

# A Unified Poland-Scheraga Model of Oligo- and Polynucleotide DNA Melting: Salt Effects and Predictive Power

Daniel Jost\* and Ralf Everaers

Laboratoire de Physique and Centre Blaise Pascal of the École Normale Supérieure de Lyon, Université de Lyon, CNRS UMR 5672, Lyon, France

**ABSTRACT** Key biological and nano-technological processes require the partial or complete association and dissociation of complementary DNA strands. We present a variant of the Poland-Scheraga model for DNA melting where we introduce a local, sequence-dependent salt correction of the nearest-neighbor parameters. Furthermore, our formulation accounts for capping and interfacial energies of helical and coiled chain sections. We show that the model reproduces experimental data for melting temperatures over the full experimental range of strand length, strand concentration, and ionic strength of the solution. In particular, we reproduce a phenomenological relation by Frank-Kamenetskii for very long chains using a parameterization based on melting curves for short oligomers. However, we also show that the parameters of the Poland-Scheraga model are still not known with sufficient precision to quantitatively predict the fine structure of melting curves. This formulation of the Poland-Scheraga model opens the possibility to overcome this limitation by optimizing parameters with respect to an extended base of experimental data for short-, medium-, and long-chain melting. We argue that the often-discarded melting data for longer oligomers exhibiting non-two-state transitions could play a particularly important role.

## INTRODUCTION

A quantitative understanding of basepairing and opening of the double-helix in DNA or RNA strands is relevant for many fundamental biological processes like transcription, replication (1), or RNA folding (2) and interference (3), as well as bio- and nanotechnological applications like DNA chips (4), DNA self-assembly (5) or guided nano-assembly of colloidal nanoparticles with DNA linkers (6,7). There are two standard theoretical descriptions of DNA thermal denaturation: the nearest-neighbor (NN) model quantitatively describes the melting of short oligonucleotides (8–10), which exhibit a two-state transition from a fully paired, double-helical complex to two separated single strands with random coil conformations; and the Poland-Scheraga (PS) model of polynucleotide melting (11,12) describes longer chains on the secondary structure level as an alternating sequence of double-stranded (helical) parts and (coiled) loops. The NN- and the PS-model have been employed routinely for several decades to describe a large variety of DNA- and RNA melting experiments, and the same formalism is widely used to investigate RNA folding (13,14). A possible point of criticism is the large number of adjustable parameters, which makes it (too) easy to fit individual sets of experiments. Not surprisingly, much effort has been devoted to the comparison of parameters extracted from different systems. Ten years ago, SantaLucia concluded, “A unified set of NN-parameters is now available for making accurate predictions of DNA oligo- and polymer thermodynamics” (10).

However, the compilation of results in SantaLucia (10) also shows a large number of unrelated boundary terms and

salt corrections to be used in different limits. Furthermore, the available descriptions were limited to two-state melting transitions of short oligomers and to stepwise melting of long polymers, i.e., to the limit where DNA denaturation becomes independent of the DNA concentration in the sample. The situation is more complicated for longer oligomers exhibiting partial internal melting before strand separation. Recent experiments of Zeng et al. (15) and Zeng and Zocchi (16) have shed some doubts on the applicability of the standard formalism in this biologically (3,17) and technologically (4,5) important limit.

In this article, we present a unified PS like model covering the full experimental range of chain lengths, strand, and salt concentrations and compare its predictions to available experimental data. In the first section, we augment a recent, unified formulation for oligomer and polymer melting (18) by a systematic extension of oligomer salt corrections (19) to the longer chains limit. Furthermore, we pay particular attention to the estimation of confidence limits on model predictions, illustrated here for a parameterization of the model, which is essentially based on published data for oligomer melting. In a second section, we discuss the generic behavior of the model, i.e., the influence of variations of strands and ionic concentrations on the result of melting experiments for DNA duplexes of different length. The following quantitative comparison to experiment data shows that the model reproduces experimental data from short oligomers to long DNA polymers. This includes the controversial non-two-state melting of long oligomers where, however, the uncertainty of the predictions becomes particularly large. We conclude that experiments along the lines of Zeng et al. (15) and Zeng and Zocchi (16) could play an important role in future attempts to improve the parameterization.

---

Submitted March 25, 2008, and accepted for publication August 5, 2008.

\*Correspondence: [daniel.jost@ens-lyon.fr](mailto:daniel.jost@ens-lyon.fr)

Editor: Kathleen B. Hall.

© 2009 by the Biophysical Society  
0006-3495/09/02/1056/12 \$2.00

---

doi: 10.1529/biophysj.108.134031

## MODEL AND METHODS

### Association equilibria and internal melting

We treat a complex  $AB$  in equilibrium with two strands  $A$  and  $B$ , each consisting of  $N$  basepairs (bps). For the association equilibrium  $A + B \leftrightarrow AB$  between bound and unbound states, we can write the law of mass action

$$\frac{(c_A/c_0)(c_B/c_0)}{c_{AB}/c_0} = \exp\left(\frac{\Delta G_0}{k_B T}\right), \quad (1)$$

where  $c_0$  is a reference concentration (usually  $c_0 = 1$  M),  $c_A$ ,  $c_B$ , and  $c_{AB}$  are, respectively, the concentration of  $A$ ,  $B$ , and  $AB$ , and

$$\Delta G_0 = \Delta G_{\text{int}} + \Delta G_{\text{mix}}^0(N) \quad (2)$$

is the Gibbs free energy difference between the bound and unbound forms at the reference concentration.  $\Delta G_0$  depends on both, the internal free energy difference,  $\Delta G_{\text{int}} = G_{AB}^{\text{int}} - G_A^{\text{int}} - G_B^{\text{int}}$ , and the difference of the mixing entropies with the solvent  $\Delta G_{\text{mix}}^0$ . For chains of length  $N$  the latter can be estimated as  $\Delta G_{\text{mix}}^0 = -k_B T \log(0.44(N-1))$  (see (18) and Appendix A of this article).

From Eq. 1 and using the definitions of the total, strand, and complex concentrations which are related via  $c_T = c_A(T) + c_B(T) + 2c_{AB}(T)$ ,  $c_A(T) = c_B(T) = (1 - \Theta_{\text{ext}}(T))c_T/2$  and  $c_{AB}(T) = \Theta_{\text{ext}}(T)c_T/2$ , we can derive the degree of association  $\Theta_{\text{ext}}(c_T, T)$  (20)

$$\Theta_{\text{ext}}(x) = 1 + x - \sqrt{x(2+x)}, \quad (3)$$

with  $x = \frac{c_0}{\alpha c_T} \exp\left(\frac{\Delta G_0}{k_B T}\right)$  ( $\alpha = 4$  for self-similarity strand  $A = B$ , otherwise  $\alpha = 1$ ).

Internal hybridization of the complex and the individual strands can be described in terms of the fraction  $\Theta_{\text{int}}(T)$  of bound basepair steps as a function of temperature. Here we assume that only the complex can form double-helical sections, hence, the experimentally observable overall fraction of bound basepair steps (see note below Eq. 4) is given by

$$\Theta(c_T, T) = \Theta_{\text{int}}(T)\Theta_{\text{ext}}(c_T, T). \quad (4)$$

(Note that experimental data used in this article for  $\Theta$  are UV absorbance data. Typically, optical absorbance curves exhibit a linear increase in the pre- and post-transition regimes due, respectively, to a slight elevation in the average stacking of the double-stranded DNA and to the unstacking of bases in the single strands (21). The  $\Theta$ -curves are obtained from a normalization of the absorbance data (22) that deletes these linear contributions.)

### Unified Poland-Scheraga model

To proceed, we need to determine the free energy  $\Delta G_{\text{int}}$ , as well as the degree of pairing  $\Theta_{\text{int}}(T)$  from a statistical mechanical description of single- and double-stranded DNA. In this article, we employ a recent formulation (18) of the PS model with a direct mapping to a lattice model.

The PS model describes DNA at the secondary structure level as a sequence of double-stranded sections and denatured loop- or end-domains. We write the association free energy in the double-stranded regions in the so-called ‘‘doublet format’’ (21) as a sum over sequence-dependent nearest-neighbor pair formation free energies  $\Delta g_{\text{NN}}(T) = \Delta h_{\text{NN}} - T\Delta s_{\text{NN}}$  for the 10 different basepair steps. Furthermore, our description includes boundary terms. We consider two different capping energies  $\omega_{AT}$  and  $\omega_{GC}$  for double-stranded chain ends with  $\omega(T) \equiv \Delta h_{\omega} - T\Delta s_{\omega}$  as well as cooperativity factors  $\sigma$  and  $\bar{\sigma}$  suppressing the opening of loop and end domains respectively.

In Fig. 1, we show the contributions of the different free energies to the total partition function of the system. The reference state is the helical state ( $Z_{\text{helix}} = 1$ ). Therefore, the contribution of an internal loop (size  $n$ ) is  $Z_{\text{loop}} = \sigma n^{-c} \exp(\beta \sum \Delta g_{\text{NN}})$  and the contribution of a free end (size  $\bar{n}$ ) is  $Z_{\text{end}} = \bar{\sigma} \bar{n}^c \exp(\beta \sum \Delta g_{\text{NN}} + \beta \omega)$ . Here,  $n^{-c}$  and  $\bar{n}^c$  account for, respectively, the number of self-avoiding polygons of length  $2n$  and the number of

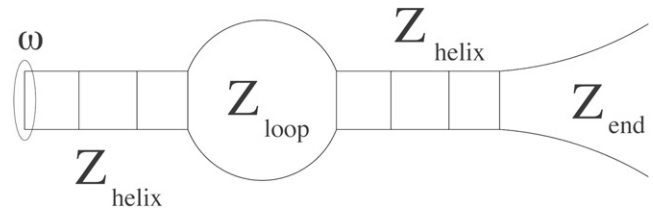


FIGURE 1 Example of a secondary structure, definition of the different contributions in the partition function.  $Z_{\text{helix}}$  is the contribution of an helical stem,  $Z_{\text{loop}}$  is the contribution of an internal loop of size  $n$ , and  $Z_{\text{end}}$  is the contribution for a free end of size  $\bar{n}$ . The term  $\omega$  refers to the capping free energy. The total partition function is the product  $Z = \exp(-\beta\omega)Z_{\text{helix}}Z_{\text{loop}}Z_{\text{helix}}Z_{\text{end}}$ .

possible conformations for a free end of length  $2\bar{n}$ . The values  $c$  and  $c'$  take into account the steric interactions in the loops and in the free ends. The values of  $c$  and  $c'$  have to be derived from polymer theory. The parameter  $c$  has been extensively discussed (11,22–25) and is equal to 1.764 for noninteracting self-avoiding loops and 2.15 for interacting self-avoiding loops. The value of  $c'$  is equal to 0.16 (26,27). The total partition function is the product of the different partial partition functions. Here we only consider the formation of native contacts in double-stranded DNA, but the general case can be discussed in the same framework (28,29). Furthermore, it is possible to treat secondary-structure formation in DNA or RNA single strands using the same formalism (13,14).

The PS model can be solved using dynamic algorithms based on recursion relations (11,12,28,29). In Appendix B, we show how the method proposed by Garel and Orland (29) for the calculation of partition functions and free energies can be adapted to the doublet format. Use of the Fixman-Freire algorithm (29,30) substantially accelerated the calculations.

### Parameters and confidence limits

Despite the large number of adjustable parameters ( $10 \Delta h_{\text{NN}}$ ,  $10 \Delta s_{\text{NN}}$ ,  $2 \Delta h_{\omega}$ ,  $2 \Delta s_{\omega}$ ,  $\sigma$ , and  $\bar{\sigma}$ ), the PS model clearly represents a drastic simplification of the true problem. The key assumption is that all nongeneric contributions to the melting free energy difference can be written as a sum of independent basepairs contributions. Furthermore, it is common to only partially consider the sequence dependence of the boundary terms. For example, since the  $5'$  and  $3'$  of the sugar-phosphate backbone are chemically different (with correspondingly modified solvation and stacking effects, etc.) there is no reason to expect identical free energy penalties for  $5' - A/3' - T$  and  $5' - T/3' - A$  ends as well as for  $5' - G/3' - C$  and  $5' - C/3' - G$  ends. Similarly, one would, in general, expect sequence- and salt-dependent forking energies; yet the available experimental data does not allow us to determine these values reliably. In general, the basic assumptions and ad hoc choices can only be corrected or justified a posteriori by evaluating the success of the model in describing (and even more importantly) predicting the results of experiments (10,21,31). Some care has thus to be taken in estimating parameter values and the associated confidence limits from comparisons to experimental data. In this article we use Monte Carlo methods (32) to account for the propagation of these errors in the calculation of model predictions, i.e., results are calculated for (and averaged over) an ensemble of models with parameters drawn from a multidimensional Gaussian distribution defined by the mean values and the covariance matrix resulting from the parameter fits.

### Nearest-neighbor parameters

Short oligomers exhibiting a two-state melting transition can be used to determine the association and capping free energies. In the NN-model,  $\Delta G_0 = \sum \Delta g_{\text{NN}} + 2\Delta g_{\text{ini}}$ , where  $\Delta g_{\text{NN}}$  and  $\Delta g_{\text{ini}}$  are pair formation and initiation NN-parameters (10). In the unified PS-model,  $\Delta G_0 = \sum \Delta g_{\text{NN}} + 2\omega + \Delta G_{\text{mix}}^0$ . Equating the two expressions yields  $\omega(T) = \Delta h_{\text{ini}} - T(\Delta s_{\text{ini}} - \Delta S_{\text{mix}}^0/2)$ , where  $\Delta S_{\text{mix}}^0 = 1.5 k_B$  accounts for the mixing entropy of oligomers at a typical size of 10 bps used in experiments employed for parameterizing the NN-model (an arbitrary standard deviation of 30% is assumed

**TABLE 1** Unified PS parameters and their standard deviations in a 1 M [Na<sup>+</sup>]-buffer

Sequence	$\Delta h$ (kcal/mol)	$\Delta s$ (cal/mol/K)
AA/TT	$-7.93 \pm 0.31$	$-22.4 \pm 1.0$
AT/TA	$-7.15 \pm 0.78$	$-20.2 \pm 2.6$
TA/AT	$-7.23 \pm 0.82$	$-21.6 \pm 2.7$
CA/GT	$-8.44 \pm 0.77$	$-22.9 \pm 2.5$
GT/CA	$-8.47 \pm 0.66$	$-22.9 \pm 2.2$
CT/GA	$-7.73 \pm 0.66$	$-20.9 \pm 2.2$
GA/CT	$-8.29 \pm 0.61$	$-22.6 \pm 2.0$
CG/GC	$-10.54 \pm 0.82$	$-27.1 \pm 2.7$
GC/CG	$-9.81 \pm 0.73$	$-24.6 \pm 2.4$
GG/CC	$-8.02 \pm 0.68$	$-19.6 \pm 2.3$
Cap with G/C	$0.08 \pm 0.99$	$-3.7 \pm 3.4$
Cap with A/T	$2.22 \pm 1.02$	$2.8 \pm 3.3$
Log $\sigma$	$-9.0 \pm 2.7$	
Log $\bar{\sigma}$	$-5.7 \pm 1.4$	
$\gamma_S$ (cal/mol/K)	$-8.9 \pm 2.7$	
$\Delta S_{\text{mix}}^0$ (cal/mol/K)	$3.0 \pm 0.9$	
$K_{s1}$ (K <sup>-1</sup> )	$(4.29 \pm 0.29) \times 10^{-5}$	
$K_{s2}$ (K <sup>-1</sup> )	$(-3.95 \pm 0.16) \times 10^{-5}$	
$K_{s3}$ (K <sup>-1</sup> )	$(9.4 \pm 0.29) \times 10^{-6}$	

for  $\Delta S_{\text{mix}}^0$ ). The corresponding total of 24 relevant parameters in the limit of short oligomers undergoing two-state melting ( $10 \Delta h_{\text{NN}}$ ,  $10 \Delta s_{\text{NN}}$ ,  $2 \Delta h_{\text{ini}} = \Delta h_{\omega}$ ,  $2$  and  $\Delta s_{\text{ini}} = \Delta s_{\omega} + \Delta S_{\text{mix}}^0/2$ ) corresponds to the number of independent parameters, which can be determined uniquely from the corresponding experiments (31,33–35).

Correlations between NN-parameters are not negligible (correlation between  $\Delta h_{\text{NN}}$  and  $\Delta s_{\text{NN}}$  approaches 99% for a given basepair step). To have access to the covariance matrix, we repeated the analysis of Allawi and SantaLucia (33) (singular value decomposition (32) and of error evaluation via the Bootstrap method (36)), which consists in optimizing the melting temperatures prediction for short two-state oligomers, using the same experimental data. Our results (Table 1) slightly deviate from Allawi's values due to the larger number of resampling trials carried out ( $>100,000$ , compared to 30 in Allawi and SantaLucia (33)).

## Salt correction

Experimental observations (19) show that changing the salt concentration from 1 M to 100 mM shifts the melting temperature by  $\sim 10$  K. These variations are described by a number of phenomenological salt corrections that are different for short oligomers and polymers (10,19). A systematic, statistical-mechanical approach should derive the observed melting temperatures from a model with salt-dependent parameters. Given that DNA is a highly charged molecule, such a dependence is not surprising and is formally the result of integrating out microscopic degrees of freedom of the DNA along with those of the solvent molecules and salt ions.

The correction most used in bioinformatics programs (DINAmelt (37), MELTING (38)) is the one given by SantaLucia (10) for oligomer NN-parameters:

$$\Delta s_{\text{NN}}([\text{Na}^+]) = \Delta s_{\text{NN}}(1\text{M}) + 0.368 \times \log[\text{Na}^+]. \quad (5)$$

This correction is sequence-independent, while Owczarzy et al. (19) have recently shown that the melting temperatures of oligomers undergoing two-state melting follow the phenomenological rule

$$\frac{d}{d(\log[\text{Na}^+])} \left( \frac{1}{T_m} \right) = (K_{s1}f(GC) + K_{s2}) + 2K_{s3} \log[\text{Na}^+], \quad (6)$$

where  $K_{s1}$ ,  $K_{s2}$ , and  $K_{s3}$  are empirical numbers of the same order (see Table 1) and  $f(GC)$  is the GC content of the sequence. Equation 6 cannot be used directly for our purposes. To infer the salt-dependence of the local NN-parameters  $\Delta h_{\text{NN}}$  and  $\Delta s_{\text{NN}}$  we proceed in two steps: firstly, we follow the literature (39,40) and assume that the complexation enthalpy is independent of salt-concentration. In contrast, the gain in mixing entropy of counterions released from molten chain sections does depend on the salt concentration (41,42). With

$$1/T_m = (\Delta S_0 + k_B \log(c_T/(c_0/\alpha)))/\Delta H_0 \quad (7)$$

for two-state melting, we can rewrite Eq. 6 in the form

$$\frac{d}{d(\log[\text{Na}^+])}(\Delta S_0) = \Delta H_0 \times \{ (K_{s1}f(GC) + K_{s2}) + 2K_{s3} \log[\text{Na}^+] \}. \quad (8)$$

Secondly, instead of applying Eq. 8 to the whole sequence, we use the local GC content  $f_i(GC)$  of a basepair step. This results in a correction of the form

$$\Delta s([\text{Na}^+]) = \Delta s(1\text{M}) + \Delta h \{ (4.29f_i(GC) - 3.95) \log[\text{Na}^+] \times 10^{-5} + 9.4 \log^2[\text{Na}^+] \times 10^{-6} \} \quad (9)$$

for the NN-pair formation and capping entropies. We have checked that our results do not change significantly if larger environments up to 5 bps are taken into account for calculating  $f_i(GC)$ . Typical values of  $\delta s([\text{Na}^+]) \equiv \Delta s([\text{Na}^+]) - \Delta s(1\text{M})$  for the different basepairs are reported in Table 2. Note that the fits of the salt correction parameters and of the NN-parameters have been realized separately. Correlations between  $K_{s1}$ ,  $K_{s2}$ , and  $K_{s3}$  exist and are considered in the error propagation analysis.

## Cooperativity

There is no a priori reason why the cooperativity parameters should not have an enthalpic contribution, or why they should be independent of DNA sequence and of the salt concentration. To date, the interfacial parameters  $\sigma$  and  $\bar{\sigma}$  are not known with sufficient precision. The cooperativity  $\sigma$  is in the range  $10^{-4}$ – $10^{-5}$  (24,25,29,43) corresponding to a loop nucleation free energy  $2\gamma \equiv -k_B T \log \sigma \equiv -2\gamma_S T$  of  $\sim 10 k_B T$ . The bare free-end formation

**TABLE 2** Salt correction from Eq. 9

Sequence	$\delta s([\text{Na}^+])$ (cal/mol/K)			
	0.01 M	0.05 M	0.1 M	0.5 M
AA/TT	$3.0 \pm 0.2$	$1.6 \pm 0.1$	$1.1 \pm 0.05$	$0.3 \pm 0.01$
AT/TA	$2.7 \pm 0.3$	$1.5 \pm 0.2$	$1.0 \pm 0.1$	$0.2 \pm 0.03$
TA/AT	$2.8 \pm 0.3$	$1.5 \pm 0.2$	$1.0 \pm 0.1$	$0.2 \pm 0.03$
CA/GT	$2.4 \pm 0.2$	$1.2 \pm 0.1$	$0.8 \pm 0.1$	$0.1 \pm 0.02$
GT/CA	$2.4 \pm 0.2$	$1.2 \pm 0.1$	$0.8 \pm 0.05$	$0.1 \pm 0.02$
CT/GA	$2.2 \pm 0.2$	$1.1 \pm 0.1$	$0.7 \pm 0.05$	$0.1 \pm 0.02$
GA/CT	$2.3 \pm 0.2$	$1.1 \pm 0.1$	$0.8 \pm 0.05$	$0.1 \pm 0.02$
CG/GC	$1.9 \pm 0.2$	$0.8 \pm 0.1$	$0.4 \pm 0.05$	$0.0 \pm 0.02$
GC/CG	$1.8 \pm 0.2$	$0.7 \pm 0.1$	$0.4 \pm 0.05$	$0.0 \pm 0.02$
GG/CC	$1.5 \pm 0.2$	$0.6 \pm 0.1$	$0.3 \pm 0.05$	$0.0 \pm 0.02$
Cap with G/C	$0.0 \pm 0.2$	$0.0 \pm 0.1$	$0.0 \pm 0.05$	$0.0 \pm 0.01$
Cap with A/T	$0.8 \pm 0.4$	$0.5 \pm 0.2$	$0.3 \pm 0.1$	$0.0 \pm 0.03$

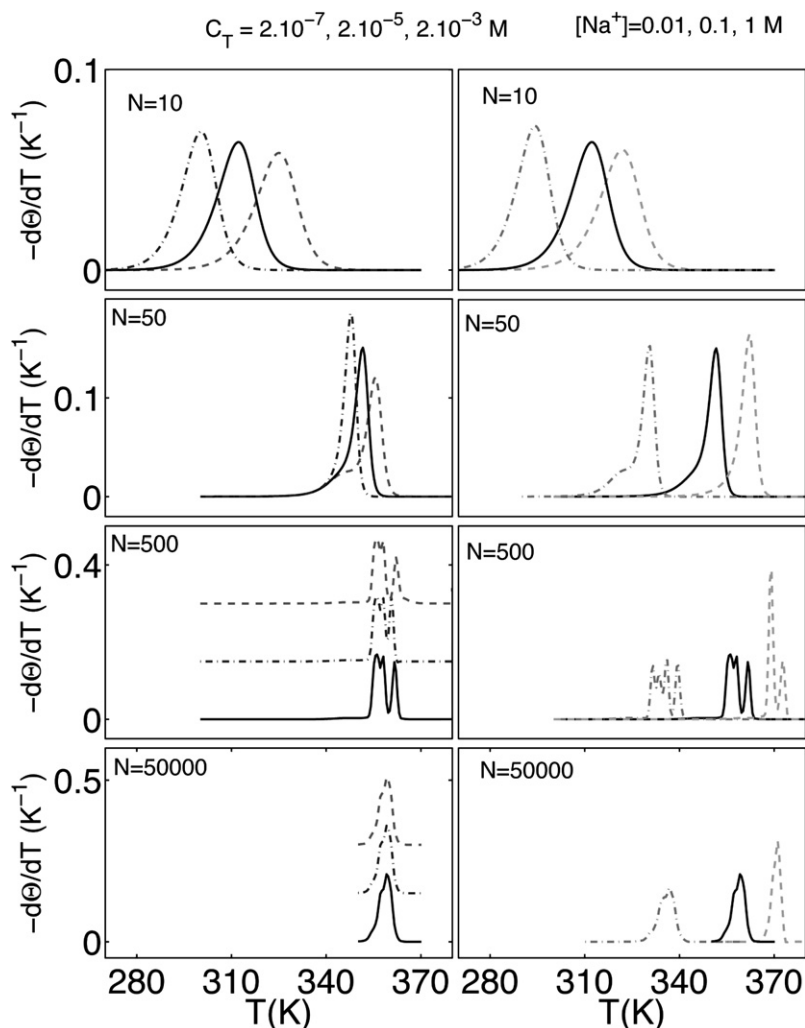


FIGURE 2  $-\text{d}\Theta/\text{d}T$  versus  $T$  for different random sequences ( $f(\text{GC}) = 0.5$ ):  $N = 10$  (first line),  $N = 50$  (second line),  $N = 500$  (third line), and  $N = 50,000$  (fourth line). (Left)  $[\text{Na}^+] = 0.1 \text{ M}$  and  $c_T = 2 \times 10^{-3} \text{ M}$  (dashed lines),  $c_T = 2 \times 10^{-5} \text{ M}$  (solid lines), and  $c_T = 2 \times 10^{-7} \text{ M}$  (dotted and dashed lines). (Right)  $c_T = 2 \times 10^{-5} \text{ M}$  and  $[\text{Na}^+] = 1 \text{ M}$  (dashed lines),  $[\text{Na}^+] = 0.1 \text{ M}$  (solid lines), and  $[\text{Na}^+] = 0.01 \text{ M}$  (dotted and dashed lines).

parameter  $\bar{\sigma}$  is of course correlated to  $\sigma$  (with the opening of a free-end there is a creation of one interface whereas, with a loop, two interfaces appear) and  $\bar{\sigma} \sim \sqrt{\sigma}$ . In the following, we assume  $\bar{\sigma} = 0.3\sqrt{\sigma}$ . The numerical values of the different parameters are listed in Table 1. The lack of data concerning  $\sigma$  forces us to assume an high arbitrary standard deviation of 30% for  $\log \sigma$  and to neglect possible correlations between  $\sigma$  and NN-parameters.

### Generic melting behavior

The generic behavior of the model is shown in Fig. 2 for random sequences ( $f(\text{GC}) = 0.5$ ) of different lengths ( $N = 10, 50, 500$ , or  $50,000$ ). In the two columns of the figure, we compare the evolution of the hybridization observable  $-\text{d}\Theta/\text{d}T$  as a function of the strand concentration  $c_T$  and the ionic strength of the  $[\text{Na}^+]$ -buffer.

Oligomer melting curves ( $N = 10, 50$ ) show one or two peaks. The height and width of the main peak are related to the chain length. For short oligomers, the transition occurs in a temperature interval of  $\sim 40 \text{ K}$  and centers around physiological temperatures (310 K). Melting curves for short- ( $N = 500$ ) and medium-sized DNA polymers exhibit several peaks due to successive domain opening. For very long chains ( $N = 50,000$ ),  $-\text{d}\Theta/\text{d}T$  becomes again featureless due to the superposition of large numbers of simultaneously occurring domain-melting events.

The strand concentration  $c_T$  has a strong influence on oligomer hybridization, but has negligible effects for long chains. Generally, larger concentrations reduce the single-strand gain in mixing entropy and stabilize the complex. The relative importance of the strand translational entropy is

highest for short chains at low concentrations resulting in two-state transitions. For longer oligomers and higher concentrations, internal melting competes with strand dissociation, and the two-state character of the transition is lost.

Variation of the salt concentration leads to comparable shifts of melting curves in all length regimes, and only to small changes in curve shapes. Low concentrations tend to stabilize bubbles and single strands (due to the decreasing screening effect of counterions (19,42)). Closer inspection reveals effects of the sequence-dependence of the salt correction (shown in Eq. 9). Lower salt concentrations tend to favor the partial opening of AT-rich domains like TATA boxes (promoter region in eukaryote genes and transcription initiator (44)).

### QUANTITATIVE COMPARISON TO EXPERIMENT—PREDICTIVE POWER

In the following, we compare available experimental data to the predictions of the model. Good agreement for the short oligomers used for the parameterization is to be expected, but not trivial: the experimentally observed two-state melting behavior (i.e., the irrelevance of fluctuations) has to be a property of the correctly parameterized model including fluctuations. The following comparisons for long polymers and intermediate chain lengths constitute true tests of the



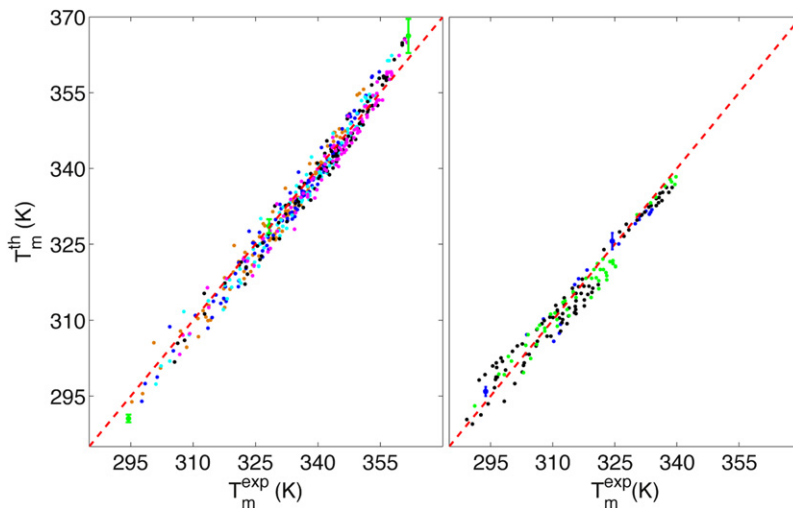


FIGURE 3 Calculated melting temperatures  $T_m^{\text{th}}$  as a function of the experimental ones  $T_m^{\text{exp}}$  for (left) 92 sequences from Owczarzy et al. (19) at five different salt concentrations ( $c_T = 2 \times 10^{-6}$  M): 69 mM (orange), 119 mM (blue), 220 mM (cyan), 621 mM (black), and 1.02 M (violet) (green dots are representative of the standard deviations); and for (right) 20 sequences from Santa-Lucia et al. (53) at various strand concentrations ( $[\text{Na}^+] = 1$  M):  $c_T \in [3.65 \times 10^{-6}, 10^{-5}]$  (blue),  $c_T \in [10^{-5}, 10^{-4}]$  (black), and  $c_T \in [10^{-4}, 6.32 \times 10^{-4}]$  (green).

predictive power of the model, since the experimental data was not used for the parameterization.

### Short oligomers (~10 bps)

The average error in the predicted melting temperatures

$$\langle \Delta T_m \rangle = \frac{1}{N} \sum_i |T_m^{\text{exp}} - T_m^{\text{th}}| \quad (10)$$

for the data sets from the literature (19,33) that we used for the parameterization is  $\langle \Delta T_m \rangle = 1.7$  K (compared to  $\langle \Delta T_m \rangle = 2.4$  K using Eq. 5 for the salt correction and to the average experimental error of 0.3 K). The more detailed comparison of calculated and experimental melting temperatures in Fig. 3 shows that there are no preferred salt or strand concentrations over the entire experimentally available range of  $[\text{Na}^+] \in [0.01 \text{ M}, 1 \text{ M}]$  and  $c_T \in [3.7 \times 10^{-6} \text{ M}, 6.3 \times 10^{-4} \text{ M}]$ : all data points are uniformly aggregated along the bisectors with error estimates corresponding to the typical deviation from the experimental value. Moreover, the two-state behavior of the transition can be shown by evaluating the maximum  $\Sigma_{\text{max}}$  of  $\Sigma = \Theta_{\text{ext}} - \Theta$  (15). If the sequence presents a two-state transition,  $\Theta_{\text{ext}} = \Theta$  or  $\Sigma_{\text{max}} = 0$  (Fig. 4). Note that the comparison is made to the full model including fluctuations, which were neglected for the parameterization. Thus, the model properly reproduces the experimental observation of two-state melting for the sequences in question. Neglecting the cooperativity factors would lead to drastically different results (18).

### Long polymers ( $\geq 10$ kbps)

In the opposite limit of very long DNA, melting curves become again relatively featureless (see Fig. 2) and can be characterized by a melting temperature, which depends on the GC-content,  $f_{GC}$ . More than 30 years ago, Frank-Kamenetskii (45) and Vologodskii et al. (46) proposed the following empirical relations:

$$T_m = T_m^{\text{AT}} + f(\text{GC})(T_m^{\text{GC}} - T_m^{\text{AT}}) \quad (11)$$

with

$$T_m^{\text{AT}} = (355.55 + 7.9 \log[\text{Na}^+]) \text{K}$$

and

$$T_m^{\text{GC}} = (391.55 + 4.89 \log[\text{Na}^+]) \text{K}.$$

We have generated random DNA sequences of length  $N = 50,000$  with  $0.4 \leq f_{GC} \leq 0.6$  within the experimental range. Fig. 5 A shows excellent agreement between our results and Eq. 11, provided our variant Eq. 9 of the Owczarzy's salt correction is used. We consider this agreement to be a key result of this work, since it presents a systematic derivation of long-chain experimental behavior from short-chain data used in the parameterization on the basis of a statistical mechanical

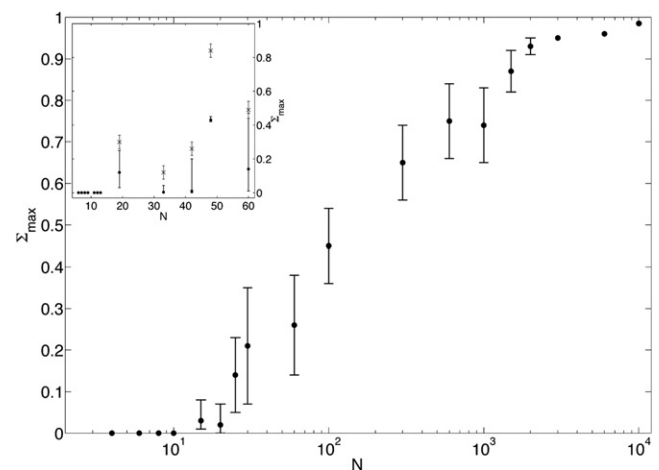


FIGURE 4  $\Sigma_{\text{max}}$  for different chain length ( $N$ ) random sequences ( $f_{GC} = 0.5$ ) in a 0.1 M  $[\text{Na}^+]$ -buffer with  $c_T = 2 \times 10^{-4}$  M (dots). The oligomers of Zeng et al. (15) in a 0.05 M  $[\text{Na}^+]$ -buffer with  $c_T = 2 \times 10^{-6}$  M (crosses, experimental results; dots, calculated results) and typical short oligomers (37) in a 1 M  $[\text{Na}^+]$ -buffer with  $c_T = 2 \times 10^{-6}$  M (dots) are shown in the inset.

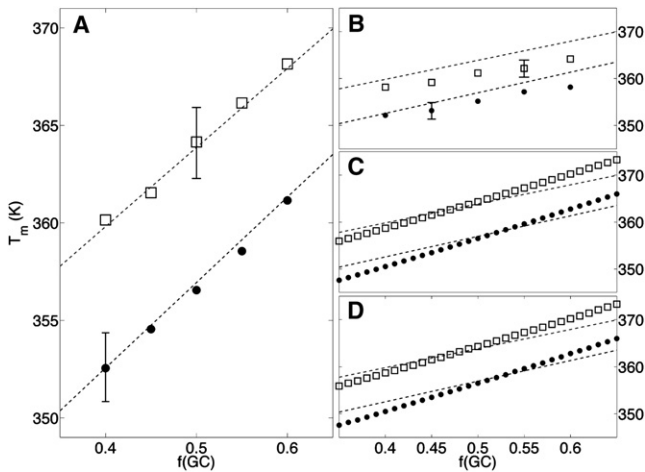


FIGURE 5 (A and B) Computed  $T_m$  with the unified PS-model for random heteropolymers (50,000 bp) with different GC content  $f(GC)$  in 74.5 mM (dots) and 220 mM (squares)  $[Na^+]$  buffer. Salt corrections are Eq. 9 (A) or Eq. 5 (B). The drawn error bars are representative of the standard deviations due to parameterization. (C and D) Theoretical  $T_m$  computed for random polymers if the transition was two-stated (C) or if the nucleation entropy was null (D); we use the salt correction we introduced in the article. Dashed lines represent the empirical relations given by Frank-Kamenetskii (45) for very long polymers (Eq. 11).

model. In particular, Fig. 5 A provides strong evidence for the validity of our local salt correction (9) for the NN parameters.

### Short polymers (100 bps $\leq \dots \leq$ 10 kbps)

Melting curves for short- and medium-sized polymers show a rich structure, but can be discussed independently of strand concentration (Fig. 2). To test the predictive power of our model, we have chosen PN/MCS-13 (note that this sequence is a 4660-bp duplex composed by a pBR322 mixed 245-bp repetitive sequence. pBR322 is associated with the primary accession number J01749 (43)). Fig. 6 shows that our model

reproduces the melting temperature of PN/MCS-13 fairly well, but fails to predict the fine structure of the curves. In particular, calculated and measured differential melting curves agree only qualitatively. Other bioinformatics programs (DINAmelt (37), MELTSIM (47), etc.) give similar results for these sequences.

How are these deviations to be interpreted? Apart from melting curves calculated using the standard parameters, Fig. 6 contains several other curves for parameter combinations drawn randomly within the correlated confidence limits from the parameterization. The curves work equally well (or badly), in particular the deviations from the experimental curve are actually within the confidence limits of the theoretical predictions and do not reveal shortcomings of the model itself.

This raises the question whether the model has any predictive power for polymer melting beyond Eq. 11. Judging from the ensemble of differential melting curves, the answer appears to be negative. However, the position-dependent melting temperature, which allows for a convenient identification of simultaneously opening basepairs, appears extremely robust with respect to the small uncertainties in the parameterization (48). Errors on local  $T_m$  are of the same order of magnitude as for oligomers. Testing these more detailed predictions using techniques based on the one-electron oxidative modifications of guanine induced by UV-laser (49) is therefore an interesting challenge.

### Long oligomers (20 bps $\leq \dots \leq$ 100 bps)

While short oligomers show two-state melting, this is no longer true for longer chains. Compared to the polymer case, the additional difficulty arises that results depend on the strand concentration (Fig. 2).

Attention recently focused on this regime with the development of an experimental protocol by Zeng et al. (15) and Zeng and Zocchi (16) for measuring the degree of strand association

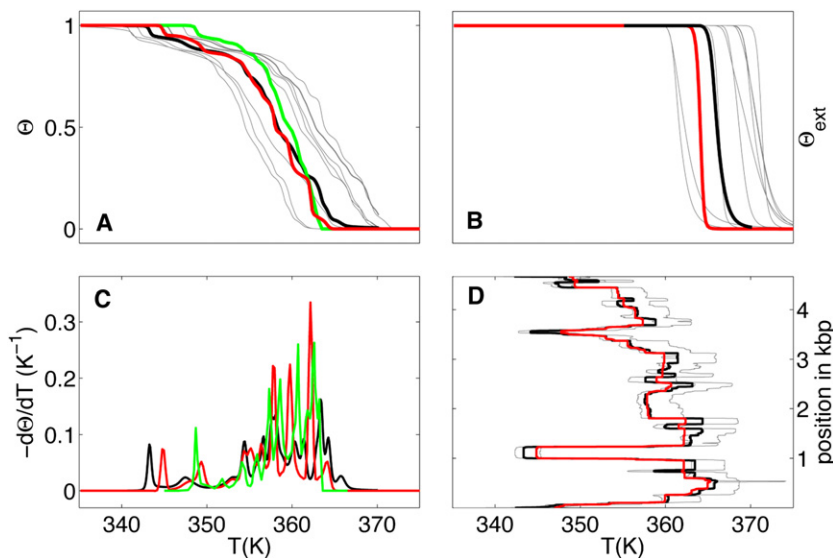


FIGURE 6 Melting curves  $\Theta$  (A),  $\Theta_{ext}$  (B),  $-d\Theta/dT$  (C), and evolution of the local basepair melting temperature (D) for PN/MCS-13 (43,47) in a 74.5 mM  $[Na^+]$  buffer. Green lines are experimental data, black lines are computed results with standard parameters, and other colored lines (gray, red) represent simulations with different random set of parameters (see text).

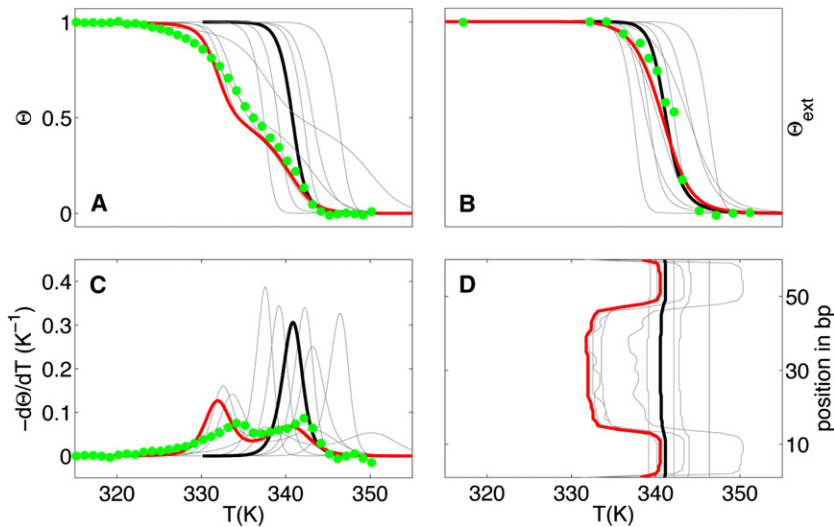


FIGURE 7 Melting curves  $\Theta$  (A),  $\Theta_{\text{ext}}$  (B),  $-\text{d}\Theta/\text{d}T$  (C), and evolution of the local basepair melting temperature (D) for the sequence *L60B36* of Zeng et al. (15) in a 50 mM  $[\text{Na}^+]$ -buffer ( $c_T = 2 \times 10^{-6}$  M). Green dots are experimental data, black lines are computed results with standard parameters and other colored lines (gray, red) represent simulations with different random set of parameters. Red curve underlines a set of parameters that reproduce well the experimental plots for *L60B36*.

$\Theta_{\text{ext}}$  independently of the degree of basepairing and stacking. In Fig. 7 we compare calculated and experimentally measured results for an oligomer *L60B36* (15) with one AT-rich domain in the center. Using the standard parameters, our model incorrectly predicts a two-state transition. However, other parameter combinations within the confidence limits correctly reproduce the opening of an internal bubble before strand separation and lead to good agreement with both  $\Theta$  and  $\Theta_{\text{ext}}$  (see Fig. 7). Fig. 4 compares the calculated and experimental results for  $\Sigma_{\text{max}}$  for the five sequences (inset) of Zeng et al. (15) and Zeng and Zocchi (16) and shows the generic behavior of  $\Sigma_{\text{max}}$  for random sequences. The large, asymmetric error bars for long oligomers highlight the particular sensitivity of the predicted behavior to the parameterization uncertainty.

## CONCLUSION

We have presented a unified Poland-Scheraga model of DNA thermal denaturation. In contrast to previously available

formulations, our description covers the entire crossover from oligo- to polynucleotide melting behavior and is applicable in the full experimental range of DNA strand and salt concentrations. We have used this Ansatz to discuss generic aspects of DNA melting and were able to obtain a systematic link between the different phenomenological sequence and salt dependences of short- and long-chain melting temperatures. Within the expected margin of error, our model reproduces experimental data for DNA of arbitrary length including the case of non-two-state-melting of longer DNA-oligomers. However, here (and to some degree in the case of domain melting of polymeric DNA) the model predictions for easily observable qualitative features are particularly sensitive to the remaining parameterization uncertainty.

Fig. 8 shows how the uncertainty in the knowledge of particular parameters (or parameter classes) affects predictions for melting curves in the various length regimes. In most cases, the dominant contribution comes from the set of 20 NN-parameters for the dinucleotide steps. They are most

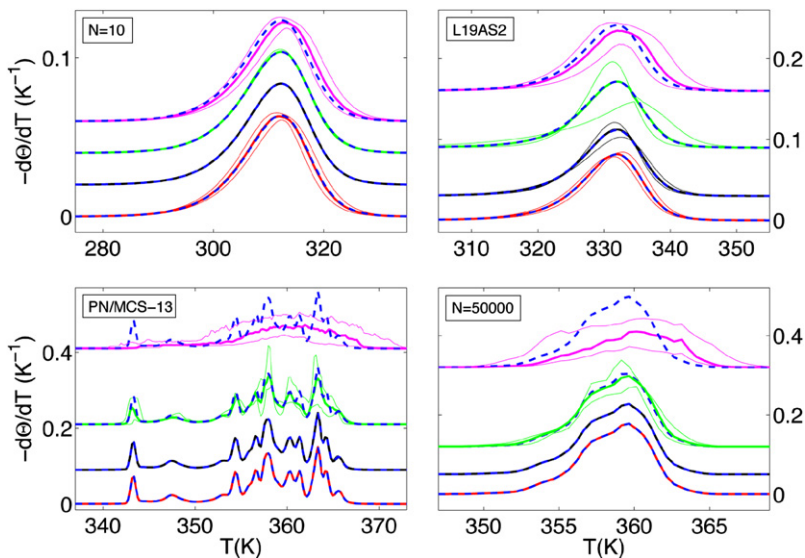


FIGURE 8 Melting curves  $-\text{d}\Theta/\text{d}T$  for short random ( $N = 10$ ) and long (*L19AS2* (15)) oligomers and for short (*PN/MCS-13*) and long random ( $N = 50,000$ ) polymers. For each chain-length DNA, we vary different parameters: initiation (red), capping (black), cooperativity (green), and NN-parameters (violet). The thick lines represents the mean curve over the variation of parameters. The thin lines are the confidence limits. The standard parameters curves are the dashed blue lines.

easily determined using data for short oligomers undergoing a two-state melting transition (10,21), since results in this case become independent of the cooperativity parameters  $\sigma$  and  $\bar{\sigma}$ . We emphasize that the Frank-Kamenetskii relations cannot be used directly to determine (linear combinations (34) of) the microscopic parameters of the PS Hamiltonian. They represent a sequence average over highly cooperative domain opening events and are not given by a trivial average over the employed NN-parameters (see Fig. 5, C and D). However, Figs. 5 A and 8 suggest that a proper comparison to averaged melting temperatures calculated from the full PS-model might be an excellent (albeit computationally expensive) strategy to refine the NN parameters. The cooperativity parameters  $\sigma$  and  $\bar{\sigma}$  affect the melting profiles of short polymers and, in particular, long oligomers. The sensitivity of these sequences has to be used to minimize errors during the parameterization process (50). The best strategy to determine their values is to devise comparative melting experiments along the lines of Blake and Delcourt (43), which allow us to isolate their effect (for more details, see Appendix C).

However, there will still exist inherent error bars that reflect the hypothesizes of the model and the ad hoc simplifications made. To reduce the intrinsic errors, one solution should be to increase the number of parameters describing chemical or physical effects not taken into account in the PS model (elasticity, basepair dependence of cooperative factors, etc.). Nevertheless, this could make the model difficult to parameterize and computationally more demanding.

For future work, our results suggest (a combination of) several strategies:

1. The analysis of the importance of the various parameter types for particular melting experiments (Fig. 8).
2. The design of (comparative) melting experiments isolating particular parameters following the framework used in Blake and Delcourt (43) (more details and examples can be found in Appendix C).
3. A simultaneous fit of the model to an extended base of experimental data for short-, medium-, and long-chain melting. This option is finally available with this formulation of the Poland-Scheraga model where all cases are treated within the same statistical-mechanical framework.

## APPENDIX A: MIXING ENTROPY AND CHEMICAL EQUILIBRIUM

To calculate the chemical equilibrium between the individual strands and the dimer, we need to account for the translational entropy, i.e., the entropy of mixing with the solvent. Using a lattice approach, one can show that the entropy of mixing per unit volume for two species P and Q is given by (51)

$$-T\Delta S_{\text{mix}}/V = k_B T \left( \frac{\Phi}{V_P} \log \Phi + \frac{1-\Phi}{V_Q} \log(1-\Phi) \right), \quad (12)$$

where  $\Phi$  is the volume fraction of species P, and  $V_P$  and  $V_Q$  are the respective molecular volumes. We have  $\Phi = c_P V_P$  and  $1-\Phi = c_Q V_Q$ , where  $c_P$  and  $c_Q$

are the respective concentrations ( $c_P = n_P/V$  with  $n_P$  the number of particles P, and the same for Q). Now we apply precedent equation to the DNA-water problem. P is either a single strand (A or B) or a double-strand (AB), and Q is water. All the studies are made at very low DNA concentration, i.e.,  $\Phi \ll 1$ , then  $(1-\Phi)/V_Q \log(1-\Phi) \approx -\Phi/V_Q = -c_P V_P/V_Q$ . Therefore, the entropy of mixing per molecule P can be obtained as

$$\begin{aligned} -T\Delta S_{\text{mix}}/n_P &= -T\Delta S_{\text{mix}}/V/c_P \\ &= k_B T \left( \log(c_P V_P) - \frac{V_P}{V_Q} \right). \end{aligned} \quad (13)$$

Finally, if concentrations are measured in units of a reference concentration  $c_0$  (typically  $c_0 = 1$  M),

$$-T\Delta S_{\text{mix}}/n_P = k_B T \left( \log\left(\frac{c_P}{c_0}\right) + \log(c_0 v_P) - \frac{v_P}{v_Q} \right), \quad (14)$$

where  $v_P$  and  $v_Q$  are the respective molar volumes.

Now we combine the entropy of mixing with the single-chain partition functions to obtain the free energy per molecule

$$G_A = G_A^{\text{int}} + k_B T \left( \log\left(\frac{c_A}{c_0}\right) + \log(c_0 v_A) - \frac{v_A}{v_{\text{water}}} \right) \quad (15)$$

(and the same for B and AB). The chemical potential per molecule can be calculated by differentiating the free energy density with respect to density,

$$\begin{aligned} \mu_A &= \frac{d}{dc_A} (c_A G_A) \\ &= G_A^{\text{int}} + k_B T \left( \log\left(\frac{c_A}{c_0}\right) + \log(c_0 v_A) - \frac{v_A}{v_{\text{water}}} + 1 \right) \end{aligned} \quad (16)$$

(and the same for B and AB). To obtain the law of mass action, we equate the chemical potentials of the bound and unbound molecules  $\mu_{\text{AB}} = \mu_A + \mu_B$ :

$$\begin{aligned} (G_{\text{AB}}^{\text{int}} - G_A^{\text{int}} - G_B^{\text{int}})/(k_B T) &= \log\left(\frac{c_A c_B}{c_{\text{AB}} c_0}\right) \\ &+ \log\left(\frac{c_0 v_A v_B}{v_{\text{AB}}}\right) \\ &- \frac{v_A + v_B - v_{\text{AB}}}{v_{\text{water}}}. \end{aligned} \quad (17)$$

By assuming that the molar volumes in the complex simply add up  $v_{\text{AB}} = (v_A + v_B)$ , this simplifies to

$$\begin{aligned} \Delta G_0 \equiv k_B T \log\left(\frac{c_A c_B}{c_{\text{AB}} c_0}\right) &= (G_{\text{AB}}^{\text{int}} - G_A^{\text{int}} - G_B^{\text{int}}) \\ &- k_B T (\log(c_0 v_{\text{AB}}/4) + 1). \end{aligned} \quad (18)$$

By defining  $G_{\text{int}} = (G_{\text{AB}}^{\text{int}} - G_A^{\text{int}} - G_B^{\text{int}})$  and  $\Delta G_{\text{mix}}^0 = -k_B T (\log(c_0 v_{\text{AB}}/4) + 1)$ , we find Eq. 2. The molar volume of the duplex is proportional to the number of dinucleotide steps in the double-strand:  $v_{\text{AB}} = (N-1)v_s$ , where  $v_s$  is the molar volume of one dinucleotide step. Numerically, we obtain  $\Delta G_{\text{mix}}^0 = -k_B T \log(0.44(N-1))$ .



## APPENDIX B: SOLVING THE UNIFIED PS MODEL

### Recursion relations

First, we consider the forward partition function  $Z_f(\alpha + 1)$  starting at base 1, ending at base  $\alpha + 1$ , bases  $\alpha$  and  $\alpha + 1$  being paired. There are three ways to have these bases closed: either the dinucleotide basepair step ( $\alpha - 1, \alpha$ ) is double-stranded, or there is a loop starting at any base  $\alpha'$  and ending at  $\alpha$ , or the duplex is completely open from base 1 to base  $\alpha$  (Fig. 9). Therefore, considering the closed state as the reference state, we can write

$$Z_f(\alpha + 1) = Z_f(\alpha) + \sigma \sum_{\alpha'=2}^{\alpha-2} (\alpha - \alpha')^{-c} e^{\beta g_{\alpha', \alpha-1}} Z_f(\alpha') + \bar{\sigma} (\alpha - 1)^{c'} e^{\beta g_{1, \alpha-1}}, \quad (19)$$

where  $(\alpha - \alpha')^{-c}$  accounts for the number of self-avoiding polygons of length  $2(\alpha - \alpha')$  (i.e., the number of loops starting at base  $\alpha'$  and ending at base  $\alpha$ );  $(\alpha - 1)^{c'}$  accounts for the number of possible conformations for a free end,  $g_{\alpha', \alpha} = \sum_{i=\alpha'}^{\alpha} \Delta g_{\text{NN}}(i, i + 1)$ , with  $\Delta g_{\text{NN}}(i, i + 1)$  the NN-stacking free energy of basepairs  $(i, i + 1)$ ;  $c$  and  $c'$  take into account the steric interactions between loop or the free end with the rest of the chain. The value of  $c$  has been extensively discussed (11,22–25) and is equal to 1.764 for noninteracting self-avoiding loops and 2.15 for interacting self-avoiding loops. The value of  $c'$  is derived from polymer theory (26) and is equal to  $0.16 = 1.16 - 1$ .

In a similar way, we consider  $Z_b(\alpha)$  the backward partition function, starting at base  $\alpha$  and ending at base  $N$ , with base  $\alpha$  being paired. So,

$$Z_b(\alpha) = Z_b(\alpha + 1) + \sigma \sum_{\alpha'=\alpha+2}^{N-1} (\alpha' - \alpha)^{-c} e^{\beta g_{\alpha, \alpha'-1}} Z_b(\alpha' + 1) + \bar{\sigma} (N - \alpha)^{c'} e^{\beta g_{\alpha, N-1}}. \quad (20)$$

Finally, we denote by  $Z_{\text{sf}}(\alpha)$  the second forward partition function, starting at base 1 and ending at base  $\alpha$ , base  $\alpha$  being closed and base  $\alpha - 1$  being opened. In the same way,  $Z_{\text{sf}}$  satisfies

$$Z_{\text{sf}}(\alpha) = \sigma \sum_{\alpha'=2}^{\alpha-2} (\alpha - \alpha')^{-c} e^{\beta g_{\alpha', \alpha-1}} Z_f(\alpha') + \bar{\sigma} (\alpha - 1)^{c'} e^{\beta g_{1, \alpha-1}}. \quad (21)$$

The probability  $p(\alpha)$  that basepair  $\alpha$  is bound can be expressed as

$$p(\alpha) = \frac{Z_{\text{sf}}(\alpha) Z_b(\alpha + 1) + Z_f(\alpha) Z_b(\alpha) + e^{-\beta(\omega_1 + \omega_N)}}{Z}, \quad (22)$$

where  $Z$  is the total partition function (Fig. 10)

$$Z = e^{-\beta(\omega_1 + \omega_N)} + e^{-\beta\omega_N} Z_f(N) + \bar{\sigma} \sum_{\alpha=2}^{N-1} (N - \alpha)^{c'} e^{\beta g_{\alpha, N-1}} Z_f(\alpha). \quad (23)$$

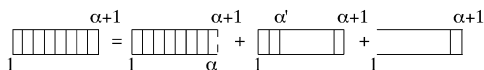


FIGURE 9 Graphical representation of the recursion relation for  $Z_f(\alpha + 1)$ .

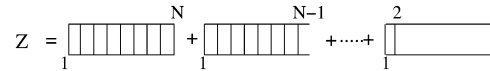


FIGURE 10 Graphical representation of total partition function  $Z$ .

We can now express  $\Theta_{\text{int}}$

$$\Theta_{\text{int}} = \frac{1}{N} \sum_{\alpha=1}^N p(\alpha) \quad (24)$$

and  $\Delta G_{\text{int}} = \sum_{i=1}^{N-1} \Delta g_{\text{NN}}(i, i + 1) - k_B T \log Z$ . The technical issue is now to solve numerically Eqs. 19–21. An answer is to simplify the recursion relations and speed up their numerical resolutions by using the Fixman-Freire algorithm (29,30).

### Fixman-Freire algorithm

We first fine-tune the recursion relations. With

$$Z_f^*(\alpha) = \exp\left(-\beta \sum_{i=1}^{\alpha-1} \Delta g_{\text{NN}}(i, i + 1)\right) Z_f(\alpha), \quad (25)$$

$$Z_b^*(\alpha) = \exp\left(-\beta \sum_{i=\alpha}^{N-1} \Delta g_{\text{NN}}(i, i + 1)\right) Z_b(\alpha), \quad (26)$$

$$Z_{\text{sf}}^*(\alpha) = \exp\left(-\beta \sum_{i=1}^{\alpha-1} \Delta g_{\text{NN}}(i, i + 1)\right) Z_{\text{sf}}(\alpha), \quad (27)$$

we derive

$$Z_f^*(\alpha + 1) = e^{-\beta \Delta g_{\text{NN}}(\alpha, \alpha+1)} Z_f^*(\alpha) + \sigma_0(\alpha) \sum_{\alpha'=2}^{\alpha-2} (\alpha - \alpha')^{-c} Z_f^*(\alpha') + \sigma_1(\alpha), \quad (28)$$

$$Z_b^*(\alpha) = e^{-\beta \Delta g_{\text{NN}}(\alpha, \alpha+1)} Z_b^*(\alpha + 1) + \sigma_0(\alpha - 1) e^{\beta \Delta g_{\text{NN}}(\alpha-1, \alpha)} \times \sum_{\alpha'=3}^N (\alpha' - \alpha - 1)^{-c} e^{-\beta \Delta g_{\text{NN}}(\alpha'-1, \alpha')} Z_b^*(\alpha') + \sigma_2(\alpha) e^{\beta \Delta g_{\text{NN}}(\alpha-1, \alpha)}, \quad (29)$$

$$Z_{\text{sf}}^*(\alpha) = \sigma \sum_{\alpha'=2}^{\alpha-2} (\alpha - \alpha')^{-c} Z_f^*(\alpha') + \bar{\sigma} (\alpha - 1)^{c-1}, \quad (30)$$

where  $\sigma_0(\alpha) = \sigma e^{-\beta \Delta g_{\text{NN}}(\alpha, \alpha+1)}$ ,  $\sigma_1(\alpha) = \bar{\sigma} e^{-\beta \Delta g_{\text{NN}}(\alpha, \alpha+1)} (\alpha - 1)^{c'}$ , and  $\sigma_2(\alpha) = \bar{\sigma} e^{-\beta \Delta g_{\text{NN}}(\alpha-1, \alpha)} (N - \alpha)^{c'}$ . The Fixman-Freire approximation consists in developing

$$x^{-c} \approx \sum_{k=1}^l a_k e^{-b_k x}. \quad (31)$$

We introduce new variables

$$e^{b_i \alpha} e^{\mu_i(\alpha)} \equiv \sum_{\alpha'=2}^{\alpha-2} e^{b_i \alpha'} Z_f^*(\alpha') \quad (32)$$

$$e^{-b_i \alpha} e^{\nu_i(\alpha)} \equiv \sum_{\alpha'=\alpha}^N e^{-b_i \alpha'} e^{-\beta \Delta_{\text{gNN}}(\alpha'-1, \alpha')} Z_b^*(\alpha'). \quad (33)$$

Therefore

$$Z_f^*(\alpha) = e^{\mu_i(\alpha)} - e^{-b_i} e^{\mu_i(\alpha-1)} \quad (34)$$

$$Z_b^*(\alpha) = e^{\beta \Delta_{\text{gNN}}(\alpha-1, \alpha)} [e^{\nu_i(\alpha)} - e^{-b_i} e^{\nu_i(\alpha+1)}], \quad (35)$$

and we obtain new recursions relations

$$\mu_i(\alpha + 1) = \mu_i(\alpha) + \log(A + B + C + D), \quad (36)$$

$$\nu_i(\alpha) = \nu_i(\alpha + 1) + \log(A' + B' + C' + D'), \quad (37)$$

with

$$\begin{aligned} A &= e^{-b_i} & B &= e^{-\beta \Delta_{\text{gNN}}(\alpha, \alpha+1)} \times \\ & & & (1 - e^{-b_i} e^{\mu_i(\alpha-1) - \mu_i(\alpha)}) \\ C &= \sigma_0(\alpha) \times & D &= \sigma_1(\alpha) e^{-\mu_i(\alpha)} \\ & \sum_{k=1}^I a_k e^{-2b_k} e^{\mu_k(\alpha-2) - \mu_i(\alpha)} & & \\ A' &= e^{-b_i} & B' &= e^{-\beta \Delta_{\text{gNN}}(\alpha-1, \alpha)} \times \\ & & & (1 - e^{-b_i} e^{\nu_i(\alpha+2) - \nu_i(\alpha+1)}) \\ C' &= \sigma_0(\alpha - 1) \times & D' &= \sigma_2(\alpha) e^{-\nu_i(\alpha+1)}, \\ & \sum_{k=1}^I a_k e^{-2b_k} e^{\nu_k(\alpha+3) - \nu_i(\alpha+1)} & & \end{aligned}$$

and for  $Z_{\text{sf}}^*$ ,

$$Z_{\text{sf}}^* = \sigma \sum_{k=1}^I a_k e^{-2b_k} e^{\mu_k(\alpha-2)} + \bar{\sigma}(\alpha - 1)^{c'}. \quad (38)$$

To solve this relation, we have to know initial conditions

$$Z_f^*(2) = e^{-\beta(\Delta_{\text{gNN}}(1,2) + \omega_1)} \quad (39)$$

$$Z_f^*(3) = e^{-\beta(\Delta_{\text{gNN}}(2,3) + \Delta_{\text{gNN}}(1,2) + \omega_1)} + \bar{\sigma} e^{-\beta \Delta_{\text{gNN}}(2,3)} \quad (40)$$

$$\begin{aligned} Z_f^*(4) &= e^{-\beta(\Delta_{\text{gNN}}(3,4) + \Delta_{\text{gNN}}(2,3) + \Delta_{\text{gNN}}(1,2) + \omega_1)} \\ &+ \bar{\sigma} e^{-\beta(\Delta_{\text{gNN}}(3,4) + \Delta_{\text{gNN}}(2,3))} \\ &+ \bar{\sigma} 2^{c'} e^{-\beta \Delta_{\text{gNN}}^0(3,4)}. \end{aligned} \quad (41)$$

Then

$$\mu_i(2) = \log(Z_f^*(2)), \quad (42)$$

$$\mu_i(3) = \log(Z_f^*(3) + e^{-b_i} Z_f^*(2)), \quad (43)$$

$$\mu_i(4) = \log(Z_f^*(4) + e^{-b_i} Z_f^*(3) + e^{-2b_i} Z_f^*(2)). \quad (44)$$

In the same way for  $\nu_i$ ,

$$Z_b^*(N) = e^{-\beta(\omega_N)}, \quad (45)$$

$$Z_b^*(N - 1) = e^{-\beta(\Delta_{\text{gNN}}(N-1, N) + \omega_N)} + \bar{\sigma}, \quad (46)$$

$$\begin{aligned} Z_b^*(N - 2) &= e^{-\beta(\Delta_{\text{gNN}}(N-2, N-1) + \Delta_{\text{gNN}}(N-1, N) + \omega_N)} \\ &+ \bar{\sigma} e^{-\beta \Delta_{\text{gNN}}(N-2, N-1)} + \bar{\sigma} 2^{c'}. \end{aligned} \quad (47)$$

Hence,

$$\nu_i(N) = \log(Z_b^*(N) e^{-\beta \Delta_{\text{gNN}}(N-1, N)}), \quad (48)$$

$$\begin{aligned} \nu_i(N - 1) &= \log(e^{-\beta \Delta_{\text{gNN}}(N-2, N-1)} Z_b^*(N - 1) \\ &+ e^{-b_i} e^{-\beta \Delta_{\text{gNN}}(N-1, N)} Z_b^*(N)), \end{aligned} \quad (49)$$

$$\begin{aligned} \nu_i(N - 2) &= \log(e^{-\beta \Delta_{\text{gNN}}(N-3, N-2)} Z_b^*(N - 2) \\ &+ e^{-b_i} e^{-\beta \Delta_{\text{gNN}}(N-2, N-1)} Z_b^*(N - 1) \\ &+ e^{-2b_i} e^{-\beta \Delta_{\text{gNN}}(N-1, N)} Z_b^*(N)). \end{aligned} \quad (50)$$

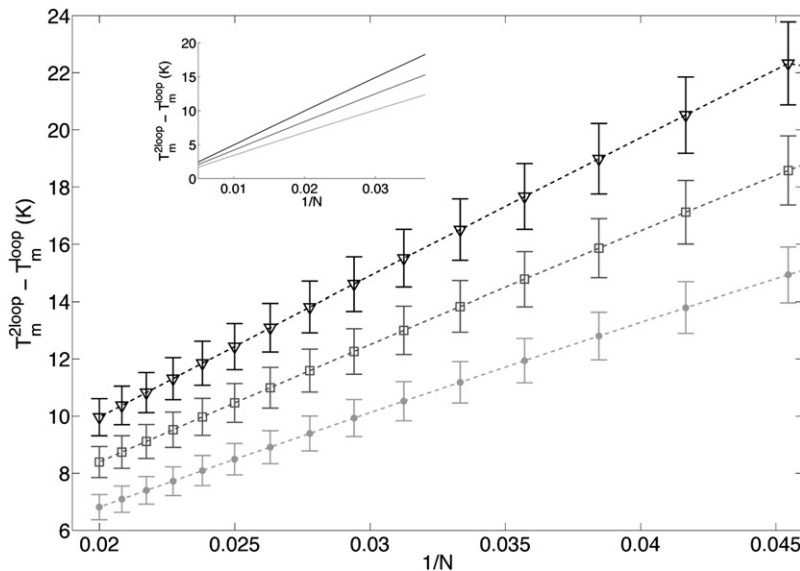


FIGURE 11  $T_m^{2loop} - T_m^{loop}$  as a function of  $1/N$  for different values of  $\gamma_S$ :  $-10.9$  cal/mol/K (triangles),  $-8.9$  cal/mol/K (squares), and  $-6.9$  cal/mol/K (dots). Error bars reflect the confidence limit on NN-parameters.

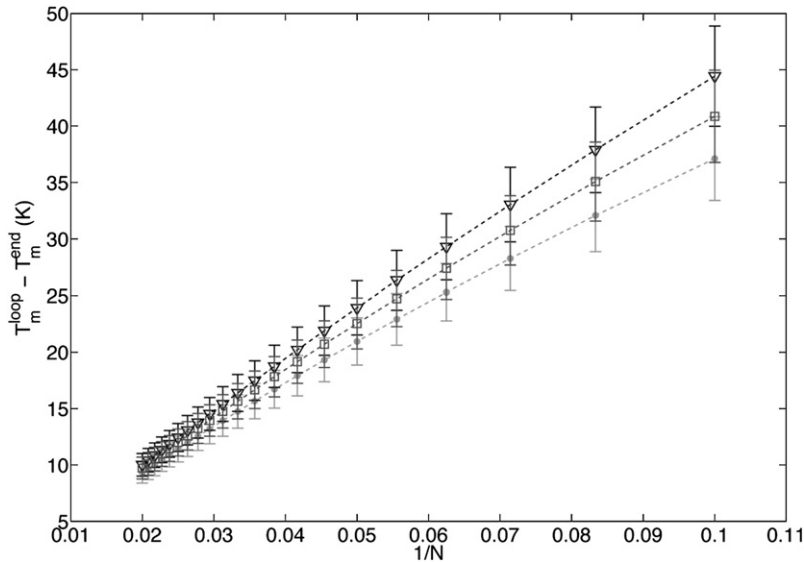


FIGURE 12  $T_m^{\text{loop}} - T_m^{\text{end}}$  as a function of  $1/N$  for different values of  $\Delta S_{\text{mix}}^0$ : 5 cal/mol/K (triangles), 3.0 cal/mol/K (squares), and 1 cal/mol/K (dots). Error bars reflect the confidence limit on NN-parameters.

For  $Z_{\text{sf}}^*$ ,

$$Z_{\text{sf}}^*(1) = 0 \quad (51)$$

$$Z_{\text{sf}}^*(2) = \bar{\sigma} \quad (52)$$

$$Z_{\text{sf}}^*(3) = \bar{\sigma} 2^{c'}. \quad (53)$$

The algorithm consists in solving recursion relations for  $\mu_i$  and  $\nu_i$ , in deducting values for  $Z_f^*$ ,  $Z_b^*$ , and  $Z_{\text{sf}}^*$ . Then it is easy to compute  $p(\alpha)$ ,  $\Theta_{\text{int}}$ , and  $\Theta_{\text{ext}}$ .

### APPENDIX C: IMPROVING THE MODEL PARAMETERIZATION

For example, to parameterize  $\gamma_S = k_{B/2} \log \sigma$ , we propose to study two types of sequences:  $S_{\text{loop}} = G_P A_N G_P$  (one central bubble) and  $S_{2\text{loop}} = G_{2P/3} A_{N/2} G_{2P/3}$  (two internal bubbles) with  $P$  large enough to neglect border effects. Following Blake and Delcourt (43), we can compute the difference of melting temperatures  $T_m^{\text{loop}} - T_m^{\text{end}}$  of the different bubbles ( $T_m^{\text{loop}}$  for  $S_{\text{loop}}$  and  $T_m^{\text{end}}$  for  $S_{2\text{loop}}$ ) for various  $N$ . As both sequences have the same basepair composition, this observable should not be very sensitive to NN-parameters and should be responsive to the energy difference between the two types of strands, which is  $\sim 2\gamma$ . Fig. 11 shows the simulated evolution of  $T_m^{\text{loop}} - T_m^{\text{end}}$  versus  $1/N$  for different  $\gamma_S$  values with the error bars due to standard deviations of NN-parameters. We remark that the observable is more and more sensitive to  $\gamma_S$  as long as  $1/N$  increases. Therefore, experiments have to be done with short bubbles ( $N$  must be big enough to permit to observe the bubble subtransition). Moreover, we notice that errors due to NN-parameters uncertainty limit the accuracy of the  $\gamma_S$  parameterization by  $\sim 10\%$ . Nevertheless, as experimental error are  $\sim 0.3$  K, a precise evaluation of  $\gamma_S$  is possible with sequences in the range  $N \in [20-40]$ .

In the same manner, to parameterize  $\Delta S_{\text{mix}}^0$ , we can study  $S_{\text{loop}}$  and  $S_{\text{end}} = A_{N/2} G_{2P} A_{N/2}$ . Fig. 12 reveals the sensitivity of  $T_m^{\text{loop}} - T_m^{\text{end}}$  according to  $\Delta S_{\text{mix}}^0$  for little  $N$ . Nonetheless, error bars do not allow a precise evaluation of  $\Delta S_{\text{mix}}^0$ . This difference with the previous example comes from the important role played by the borders. Indeed, NN-initiation parameters have large standard deviations that automatically reflect on the theoretical errors. Without an improvement on the errors of the initiation enthalpies and entropies, there is no hope to parameterize  $\Delta S_{\text{mix}}^0$  with these types of experiments.

Moreover, as seen in Fig. 4, the two-state parameter  $\Sigma_{\text{max}}$  is highly  $\sigma$ -dependent for long oligomers. Therefore, more experimental data for long oligomers would allow better estimates for the cooperativity factors by including the  $\Sigma_{\text{max}}$  values in the global parameterization process.

R.E. acknowledges support from the chair of excellence program of the Agence Nationale de la Recherche (ANR/France).

### REFERENCES

- Calladine, C., H. Drew, B. Luisi, and A. Travers. 2004. Understanding DNA; The Molecule and How It Works. Elsevier Academic Press, New York.
- Schuster, P. 2006. Prediction of RNA secondary structures: from theory to models and real molecules. *Rep. Prog. Phys.* 69:1419–1477.
- Fire, A., S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, and C. C. Mello. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*. 391:806–811.
- Shoemaker, D. D., E. E. Schadt, C. D. Armour, Y. D. He, P. Garrett-Engel, et al. 2001. Experimental annotation of the human genome using microarray technology. *Nature*. 409:922–927.
- Seeman, N. C. 2003. Biochemistry and structural DNA nanotechnology: an evolving symbiotic relationship. *Biochemistry*. 42:7259–7269.
- Nykypanchuk, D., M. M. Maye, D. van der Lelie, and O. Gang. 2008. DNA-guided crystallization of colloidal nanoparticles. *Nature*. 451:549–552.
- Park, S. Y., A. K. R. Lytton-Jean, B. Lee, S. Weigand, and G. C. Schatz. 2008. DNA-programmable nanoparticle crystallization. *Nature*. 451: 553–556.
- Crothers, D. M., and B. H. Zimm. 1964. Theory of the melting transition of synthetic polynucleotides evaluation of the stacking free energy. *J. Mol. Biol.* 9:1–9.
- DeVoe, H., and I. Tinoco, Jr. 1962. The stability of helical polynucleotides: base contributions. *J. Mol. Biol.* 4:500–517.
- SantaLucia, J., Jr. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA*. 95:1460–1465.
- Poland, D., and H. A. Scheraga. 1966. Phase transition in one dimension and the helix-coil transition in polyamino acids. *J. Chem. Phys.* 45:1456–1464.
- Poland, D., and H. A. Scheraga. 1970. Theory of Helix-Coil Transition in Biopolymers. Academic Press, New York.

13. Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31:3406–3415.
14. Hofacker, I. L. 2003. Vienna RNA secondary structure server. *Nucleic Acids Res.* 31:3429–3431.
15. Zeng, Y., A. Montrichok, and G. Zocchi. 2004. Bubble nucleation and cooperativity in DNA melting. *J. Mol. Biol.* 339:67–75.
16. Zeng, Y., and G. Zocchi. 2006. Mismatches and bubbles in DNA. *Biophys. J.* 90:4522–4529.
17. Daneholt, B. 2006. Advanced Information on The Nobel Prize Generally in Physiology or Medicine 2006—RNA Interference. The Nobel Assembly at Karolinska Institutet, Stockholm, Sweden.
18. Everaers, R., S. Kumar, and C. Simm. 2007. Unified description of poly- and oligonucleotide DNA melting: nearest-neighbor, Poland-Scheraga, and lattice models. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 75:041918–041930.
19. Owczarzy, R., Y. You, B. G. Moreira, J. A. Manthey, L. Huang, et al. 2004. Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures. *Biochemistry.* 43:3537–3554.
20. Wartell, R. M., and A. S. Benight. 1985. Thermal denaturation of DNA molecules: a comparison of theory with experiment. *Phys. Rep.* 126:67–107.
21. Owczarzy, R., P. M. Vallone, F. J. Gallo, T. M. Paner, M. J. Lane, et al. 1998. Predicting sequence-dependent melting stability of short duplex DNA oligomers. *Biopolymers.* 44:217–239.
22. Fisher, M. E. 1966. Effects of excluded volume on phase transitions in biopolymers. *J. Chem. Phys.* 45:1469–1473.
23. Kafri, Y., D. Mukamel, and L. Peliti. 2002. Melting and unzipping of DNA. *Eur. Phys. J. B.* 27:135–146.
24. Kafri, Y., D. Mukamel, and L. Peliti. 2003. Kafri, Mukamel, and Peliti reply. *Phys. Rev. Lett.* 90:159802.
25. Blossey, R., and E. Carlon. 2003. Reparameterizing the loop entropy weights: effect on DNA melting curves. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 68:061911–061918.
26. de Gennes, P. -G. 1979. Scaling Concepts in Polymer Physics. Cornell University Press, Ithaca, NY.
27. Vanderzande, C. 1998. Lattice Models of Polymers. Cambridge University Press, Cambridge, UK.
28. Dimitrov, R. A., and M. Zuker. 2004. Prediction of hybridization and melting for double-stranded nucleic acids. *Biophys. J.* 87:215–226.
29. Garel, T., and H. Orland. 2004. Generalized Poland-Scheraga model for DNA hybridization. *Biopolymers.* 75:453–467.
30. Fixman, M., and J. J. Freire. 1977. Theory of DNA melting curves. *Biopolymers.* 16:2693–2704.
31. Goldstein, R. F., and A. S. Benight. 1992. How many numbers are required to specify sequence-dependent properties of polynucleotides? *Biopolymers.* 32:1679–1693.
32. Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1996. Numerical Recipes in Fortran 77: the Art of Scientific Computing. Cambridge University Press, Cambridge, UK.
33. Allawi, H. T., and J. SantaLucia, Jr.. 1997. Thermodynamics and NMR of internal G·T mismatches in DNA. *Biochemistry.* 36:10581–10594.
34. Gray, D. M. 1997. Derivation of nearest-neighbor properties from data on nucleic acid oligomers. I. Simple sets of independent sequences and the influence of absent nearest neighbors. *Biopolymers.* 42:783–793.
35. Gray, D. M. 1997. Derivation of nearest-neighbor properties from data on nucleic acid oligomers. II. Thermodynamic parameters of DNA/RNA hybrids and DNA duplexes. *Biopolymers.* 42:795–810.
36. Efron, B., and R. Tibshirani. 1993. An Introduction to the Bootstrap. Chapman & Hall, London, UK.
37. Markham, N. R., and M. Zuker. 2005. DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.* 33:W577–W581.
38. Le Novere, N. 2001. MELTING, computing the melting temperature of nucleic acid duplex. *Bioinformatics.* 17:1226–1227.
39. Erie, D., N. Sinha, W. Olson, R. Jones, and K. Breslauer. 1987. A dumbbell-shaped, double-hairpin structure of DNA: a thermodynamic investigation. *Biochemistry.* 26:7150–7159.
40. Record, M. T., and T. M. Lohman. 1978. A semiempirical extension of polyelectrolyte theory to the treatment of oligoelectrolytes: application to oligonucleotide helix-coil transitions. *Biopolymers.* 17:159–166.
41. Manning, G. S. 1969. Limiting laws and counterion condensation in polyelectrolyte solutions. *J. Chem. Phys.* 51:924–933.
42. Manning, G. S. 1972. On the application of polyelectrolyte “limiting laws” to the helix-coil transition of DNA. I. Excess univalent cations. *Biopolymers.* 11:937–949.
43. Blake, R. D., and S. G. Delcourt. 1998. Thermal stability of DNA. *Nucleic Acids Res.* 15:3323–3332.
44. Smale, S. T., and J. T. Kadonaga. 2003. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* 72:449–479.
45. Frank-Kamenetskii, M. 1971. Simplification of the empirical relationship between melting temperature of DNA, its GC content and concentration of sodium ions in solution. *Biopolymers.* 10:2623–2624.
46. Vologodskii, A., B. Amirkyan, Y. Lyuchenko, and M. Krank-Kamenetskii. 1984. Allowance for heterogeneous stacking in the DNA helix-coil transition theory. *J. Biomol. Struct. Dyn.* 2:131–148.
47. Blake, R.D., J.W. Bizzaro, J.D. Blake, G.R. Day, S.G. Delcourt, et al. 1999. Statistical mechanical simulation of polymeric DNA melting with MELTSIM. *Bioinformatics.* 15:370–375.
48. Yeramian, E., and L. Jones. 2003. GeneFizz: a web tool to compare genetic (coding/non-coding) and physical (helix/coil) segmentations of DNA sequences. Gene discovery and evolutionary perspectives. *Nucleic Acids Res.* 31:3843–3849.
49. Spassky, A., and D. Angelov. 2002. Temperature-dependence of UV laser one-electron oxidative guanine modifications as a probe of local stacking fluctuations and conformational transitions. *J. Mol. Biol.* 323:9–15.
50. Bevington, P. R., and D. K. Robinson. 1992. Data Reduction and Error Analysis for the Physical Sciences. McGraw-Hill, New York.
51. Rubinstein, M., and R. H. Colby. 2003. Polymer Physics. Oxford University Press, New York.
52. Cantor, C. R., and P. R. Schimmel. 1980. Biophysical Chemistry. Freeman, New York.
53. SantaLucia, J., Jr., H. T. Allawi, and P. A. Seneviratne. 1996. Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry.* 35:3555–3562.