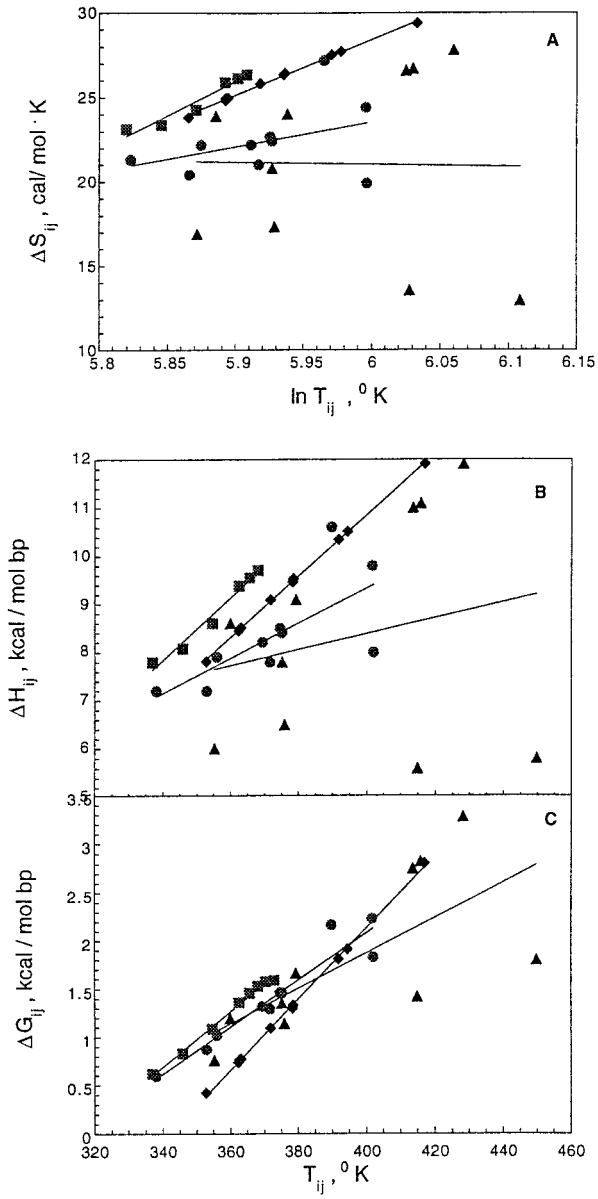

FIGURE 1 Correlation between thermodynamic melting parameters for the nearest-neighbor doublets (NN) and their melting temperatures T_{ij}. ◆, Polymeric DNA data (Blake and Delcourt, 1998) obtained by van't Hoff analysis of the optical melting profiles. ●, Oligomeric DNA data from Santalucia (1998) obtained by analysis of T_m dependence versus oligomer strand concentration. ▲, A combination of oligomeric and polymeric DNAs measured by many different methods (Breslauer et al., 1986). ■, Data for polymeric calf thymus DNA (Gruenwedel, 1974), the melting temperature of which was varied by changing solution ionic strength. In this case the slopes reflect the actual heat capacity change in the course of DNA melting transition, whereas this relationship is indirect for the NN data.

tion enthalpy and entropy of a DNA fragment are always measured at the particular melting temperature T_m of this fragment, which is determined by its whole sequence. Therefore, the derived ΔH_{ij} and ΔS_{ij} for NN doublets will have thermal contributions dependent on the sequence of the entire fragment. Reported ΔH_{ij} and ΔS_{ij} parameters are obtained by averaging thermodynamic data on a set of DNA fragments. It is this averaging that introduces the apparent differences between the polymer and oligomer DNA melting thermodynamics.

To a very good first approximation, the T_m of a DNA fragment is its average over the T_m's of the NN doublets:

$$T_m = \sum f_{ij} T_{ij}. \tag{2}$$

(This relation is strictly true only if ΔS is independent of nearest-neighbor composition. While this has to be assumed when complex optical melting profiles are analyzed, its universal validity has not been demonstrated. However, our analysis herein shows that the assumption is at least self-consistent.) The related thermal contribution to the individual ΔH_{ij} parameter is then $f_{ij}\Delta C_p(T_{ij} - T^0)$, where T^0 is the standard melting temperature (Rouzina and Bloomfield, 1999). With $\langle f_{ij}\rangle \approx 0.1$ for a random mixture of nearest neighbors, $\Delta C_p \approx 100$ cal/mol K, and $(T_{ij} - T^0) \approx 10$ K, this thermal contribution amounts to ~100 cal/mol.

On the other hand, melting of polymeric DNA of a given base composition proceeds via cooperative unwinding of large (~200 bp) domains enriched with NN doublets of a particular composition and sequence, the melting temperature T_{ij} of which strongly correlates with the melting temperature of the whole fragment. In the limiting case of melting of pure fragments containing only one kind of NN, this correlation would be at maximum; i.e., $\langle f_{ij}\rangle = 1$, $T_{ij} = T_m$, and the thermal contribution to ΔH_{ij} is the largest possible: $\Delta C_p(T_{ij} - T^0)$. For polymer domains of heterogeneous sequence this correlation is weaker, lying between the random composition case and the pure NN, such that

$$0.1 \lesssim \eta \leq 1, \tag{3}$$

where $\eta = \langle f_{ij}\rangle$. The corresponding thermal contribution to NN enthalpy of polymeric DNA is $\eta\Delta C_p(T_{ij} - T^0)$.

We will assume that the differences between various NN doublets are purely enthalpic, so that

$$\delta H_{ij} = \Delta S(T_{ij} - T^0), \quad \delta S_{ij} = \Delta S = \Delta S^0 + \delta S. \tag{4}$$

Here δS is the NN-type independent entropy change, such as arises from polyelectrolyte effects when the salt concentration is less than 1 M; it is small compared to ΔS^0 in most situations (Rouzina and Bloomfield, 1999). Thus for $I = 0.075$ M salt, in which most of experiments analyzed by Blake and Delcourt (1998) were done, $\delta S = 1.24$ cal/mol K. Then the total enthalpy and entropy of melting for a particular NN doublet melting are

$$\begin{aligned}\Delta H_{ij} &= \Delta H^0 + \delta H_{ij} + \eta\Delta C_p(T_{ij} - T^0)\\ &= -T^0(\delta S + \eta\Delta C_p) + T_{ij}(\Delta S^0 + \delta S + \eta\Delta C_p),\end{aligned} \tag{5}$$

where the second equality follows from $\Delta G_{ij} = 0$ at T_{ij}, and

$$\Delta S_{ij} = \Delta S^0 + \delta S + \eta\Delta C_p \ln(T_{ij}/T^0). \tag{6}$$

These expressions predict much stronger variation of the NN parameters with T_{ij} than could be expected without the heat capacity change. That is, if $\Delta C_p = 0$,

$$\Delta H_{ij} = -T^0\delta S + T_{ij}(\Delta S^0 + \delta S) \approx 25T_{ij} \text{ cal/mol}, \quad (7)$$

and

$$\Delta S_{ij} = \Delta S^0 + \delta S \approx 25 \text{ cal/mol K}. \quad (8)$$

Experimental data for three sets of NN parameters as functions of melting temperature T_{ij} are presented in Fig. 1, *A* and *B*. They can be approximated with the linear functions in Table 1. There is an obvious trend in the first three data sets, which summarize the effects of sequence. While the experimental ranges of variation of ΔH_{ij} and ΔS_{ij} among the NN doublets are of similar magnitude for all data sets, the correlations with the melting temperatures T_{ij} are progressively weaker from polymer through oligomer to mixed polymer-oligomer. The inconsistent signs for the polymer-oligomer parameters are artifacts arising from the poor fit of widely scattered data to linear functions. This poor fit may simply reflect the lack of computational coupling of ΔH and ΔS in the analysis of the calorimetric data as opposed to the computationally coupled, van't Hoff analysis of the optical data. For discussions of this point see Krug et al. (1976), Owczarzy et al. (1997), and Plum et al. (1995).

At the same time, the melting temperatures of the NN doublets and the free energies at 37°C are quite similar for all data sets (Fig. 1 *C*). This means that even when variations in ΔH_{ij} and ΔS_{ij} are not correlated with DNA sequence, they are still correlated with each other because of thermal contributions of $\sim\Delta C_p(T_m - T)$ (Fig. 2). This implies that the shift of T_m of a fragment is caused in significant part by factors other than change in DNA sequence, such as some variation in solution conditions. For oligomers, such variation could be due to end effects, i.e., energies of ~1 kcal/mol bp associated with helix initiation (Santalucia, 1998) but unrelated to average DNA content.

When the linear relations in the first three entries in Table 1 are fitted to Eqs. 5 and 6, one obtains self-consistent values with fewer parameters than the constraints in the fit;

TABLE 1 Compilation of enthalpies and entropies as functions of temperature for various DNAs, with correlation coefficient *N*

Source	ΔH, ΔS	R	Reference
Polymeric DNA	$\Delta H_{ij} = -9516 + 52.9T_{ij}$	1.000	a
	$\Delta S_{ij} = -127.3 + 26.1 \ln T_{ij}$	0.999	
Oligomeric DNA	$\Delta H_{ij} = -5255.5 + 36.5T_{ij}$	0.700	b
	$\Delta S_{ij} = -63.6 + 14.5 \ln T_{ij}$	0.380	
Poly and oligo	$\Delta H_{ij} = 1781.0 + 16.5T_{ij}$	0.220	c
	$\Delta S_{ij} = 29.6 - 1.43 \ln T_{ij}$	0.020	
Calf thymus	$\Delta H_m = -14340 + 65.3T_m$	0.986	d
	$\Delta S_m = -212.3 + 40.4 \ln T_m$	0.967	

a, Blake and Delcourt (1998).
b, Santalucia (1998).
c, Breslauer et al. (1986).
d, Gruenwedel (1974).

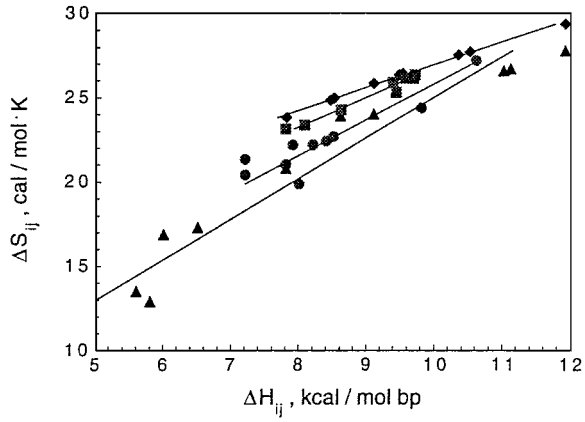

FIGURE 2 Entropy-enthalpy correlation of NN doublet melting from all data sets in Fig. 1 (the symbols have the same meanings). Although ΔH and ΔS each have a range of variation of 100% or more, they are always strongly coupled, supporting the thermal origin of these variations.

it is clear that the last data set is close to the situation with $\Delta C_p = 0$, described by Eqs. 7 and 8. The negative $\eta\Delta C_p$ value and very large T^0 are artifacts produced by the very poor linear fit to highly scattered data.

For comparison we also present in Fig. 1 and Table 1 data from a calorimetric study of calf thymus DNA (Gruenwedel, 1974) with T_m varied by salt. We see that variation of the thermodynamic parameters with temperature is even stronger than for polymeric DNAs of different base composition, while the quality of the fits to linear functions is very high. The reason for this is that all of the variation in ΔH and ΔS with temperature comes from the heat capacity contribution (Rouzina and Bloomfield, 1999):

$$\Delta H = \Delta H^0(1 + \gamma_H) + \Delta C_p(T_m - T^0), \quad (9)$$

and

$$\Delta S = \Delta S^0(1 + \gamma_H) + (\Delta C_p - \Delta S^0)\ln(T_m/T^0). \quad (10)$$

In Eq. 10 we used Eqs. 8–11 of the accompanying manuscript (Rouzina and Bloomfield, 1999) to relate the relative entropy variation γ_S to the relative melting temperature shift $\gamma_S = \gamma_H - t_m$ for the value of γ_H appropriate to the fixed DNA composition. No averaging over DNA composition was involved. Therefore the actual heat capacity contribution ΔC_p was obtained directly. Comparing Eqs. 9 and 10 to the experimental values in Table 1, we arrive at the values

$$\Delta C_p = 65.3 \text{ cal/mol K}, \quad \Delta S^0 = 24.9 \text{ cal/mol K},$$
$$T^0 = 345 \text{ K}, \quad (11)$$
$$\Delta H^0 = 7949 \text{ cal/mol}, \quad \delta H = 400 \text{ cal/mol}.$$

The value for δH agrees with the 50% CG content of calf thymus DNA (Rouzina and Bloomfield, 1999).

The apparently similar strong dependence of the melting enthalpy on temperature in the cases of NN type or salt variation has two quite different origins. In the case of composition variation, enthalpy varies both directly and

TABLE 2 Fitting parameters from Eqs. 5 and 6 for the data in Table 1

Source	$\eta\Delta C_p$ (cal/mol K)	T^0 (K)	$\Delta S^0 + \delta S$ (cal/mol K)
Polymeric DNA	26.1	344	25 ± 1.24
Oligomeric DNA	14.5	362	22
Poly and oligo	−1.43	1245	18

because of the indirect, statistically averaged, thermal contribution, i.e.,

$$\frac{\partial \Delta H_{ij}}{\partial T_{ij}} = \Delta S + \eta \Delta C_p , \tag{12}$$

while in the case of salt variation and fixed DNA composition all of the enthalpy variation comes from the thermal contribution:

$$\frac{\partial \Delta H}{\partial T_m} = \Delta C_p . \tag{13}$$

Assuming that the value of ΔC_p is 65 cal/mol K, we can estimate that the correlation factor $\eta = 0.40$, 0.22, and 0 from top to bottom in Table 2. Thus, as expected, the polymeric data show the largest thermal contribution to thermodynamic parameters, determined by DNA sequence. The two other data sets had progressively less correlation of the thermal contribution with the sequence.

Finally, we want to comment on the ratio

$$2\frac{k_B T_m^2}{\Delta H} = \frac{2}{\Delta n}\frac{\partial T_m}{\partial \ln I}. \tag{14}$$

which is frequently claimed to be a universal constant (Blake and Delcourt, 1998; Privalov et al., 1969; Record et al., 1978). Equation 14 is a general thermodynamic relation, which follows from Eqs. 19 and 35 of the accompanying paper (Rouzina and Bloomfield, 1999). Then using Eq. 34 of that paper, one obtains

$$2\frac{k_B T_m^2}{\Delta H} = 2\frac{k_B T^0}{\Delta S^0}\left[1 - \gamma_H^0 (X_{GC} - 1)(\alpha - 1)\right]. \tag{15}$$

This ratio is independent of salt concentration, and for DNA of base composition $X_{GC} = 0.5$ and $\alpha = 2.6$ it equals ~60. The ratio does not have to be the same for DNAs of different compositions. Nevertheless, in the particular case of the NN parameters given by Blake and Delcourt (1998), we see that with the substitution $\alpha \rightarrow \eta\alpha = \eta\Delta C_p/\Delta S^0 = 26/25 \approx 1$, the coefficient of X_{GC} in Eq. 15 is essentially zero, so the ratio is nearly composition-independent with a value of 56, which agrees well with the measured value of 55 ± 2 (Blake and Delcourt, 1998). Once again, there is no need to invoke a variable number of counterions bound, Δn, to different DNA sequences to explain this observation.

We are grateful to Prof. Kenneth Breslauer for helpful comments.

This research was supported in part by National Institutes of Health research grant GM28093.

REFERENCES

Blake, R. D., and S. G. Delcourt. 1998. Thermal stability of DNA. *Nucleic Acids Res.* 26:3323–3332.

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

Delcourt, S. G., and R. D. Blake. 1991. Stacking energies in DNA. *J. Biol. Chem.* 266:15160–15169.

Doktycz, M. J., R. F. Goldstein, T. M. Paner, F. J. Gallo, and A. S. Benight. 1992. Studies of DNA dumbbells. I. Melting curves of 17 DNA dumbbells with different duplex stem sequences linked by T4 endloops: evaluation of the nearest-neighbor stacking interactions in DNA. *Biopolymers.* 32:849–864.

Gotoh, O., and Y. Tagashira. 1981. Stabilities of nearest-neighbor doublets in double-helical DNA determined by fitting calculated melting profiles to observed profiles. *Biopolymers.* 20:1033–1042.

Gruenwedel, D. W. 1974. III. A calorimetric investigation of the transition enthalpy of calf thymus DNA in Na_2SO_4 solutions of varying ionic strength. *Biochim. Biophys. Acta.* 340:16–30.

Krug, R. R., W. G. Hunter, and R. A. Grieger. 1976. Enthalpy-entropy compensation. I. Some fundamental statistical problems associated with the analysis of van't Hoff and Arrhenius data. *J. Phys. Chem.* 80: 2335–2341.

Owczarzy, R., P. M. Vallone, F. J. Gallo, T. M. Paner, M. J. Lane, and A. S. Benight. 1997. Predicting sequence-dependent melting stability of short duplex DNA oligomers. *Biopolymers.* 44:217–239.

Plum, G. E., A. P. Grollman, F. Johnson, and K. J. Breslauer. 1995. Influence of the oxidatively damaged adduct 8-oxodeoxyguanosine on the conformation, energetics, and thermodynamic stability of a DNA duplex. *Biochemistry.* 34:16148–16160.

Privalov, P. L., O. B. Ptitsyn, and T. M. Birshtein. 1969. Determination of stability of the DNA double helix in an aqueous medium. *Biopolymers.* 8:559–571.

Record, M. T., Jr., C. F. Anderson, and T. M. Lohman. 1978. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Q. Rev. Biophys.* 11:103–178.

Rouzina, I., and V. A. Bloomfield. 1999. Heat capacity effects on the melting of DNA. 1. General aspects. *Biophys. J.* 77:000–000.

Santalucia, J. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA.* 95:1460–1465.

Sugimoto, N., S. Nakano, M. Yoneyama, and K. Honda. 1996. Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucleic Acids Res.* 24:4501–4505.

Vologodskii, A. V., B. R. Amirikyan, Y. L. Lyubchenko, and M. D. Frank-Kamenetskii. 1984. Allowance for heterogeneous stacking in the DNA helix-coil transition theory. *J. Biomol. Struct. Dyn.* 2:131–148.