Unified description of poly- and oligonucleotide DNA melting: Nearest-neighbor, Poland-Sheraga, and lattice models

Ralf Everaers*
Max-Planck-Institut für Physik komplexer Systeme, Nöthnitzer Strasse 38, 01187 Dresden, Germany and Laboratoire de Physique, ENS Lyon, 46, allée d'Italie, 69364 Lyon cedex 07, France

Sanjay Kumar
Max-Planck-Institut für Physik komplexer Systeme, Nöthnitzer Strasse 38, 01187 Dresden, Germany and Department of Physics, Banaras Hindu University, Varanasi 221 005, India

Christian Simm
Max-Planck-Institut für Physik komplexer Systeme, Nöthnitzer Strasse 38, 01187 Dresden, Germany
(Received 16 August 2006; published 27 April 2007)

We show that a simple lattice model can provide a unified description of the thermal denaturation of DNA oligomers and polymers. The model quantitatively reproduces experimental melting curves and reduces in limiting cases to the nearest-neighbor model and a suitable modified Poland-Sheraga model. Our results support the interpretation of the cooperativity parameter σ for bubble opening in terms of an interfacial (forking) free energy which also affects chain opening from the ends. The lattice model treats long-ranged excluded volume interactions between all parts of the molecule explicitly, provides access to an ensemble of three dimensional structures (and hence the response to external mechanical forces), and includes fluctuations in situations without pre-determined secondary structure.

DOI: 10.1103/PhysRevE.75.041918 PACS number(s): 87.14.Gg, 87.15.Aa, 82.60.−s, 64.75.+g

I. INTRODUCTION

Key biological processes such as transcription and replication of genetic information require the opening of the double-helical (Watson-Crick) complex of complementary DNA strands [1]. For this reason, the related process of thermally induced denaturation has been studied for a long time [2]. Depending on the GC content of long chains, melting occurs in 1 M NaCl solutions in a temperature interval between 355 to 390 K [3]. A detailed analysis of differential melting curves for intermediate length chains reveals the existence of discrete peaks corresponding to the successive opening of AT-rich regions of the chain [4]. Both standard theoretical descriptions of DNA thermal denaturation, the nearest-neighbor (NN) model of oligonucleotide melting [5,6] and the Poland-Sheraga (PS) model of polynucleotide melting [7], are based on sequence-dependent parameters describing the free energy gain per nearest neighbor pair of stacked and paired bases in the double helix. Extensive experiments have led to a unified parametrization [5] at least as far as the nearest-neighbor parameters are concerned.

As the more recent Peyrard-Bishop model [8], the standard approach does not explicitly consider the embedding of the molecule in three-dimensional space and the resulting long-range (in terms of chemical distance) interactions. The approximation made in the PS model, to account for the entropy reduction due to intrabubble [9] and interbubble [10,11] excluded volume interactions through a single universal exponent c, is still under debate [12].

Is it possible to preserve the polymer model underlying the PS approach? Explicit models of associating polymers [13–19] automatically account for the subtle (but universal) polymer contributions to the free energy. In addition, they include fluctuations in situations without predetermined secondary structure, generate an ensemble of three-dimensional structures and provide direct acess to the response to external mechanical forces [20,21]. The price one has to pay is an enormous increase in computing time required for enumeration or simulation studies. For generic lattice models current resources allow the investigation of chain lengths up to $N=20$ [13–15] and $N=10^4$, respectively [16,17]. These studies have provided insight into the closure of hairpin conformations, force-induced unzipping and the effect of chain stiffness and excluded volume interactions on the order of the melting transition.

In the present paper we show that simple lattice models of the type studied in Refs. [16,17] can *quantitatively* reproduce the melting of DNA oligomers and polymers with the same precision as the standard PS-approach without introducing a host of new adjustable parameters. The idea is (i) to parametrize the polymer model on the basis of experimentally well-studied limiting cases of the standard approach and (ii) to consider its application in more complex situations where the approximations of the PS approach are controversial. Here we concentrate on the first point. Interestingly, our results suggest a reinterpretation of boundary terms in the PS model, leading to a unified theoretical treatment of oligonucleotide melting and of domain opening in polymeric DNA.

The paper is structured as follows. In Sec. II, we review the analysis and theoretical description of experimental melting curves in the framework of the nearest-neighbor and the Poland-Sheraga model. In Sec. III we define the lattice model (Sec. III A) and use limiting cases with a straightfor-

*Email address: ralf.everaers@ens-lyon.fr

ward connection to the standard description (Sec. III B) to derive the parameters of the lattice model (Sec. III C). Results from the exact enumeration of the lattice model are presented in Sec. IV. The discussion in Sec. V focuses on the validation of the lattice model and a proposal for the modification of the PS model. We conclude in Sec. VI. In the appendix, we analyze the role of fluctuations in situations used to experimentally determine the parameters of the standard model. The resulting analytical expressions provide a bridge between numerical solutions of the PS model for medium-sized domains and our exact enumeration data.

II. THEORETICAL BACKGROUND

In the following, we treat a complex "*AB*" in equilibrium with two strands "*A*" and "*B*" consisting of $N+1$ bases or N base-pair steps, respectively. We have to consider the following.

(1) The degree of association $\Theta_{\mathrm{ass}}(c_T,T)$, Eq. (2), for the chemical equilibrium $A+B \leftrightarrow AB$ between bound and unbound states as a function of temperature T and of the total strand concentration c_T.

(2) "Internal" melting of the complex and the individual strands, i.e., the numbers $n_{\epsilon,A}(T)$, $n_{\epsilon,B}(T)$, $n_{\epsilon,AB}(T)$ and fractions $\Theta_{\mathrm{int},A}(T)=n_{\epsilon,A}(T)/(N/2)$, $\Theta_{\mathrm{int},B}(T)=n_{\epsilon,B}(T)/(N/2)$, and $\Theta_{\mathrm{int},AB}(T)=n_{\epsilon,AB}(T)/N$, of bound base-pair steps as a function of temperature.

The experimentally observable overall fraction of bound base pair steps derives from the degree of internal melting of the three components weighted by their respective concentrations

$$\Theta(c_T,T)=\frac{1}{2}\Theta_{\mathrm{int},A}(T)[1-\Theta_{\mathrm{ass}}(c_T,T)]+\frac{1}{2}\Theta_{\mathrm{int},B}(T)\times[1-\Theta_{\mathrm{ass}}(c_T,T)]+\Theta_{\mathrm{int},AB}(T)\Theta_{\mathrm{ass}}(c_T,T), \quad (1)$$

where the total, strand, and complex concentrations are related via $c_T=c_A(T)+c_B(T)+2c_{AB}(T)$, $c_A(T)=c_B(T)=[1-\Theta_{\mathrm{ass}}(T)]c_T/2$, and $c_{AB}(T)=\Theta_{\mathrm{ass}}(T)c_T/2$.

New quenching techniques [22–24] allow one to measure Θ and Θ_{ass} separately. For the purposes of the present paper, it is sufficient to consider simple limiting cases where melting is either due to strand separation or domain opening. Most of the theoretical results are standard, except for our use of the entropy of mixing (4), the approximations Eqs. (A14)–(A17) for the effective cooperativity parameter observed in experiments on medium-sized bubble opening, and the estimate Eq. (A20) of the maximum total strand concentration up to which the PS model reduces to the NN model for oligomer melting.

A. Law of mass action, entropy of mixing

The degree of association $\Theta_{\mathrm{ass}}(c_T,T)$ given by

$$\Theta_{\mathrm{ass}}(x)=1+x-\sqrt{x(2+x)} \quad (2)$$

with $x=\frac{c_0}{c_T}\exp\left(\frac{\Delta F_0}{k_BT}\right)$ derives from the law of mass action

$$\frac{(c_A/c_0)(c_B/c_0)}{c_{AB}/c_0}=\exp\left(\frac{\Delta F_0}{k_BT}\right), \quad (3)$$

where c_0 is a reference concentration (usually $c_0=1$ M) and ΔF_0 the free energy difference between the bound and unbound forms at the reference concentration. ΔF_0 equals the difference $F_{AB}-F_A-F_B$ of the (internal) free energies of the double and single stranded forms up to an additive constant and changes in an obvious manner, if a different reference concentration is used. Using the entropy of mixing between a large molecule and the solvent (see, e.g., Ref. [25]) and assuming that the molecular volumes in the complex simply add up $v_{AB}=v_A+v_B=2v_A$, we find

$$\Delta F_0=F_{AB}-F_A-F_B-k_BT\ln(c_0v_{AB}e) \quad (4)$$

$$=\Delta F_{\mathrm{int}}-k_BT\ln(1.75N). \quad (5)$$

B. Analysis of melting curves

Experimental melting curves are often analyzed assuming two-state transitions between (ensembles of) microstates and characterized by temperature independent enthalpy and entropy differences $\Delta H_{(0)}$ and $\Delta S_{(0)}$ [5]. In the case of oligomer melting $\Theta_{\mathrm{int},A}=\Theta_{\mathrm{int},B}=0$, and $\Theta_{\mathrm{int},AB}=1$ together with Eq. (2) and $\Delta F_0(T)=\Delta H_0-T\Delta S_0$ can be solved for the melting temperature implicitly defined via $\Theta(T_m)\equiv 1/2$:

$$T_m=\frac{\ln \Delta H_0}{\Delta S_0+k_B\ln(c_T/c_0/4)}. \quad (6)$$

There are two ways to determine ΔH_0 and ΔS_0: either from individual melting curves by using

$$\Delta H_0=6k_BT_m^2\Theta'(T_m), \quad (7)$$

$$\Delta S_0=\Delta H_0/T_m \quad (8)$$

or via Eq. (6) by measuring melting temperatures at two different concentrations. Agreement within 10% is usually considered as confirmation of the two-state character of the transition.

In the case of internal two-state melting, the probability of the closed state is given by

$$\Theta_{\mathrm{closed}}=\frac{1}{1+\exp[(\Delta H-T\Delta S)/k_BT]}. \quad (9)$$

The melting temperature is implicitly defined as $\Theta_{\mathrm{closed}}\equiv 1/2$ or

$$\Delta H=T_m\Delta S. \quad (10)$$

If the closed state is fully paired, $\Theta_{\mathrm{int,closed}}=1$ and the open state fully unpaired, $\Theta_{\mathrm{int,open}}=0$, the closing probability equals the observable pairing probability. The melting curve and the differential melting curve, $d\Theta/dT\equiv\Theta'(T)$, are then given by

$$\Theta(T)=\frac{1}{1+\exp[(\Delta H-T\Delta S)/k_BT]}, \quad (11)$$

$$\Theta'(T) = \frac{\Delta H}{4k_B T^2} \Big/ \cosh\left[\frac{\Delta H}{2k_B T}\left(1 - \frac{T}{T_m}\right)\right], \tag{12}$$

$$\Theta'(T_m) = \frac{\Delta H}{4k_B T_m^2}. \tag{13}$$

For ideal two-state melting $T^2\Theta'(T)$ has a minimum at T_m. Thus, once again one can check the two-state character of the transition by comparing results obtained for two independent criteria for the melting temperature.

C. Nearest-neighbor model

In the nearest-neighbor (NN) model [5] the formation of double-stranded complexes is the consequence of a sequence-dependent free-energy gain $\Delta h_{\mathrm{NN}} - T\Delta s_{\mathrm{NN}}$ per stacked pair of neighboring base pairs. Furthermore, there are initiation free energies $\Delta h^0_{\mathrm{ini}} - T\Delta s^0_{\mathrm{ini}}$ for the two ends of the double-stranded complex:

$$\Delta G_0 = (\Delta h^0_{\mathrm{ini},1} + \Delta h^0_{\mathrm{ini},2} + \Delta H_{\mathrm{NN}}) - T(\Delta s^0_{\mathrm{ini},1} + \Delta s^0_{\mathrm{ini},2} + \Delta S_{\mathrm{NN}}), \tag{14}$$

$$\Delta H_{\mathrm{NN}} = \sum_{i=1}^{N} \Delta h_{\mathrm{NN},i}, \tag{15}$$

$$\Delta S_{\mathrm{NN}} = \sum_{i=1}^{N} \Delta s_{\mathrm{NN},i}. \tag{16}$$

An alternate source of information about the nearest-neighbor terms are melting experiments with very long chains where the translational entropy gain of single strands plays no role. Both routes have converged to comparable values for the 10 NN dimer duplexes with Watson-Crick base pairs [5]. The values (as well as the initiation free energies in the case of oligonucleotide melting) are sequence dependent and correlated with GC content $0 \leqslant f_{\mathrm{GC}} \leqslant 1$. Our Fig. 1 illustrates this point (and the remaining experimental uncertainties) by a comparison of estimates for melting temperatures based on oligonucleotide [26], domain [27], and very long chain [3] data.

D. Poland-Sheraga model

While some salient features of DNA melting are thus relatively easy to grasp in the limits of short and very long chains, the intermediate case turns out to be considerably more complex. Experimentally observed differential melting curves exhibit a number of discrete peaks corresponding to the successive opening, fusion, and enlargement of AT-rich sections of the chain [2]. In particular, the melting temperature of a particular domain is significantly higher if positioned as an interior "loop domain" than as an "end domain."

The Poland-Sheraga (PS) model describes DNA as an alternating sequence of double stranded (helical) parts and (coiled) loops. The free energy of a particular sequence is calculated from the nearest-neighbor model and usually given relative to the double stranded ground state. As a consequence, double stranded segments contribute a factor of 1 to the statistical weight of a DNA conformation. The dominant contribution of molten sections is due to the familiar comparison of the free energy of double stranded helices and isolated single strands. In the case of a frayed chain end, one usually writes

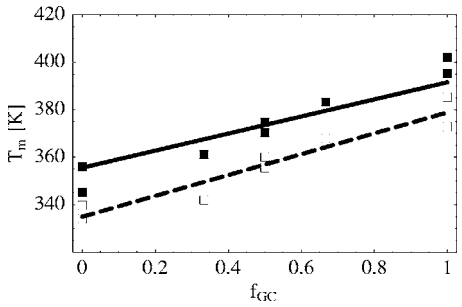

FIG. 1. DNA melting temperatures as a function of GC content f_{GC} at two different salt concentrations. Filled symbols are calculated from $T_m = \Delta h_{\mathrm{NN}}/\Delta s_{\mathrm{NN}}$ for oligonucleotide NN parameters [26] for 1 M salt solutions. Open symbols are calculated from $T_m = \Delta h_{\mathrm{NN}}/\Delta s_{\mathrm{NN}}$ for the MELTSIM [27] polymer NN parameters for 0.0745 M salt solutions. The lines represent the empirical relation for the melting temperatures of very long chains [3] $T_m = T_M^{AT} + f_{\mathrm{GC}}(T_M^{\mathrm{GC}} - T_M^{AT})$ with $T_M^{AT} = \{355.55 + 7.9\ \ln([\mathrm{Na}^+])\}$ K and $T_M^{\mathrm{GC}} = \{391.55 + 4.89\ \ln([\mathrm{Na}^+])\}$ K. Data points are shown for eight linear combinations [28] of the ten Watson-Crick dinucleotide steps.

$$Z^{\mathrm{PS}}_{\mathrm{end}}(N) = \exp\left(\sum_{i=1}^{N} \frac{\Delta h_i - T\Delta s_i}{k_B T}\right) \tag{17}$$

without any further corrections, while the constraint that the two strands of a molten domain in the interior of the complex have to form a closed loop leads to an entropy reduction

$$Z^{\mathrm{PS}}_{\mathrm{loop}}(N) = \sigma N^{-c} \exp\left(\sum_{i=1}^{N} \frac{\Delta h_i - T\Delta s_i}{k_B T}\right). \tag{18}$$

The cooperativity parameter $\sigma \approx 10^{-5}$ in Eq. (18) needs to be determined experimentally [4,29]. The loop exponent c has to be derived from polymer physics [7,9–11] and determines the order of the melting transition [30]. Current estimates for c vary between $c = 1.75$ for opening of isolated loops [9] and $c = 2.15$ for simultaneous opening of many bubbles close to the overall melting temperature [11]. Due to the long-ranged power-law correction to the loop entropy, the PS-model is more difficult to solve than the (otherwise similar) one-dimensional Ising model. For particular sequences one has to rely on efficient [31] numerical schemes as implemented, for example, in the freely available MELTSIM code [27]. There are also a few analytical results for homogeneous sequences [7,32–34]. In the Appendix, we solve the PS model in the zero- and the one-bubble approximations, respectively, for three situation considered in this article: oligomer melting, opening of end domains and opening of loop domains.

III. MODEL AND METHOD

In the following we introduce a lattice model with short-range interactions as a complement to the two standard de-

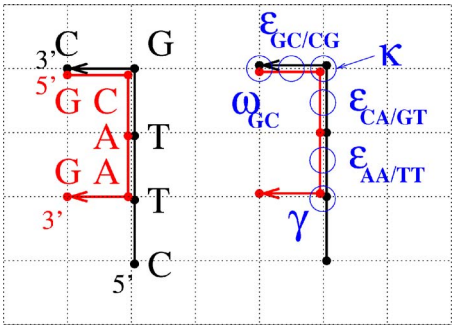

FIG. 2. (Color online) Illustration of the lattice model defined in the text. On the LHS we show a partially bound complex formed by two antiparallel DNA strands; the RHS illustrates the various contributions to the free energy associated with the complex.

scriptions of DNA melting. We derive all DNA specific parameters of the lattice model from the nearest-neighbor and the Poland-Sheraga descriptions.

(i) We translate the sequence dependent free-energy gain $\Delta h_{\mathrm{NN}} - T\Delta s_{\mathrm{NN}}$ per stacked pair of neighboring base-pairs to a pairing free energy for overlapping chain sections in the lattice model.

(ii) We express the initiation terms $\Delta h^0_{\mathrm{ini}} - T\Delta s^0_{\mathrm{ini}}$ as the sum of the entropy of mixing at standard conditions and a (capping) free energy penalty for the ends of double-helical chain sections.

(iii) We interpret the PS cooperativity parameter σ for loop formation in terms of an interfacial (forking) free energy between coiled and helical sections of a double-stranded complex.

The lattice model has no adjustable parameter corresponding to the loop exponent c, because it implicitely accounts for generic polymer contributions to the entropy loss due to loop closure, as well as intraloop and interloop excluded volume interactions. In an explicit polymer model, cooperativity generated through a forking free energy is the only choice compatible with Eq. (18). Note that the value of this *local* interfacial energy cannot depend on the *global* shape of the coiled section, i.e., on whether or not the coiled section forms a loop or an end domain. This will lead us to reconsider the absence of prefactors in Eq. (17) in the standard formulation of the PS model.

The section is organized as follows. We define the lattice model in Sec. III A and consider a number of useful limiting case in Sec. III B. In Sec. III C we present our parametrization of the model. In Sec. III D we briefly discuss the enumeration methods used to study oligomer and domain melting.

A. Definition of the lattice model

We use a variant of the lattice model proposed by Causo *et al.* [16]. DNA strands are modeled as self- and mutually avoiding walks on a simple cubic lattice (Fig. 2). We associate the position of bases with the lattice sites. Two walks are allowed to overlap on a lattice site, if and only if they can form a Watson-Crick base-pair (anti-parallel orientation, complementary base). We assign a free energy

$$\epsilon(T) = \epsilon_H - T\epsilon_S \tag{19}$$

to a pair of stacked neighboring Watson-Crick base-pairs, i.e., to doubly occupied *bonds* on the lattice. Note that the lattice model is a coarse-grained description of DNA. Integrating out the microscopic degrees of freedom of the molecule and the solvent leads to a temperature dependent Hamiltonian for the remaining degrees of freedom. Similarly to Eq. (14) for the NN model, we only consider the simplest functional form of this temperature dependence.

In addition to the pairing energy, we allow for a capping free energy

$$\omega(T) = \omega_H - T\omega_S \tag{20}$$

at free ends of paired strands and an interfacial or forking free energy

$$\gamma(T) = \gamma_H - T\gamma_S \tag{21}$$

between paired and unpaired sections in the complex. Finally, we include a free energy penalty

$$\kappa(T) = \kappa_H - T\kappa_S \tag{22}$$

for bending of double-helical chain sections.

Relating the lattice model to the PS model is, in principle, simple, since both are defined on the same length scale and have similar adjustable parameters. Essentially, we have to count and group microstates of the lattice model and match the corresponding contributions to the partition function of the PS model term by term. Stated differently, we invert the line of argument that led to the PS model to parametrize the lattice model.

B. Limiting cases

1. Double strands

The free energy of the unique (straight) ground state of a double strand is given by

$$H_{Ds} = N\epsilon_H + 2\omega_H, \tag{23}$$

$$S_{Ds} = N\epsilon_S + 2\omega_S. \tag{24}$$

Double stranded DNA is a fairly stiff molecule with a temperature dependent Kuhn length $l_K(T)$ of the order of $l_K(300\ \mathrm{K}) = 300$ bp. In order to match this stiffness with a nonreversal random walk on a simple cubic lattice (excluded volume effects may be neglected in this case), we have to suppress bending with a Boltzmann factor of

$$\exp\left(-\frac{\kappa(T)}{k_B T}\right) = \frac{1}{2[l_K(T)/\mathrm{bp} - 1]}. \tag{25}$$

In principle, bending contributes a multiplicative factor of

$$\exp\left(\frac{2(N-1)}{l_K(T)/\mathrm{bp}}\right) \tag{26}$$

to the partition function of a double stranded segment of N bonds. In practice, the corresponding contribution to free energy is of order of 1% and can be neglected for the parametrization of the model.

2. Single strands

The number of SAWs as a function of their length N has the asymptotic form $\mu^N N^{\gamma-1}$ with $\gamma=1.16$ in three dimensions and $\mu=4.68$ for the cubic lattice [35]. In our model, the partition function for single strands is approximately given by

$$Z_{SS} = 0.2\mu^N N^{\gamma-1}, \tag{27}$$

$$H_{SS} = 0, \tag{28}$$

$$S_{SS} = k_B[N\ln(\mu) + (\gamma-1)\ln(N) - 1.6], \tag{29}$$

where we have determined an overall prefactor by fitting enumeration data for walks up to length $N=22$.

3. Bubbles

The number of self-avoiding polygons as a function of their length $2N$ has the asymptotic form $\mu^{2N}N^{-3\nu}$ with $\nu=0.588$ in three dimensions and $\mu=4.68$ for the cubic lattice [35]. In present case, the partition function for a bubble may be approximated by

$$Z_{\text{loop}} = \frac{8}{3} \times 0.16\mu^{2N}N^{-3\nu}\exp\left(-\frac{2\gamma(T)}{k_BT}\right), \tag{30}$$

$$H_{\text{loop}} = 2\gamma_H, \tag{31}$$

$$S_{\text{loop}} = 2\gamma_S + k_B[2N\ln(\mu) - 3\nu\ln(N) - 0.84], \tag{32}$$

where we have determined an overall prefactor by fitting enumeration data for polygons up to length $N=22$ [36]. The additional prefactor of 8/3 accounts approximately for the number of possibilities to attach straight (double stranded) sections at opposite sites of the polygon.

4. Open ends

We now apply the same logic to a situation where a double stranded complex opens from one end. The number of conformations for N open base-pair steps equals the number of SAWs of length $2N$ multiplied (approximately) by a factor of 4 for the orientation of the double stranded stem.

$$Z_{\text{end}} = 4 \times 0.2\mu^{2N}(2N)^{\gamma-1}\exp\left(-\frac{\gamma(T)}{k_BT}\right), \tag{33}$$

$$H_{\text{end}} = \gamma_H, \tag{34}$$

$$S_{\text{end}} = \gamma_S + k_B[2N\ln(\mu) + (\gamma-1)\ln(2N) - 0.11]. \tag{35}$$

For a closed end or $N=0$ one has to replace $\gamma_{H/S}$ in the above expression by $\omega_{H/S}$ and the factor of 4 for the orientation of the stem by a factor of 6. Note that standard PS theory does not account for the interfacial energy. The power law correction was recently introduced in a theoretical study [33].

C. Parametrization

The pairing and end capping free energy of the lattice model can be determined using experimental data for oligomer or polymer melting [5]. In order to do so, we (i) assume a two-state melting transition between a fully paired double stranded complex and two single strands and (ii) expand the free energy difference for the lattice model (including our estimate for the entropy of mixing) to linear order in N around a typical experimental value $N=10$. The resulting expressions for initialization and nearest-neighbor entropies and enthalpies can be matched directly with tabulated values for the NN model [5]

$$\epsilon_H = \Delta h_{\text{NN}}, \tag{36}$$

$$\epsilon_S = \Delta s_{\text{NN}} + 2k_B\ln(\mu) = \Delta s_{\text{NN}} + 3.0k_B, \tag{37}$$

$$\omega_H = \Delta h_{\text{ini}}, \tag{38}$$

$$\omega_S = \Delta s_{\text{ini}} - 2.3k_B. \tag{39}$$

To parametrize the interfacial energy between coiled and helical regions we have to match terms with the Poland-Sheraga theory. In the case of the opening of a single small or medium sized bubble the loop exponent $c=3\nu$ is noncontroversial and Eq. (18) has to match the corresponding expression (30) for the lattice model with energies calculated relative to the double stranded ground state

$$\exp\left(\sum_{i=1}^{N}\frac{\epsilon_{H,i} - T\epsilon_{S,i}}{k_BT}\right)Z_{\text{loop}} = \left[\frac{8}{3}\times 0.16\exp\left(-\frac{2\gamma(T)}{k_BT}\right)\right]N^{-3\nu} \times\left[\mu^{2N}\exp\left(\sum_{i=1}^{N}\frac{\epsilon_{H,i} - T\epsilon_{S,i}}{k_BT}\right)\right].$$

The power law corrections are obviously identical and the nearest-neighbor terms agree by construction (37) and (36), so that

$$\sigma \equiv 0.45\exp\left(-\frac{2\gamma(T)}{k_BT}\right). \tag{40}$$

In the literature [4,27,29], the PS cooperativity parameter is given as a temperature and sequence independent number. In the present context, we see no principle reason why $\gamma(T)$ should have a simpler structure than $\omega(T)$. From a practical point of view, however, fluctuations (see the Appendix) and the smaller accessible temperature interval probably render the experimental determination of these dependencies even more difficult for $\gamma(T)$ than for $\omega(T)$. More interesting than these details is a comparison of the order of magnitude of the forking and capping energies. For a typical value of $\sigma \approx 0.9\times 10^{-5}$

$$\frac{\gamma(370\text{ K})}{k_B 370\text{ K}} = -\frac{1}{2}\ln(\sigma/0.45) = 5.3 \pm 0.3 \tag{41}$$

is larger, but comparable to [26]

$$\frac{\omega_{AT}(370\text{ K})}{k_B 370\text{ K}} = 3.4 \pm 1.8, \tag{42}$$

$$\frac{\omega_{GC}(370\ \mathrm{K})}{k_B 370\ \mathrm{K}} = 3.1 \pm 1.5. \tag{43}$$

As we will show, these features have important consequences for the behavior of the lattice model.

Finally, we assume with no good reason that the origin of the bending rigidity is purely energetic $l_K(T) = (300\ \mathrm{K}/T)300$ bp and

$$\kappa = 1900 k_B\ \mathrm{K}. \tag{44}$$

D. Exact enumeration

We study three different situations using enumeration techniques: (1) two strands originating from a common origin, (2) two strands originating from a rigid, semi-infinite, double-stranded chain, and (3) two strands originating from and terminating in two nonintersecting, rigid, semi-infinite double-stranded chains. For each microstate Γ we count the number of stacked base-pair pairs $n_\epsilon(\Gamma)$ of interfaces between double and single stranded segments $n_\gamma(\Gamma)$, of free ends of double stranded segments $n_\omega(\Gamma)$, and of kinks in double stranded segments $n_\kappa(\Gamma)$. The partition function, free energy, and degree of pairing of double-stranded complexes are calculated by summing over all microstates Γ with at least one stacked base-pair pair $n_\epsilon(\Gamma) > 0$:

$$H(\Gamma,T) = n_\epsilon(\Gamma)\epsilon(T) + n_\gamma(\Gamma)\gamma(T) + n_\omega(\Gamma)\omega(T) + n_\kappa(\Gamma)\kappa(T), \tag{45}$$

$$Z(T) = \sum_\Gamma \exp[-H(\Gamma,T)/(k_B T)], \tag{46}$$

$$F(T) = -k_B T \ln[Z(T)], \tag{47}$$

$$\langle n_\epsilon(T)\rangle = \frac{\sum_\Gamma n_\epsilon(\Gamma)\exp[-H(\Gamma,T)/(k_B T)]}{Z(T)}. \tag{48}$$

In the case of oligonucleotide denaturation, melting curves are calculated from Eq. (1) using, respectively, Eqs. (2) and (4) for the chemical equilibrium and Eqs. (28) and (29) for the free energy of single strands. The use of enumeration data for the scenario (1) in this case involves a small approximation; in principle one should enumerate a series of complex conformations where the first paired bond is located at positions $i = 1, \ldots, N$. Here we include specific binding by only allowing monomers of the same index to overlap on a lattice site, but neglect the heterogeneity of the pairing energy [16]. However, we emphasize that the exclusion of secondary structure (hairpin) formation in single strands is *not* a limitation of the lattice model as such, but a justified simplification for the oligonucleotide sequences used for the parametrization of the nearest-neighbor model. For the data analysis we follow the experimental practice to analyze melting curves assuming two state transitions between (ensembles of) microstates and characterized by temperature independent enthalpy and entropy differences $\Delta H_{(0)}$ and $\Delta S_{(0)}$.

IV. RESULTS

A. Oligomer melting

A well understood aspect of DNA melting is the thermal denaturation of oligonucleotides. In this limit, the transition is controlled by the competition between the translational entropy of separated strands and the free energy gain due to pairing and stacking in the double stranded complex. The experimental evidence [5,6] is in quantitative agreement with a two-state transition Eq. (6) with concentration independent free energy differences $\Delta H_0 - T\Delta S_0$ between the double stranded and single stranded forms described by the nearest-neighbor model Eqs. (14)–(16) [5,6]. Thus, oligonucleotides with up to a dozen base pairs typically dissociate at temperatures where partially bound intermediate states play no role.

In Fig. 3 we compare results derived from our enumeration data for the lattice model to the predictions from the nearest-neighbor model using the standard parameters ($\Delta h_{AA/TT} = -3979 k_B$ K, $\Delta s_{AA/TT} = -11.18 k_B$, $T_{m,\infty} = \Delta h_{AA/TT}/\Delta s_{AA/TT} = 356$ K, $\Delta h_{\mathrm{ini}wA/T} = 1159 k_B$ K, $\Delta s_{\mathrm{ini}wA/T} = 2.07 k_B$) for DNA in 1 M NaCl solution from Allawi and SantaLucia [26]. The top row shows differential melting curves, followed by the chain length dependence of the values for $T_m(N)$, $\Delta H(N)$, and $\Delta S(N)$ derived from a two-state melting analysis of the peak positions and heights. Results for the lattice model shown on the left-hand side (LHS) of Fig. 3 were calculated using our parameterization from Sec. III C. As assumed by the NN-model, the peaks in the differential melting curves for strand association $d\Theta_{\mathrm{ass}}(T)/dT$ and the degree of pairing in the complex $d\Theta_{AB,\mathrm{int}}(T)/dT$ are indeed well separated: Strand separation occurs at temperatures where excited states of the complex play no role. As a consequence, the measured melting temperatures T_m as well as the enthalpy and entropy differences ΔH_0 and ΔS_0 shown in the other parts of Fig. 3 are in excellent agreement with the nearest-neighbor model on which the parametrization was based.

The corresponding graphs on the right-hand side (RHS) of Fig. 3 were obtained for a parametrization where the forking energy was set to zero. In this case, partially molten states of the complex become relevant at much lower temperatures. But while the peak in $d\Theta_{AB,\mathrm{int}}(T)/dT$ broadens and shifts to lower temperatures, there is an antagonistic effect on the association equilibrium. The access to additional, partially unbound states lowers the free energy of the complex, so that the strand dissociation occurs at higher temperatures. The combined effect on the observed melting temperatures is moderate, but results for $\Delta H_0(N)$ and $\Delta S_0(N)$ obtained at different total strand concentrations show clear indications of non-two-state melting.

B. Domain melting

The opening of isolated, medium-sized (a few dozen to a few hundred bps) AT-rich domains is a standard test case of the PS model and of particular interest for the experimental determination of the cooperativity parameter σ. Two cases need to be distinguished: Opening of an *end* domain flanked on *one* side by a GC-rich barrier domain and opening of an

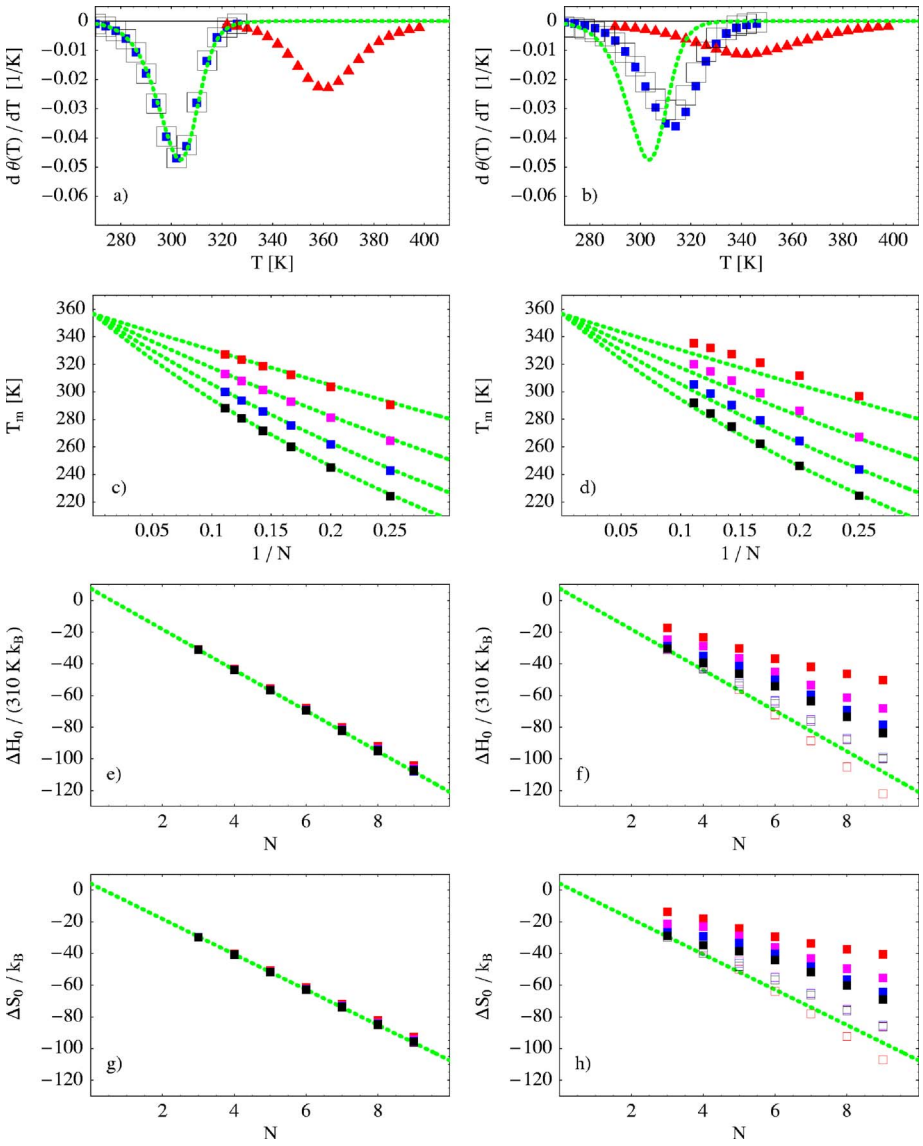

FIG. 3. (Color online) Melting of poly-A/poly-T oligonucleotides. SantaLucia [26] nearest neighbor model (lines) in comparison to data points calculated from exact enumeration results for the lattice model using the corresponding parametrization. Left column: including our free energy penalty for forking; right column: neglecting this term as in current PS parametrizations. (a), (b) Differential melting curves for oligomers of length $N=7$ at a total strand concentration of $c_T=10^{-4}$ M: degree of pairing in the complex $\Theta_{\text{int}}(T)$ (triangles), degree of strand association $\Theta_{\text{ass}}(T)$ (filled boxes), total degree of pairing $\Theta(T)$ (open boxes). (c), (d) Chain length dependence of the melting temperature T_m for total strand concentration of (from top to bottom) $c_T=10^{-2}$ M, $c_T=10^{-4}$ M, $c_T=10^{-6}$ M, $c_T=10^{-8}$ M. Panels (e)–(h) show the corresponding results of the two-state melting analysis for ΔH_0 and ΔS_0; filled symbols indicate data from the analysis of individual melting curves Eqs. (7) and (8); open symbols represent results from a comparison of melting temperatures at different concentrations, Eq. (6). Agreement of the two data sets is a consistency check for the assumed two-state character of the transition at a given concentration.

internal (loop) domain flanked on *both* sides by barrier domains with higher melting temperatures. In Fig. 4, we compare a number of approximate solution of the PS-model derived in the Appendix to data for the lattice model (RHS of Fig. 4) and results obtained from MELTSIM (LHS of Fig. 4) using the standard parameters ($\Delta h_{AA/TT}=-4257k_B$ K, $\Delta s_{AA/TT}=-12.52k_B$, $T_{m,\infty}=\Delta h_{AA/TT}/\Delta s_{AA/TT}=340$ K, $T_{m,\infty}=\Delta h_{GG/CC}/\Delta s_{GG/CC}=373$ K, $c=1.75$, $\sigma=0.9\times10^{-5}$) for DNA in 0.0745 M NaCl-solution. The theoretical expressions used in the two parts of the figure are identical. Results were plotted separately due to the difference in the accessible domain sizes.

From top to bottom the four rows in Fig. 4 show differential melting curves and values for $T_m(N)$, $\Delta H(N)$, and $\Delta S(N)$ derived from a two-state melting analysis of the peak positions and heights. The most important qualitative observation concerns the domain size dependence of the melting temperature. For internal domains T_m shows a strong increase with inverse domain size. For end domains T_m is essentially independent of domain size. The comparison to the theoretical expressions shows that, in contrast to oligomer melting, fluctuations may not be neglected in the analysis of domain opening. The results Eqs. (A14)–(A17) of the one-bubble approximation for loop opening derived in the Appendix are in excellent agreement with both data sets. Similarly, we find excellent agreement between the results Eqs. (A9)–(A11) of the zero-bubble approximation for end-domain opening and the MELTSIM data. In contrast, there are small, but visible discrepancies between the results for the lattice model and the theoretical expressions. Note, however, that following the MELTSIM convention, the theory does *not* account for interfacial and end capping energies.

V. DISCUSSION

In this paper we add a lattice model to the two standard descriptions of the thermal denaturation of DNA, the nearest-neighbor model and the Poland-Sheraga model. All three descriptions are intimately related. Both the PS model and the lattice model include the effect of fluctuations in the calculation of experimental observables. Both models derive their parameters from the NN model, the main difference being

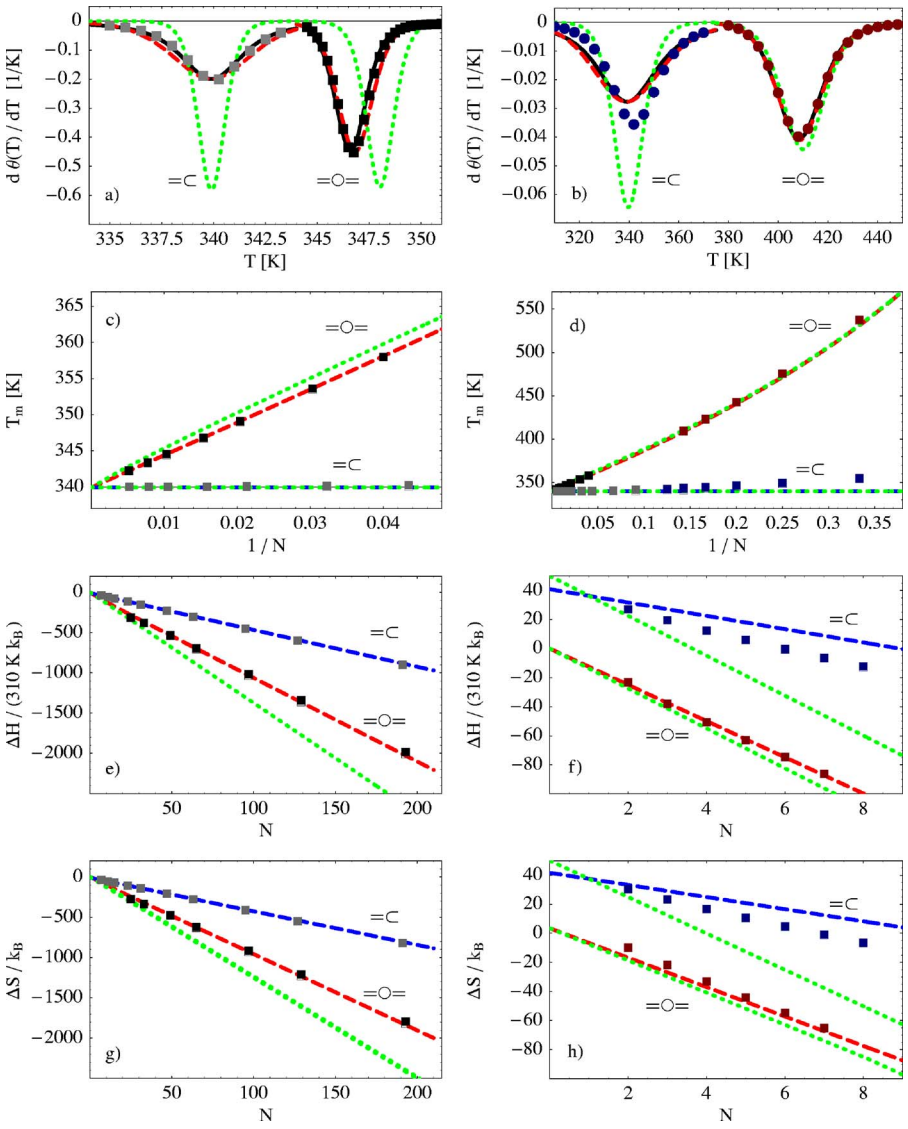

FIG. 4. (Color online) Opening of AT loop (=○=) and end (=⊂) domains under MELTSIM [27] standard conditions. Symbols left column: Results obtained from the MELTSIM [27] numerical solution of the PS theory. Symbols right column: Results from the enumeration of the lattice model. Solid lines: Exact summation of the zero- and one-bubble approximations (A8) and (A13); dashed lines: two-state approximation accounting for fluctuation (A9), (A10), (A16), and (A17); dotted lines: two-state approximation neglecting fluctuation Eqs. (A1), (A2), (A4), and (A5). Top row: differential melting curves for domain sizes (a) $N=64$ and (b) $N=7$, respectively; AT-domain-size dependence of (c), (d) melting temperature, (e), (f) apparent ΔH, and (g), (h) apparent ΔS. Data for end-domain opening in panels (f) and (h) were shifted by 50 units for better visibility.

that the lattice model assigns loop cooperativity to a large extent to a free energy penalty for "forking." In contrast to the current parametrization of the PS model, this free energy penalty (as well as the capping free energy) also affects chain opening from the ends. In this section, we address the following questions.

(1) Does the lattice model agree with the experimental evidence as well as the standard approach?

(2) Why has the introduction of the forking energy practically no effect in the case of end domain melting?

(3) Can we modify the PS model to obtain a unified description?

(4) Is there experimental evidence that the modified PS model offers a superior description of DNA?

(5) If so, what is the use of the lattice model itself?

A. Validation of the lattice model

Our results for oligomer melting in Fig. 3 were calculated for temperature and concentration ranges as well as chain lengths ($N+1 \leqslant 10$ bp) corresponding to the experimental situation under which the nearest-neighbor model was parameterized. The theoretical curves may thus be taken as a faithful representation of experimental data which could, in principle, be included directly into the graphs. The perfect agreement with the results for the lattice model displayed on the LHS of Fig. 3 constitutes a first validation of our parametrization presented in Sec. III C.

In the case of the opening of an isolated internal domain, there is again excellent agreement between our results for the lattice model and the analytical solution of the PS model (Fig. 4). In particular, we correctly reproduce the experimentally observed cooperativity parameter for loop opening with our choice of the forking free energy (40). However, due to the limited domain sizes accessible in enumeration studies, these results cannot be compared directly to experiment. In practice, it is impossible to observe the opening of loop domains whose melting temperature is too close to (or even higher than) the melting temperature of the barrier domains. For the MELTSIM curves we found that discrete peaks corresponding to the melting of an internal AT domain could be reliably isolated for $N \geqslant 20$ in good agreement with direct experimental observations [23,24]. We therefore used semi-

infinite, rigid double-stranded barriers to measure loop opening for the lattice model.

On first sight, the corresponding results for end domain opening (blue symbols and corresponding lines in Fig. 4) seem slightly disappointing. The agreement with the theoretical melting curve is not as perfect as for oligomer and loop melting, but still fair with nearly coinciding melting temperatures and a 20% error in the peak height. On closer inspection, however, the agreement appears remarkably good, given that (i) the case of end domain opening was not considered for the parametrization of the lattice model, and (ii) the theoretical curves are calculated following the usual approximation (17) in Poland-Sheraga models to neglect intradomain excluded volume interactions, forking and end capping energies included in the lattice model, (33).

The reason why the considerable free energy penalties for forking and end capping in the lattice model have such a small effect on the opening of end domains is easy to understand: if both terms are *identical,* they contribute a constant and inconsequential prefactor to the PS partition function (A7). For our parametrization the difference is less than $2k_BT$ and plays no role for larger domains where the contribution of the ground state to the partition function becomes negligible in spite of its slightly increased statistical weight. We note further that the difference between the interfacial and the capping energies (40)–(43) for the lattice model is comparable to the experimental error.

Experiments on end-domain opening [4] have focused on the position dependence of the melting temperature of a given domain, while the peak heights have not been tested with comparable rigor (see, for example, Fig. 8 in Ref. [4]). Since both models make practically indistinguishable predictions for $T_m(N) \approx T_m(\infty)$, further experiments would seem to be needed to resolve the remaining, small discrepancy.

B. Fluctuations, forking, and the PS model

Given the negligible effect in polymeric DNA, why do we think that it is worthwhile to include the forking free energy at fraying chain ends into the PS model? After all, the NN and the PS model do already describe experimental data for oligomer and domain melting within a few percent accuracy [6,27] and with closely agreeing NN parameters Δh_{NN} and Δs_{NN} (see Fig. 1 and Ref. [5]).

One problem of the standard approach becomes apparent, if one considers oligomer melting within the more general PS model. This may appear futile, since experiments show that fluctuations may be neglected. However, self-consistency requires this irrelevance to be a property of the correctly parametrized model including fluctuations—as in the case of our lattice model (LHS of Fig. 3). In the Appendix we show that the standard parametrization of the PS model does not fulfill this requirement. The predicted upper concentration limit (A20) for two-state melting of 10^{-N} M for oligomers of length N is far below the experimental range.

It is instructive to compare the effect of the boundary terms for end(-domain) opening and oligomer melting. A huge forking energy suppresses internal fluctuations, entailing domain-size dependent melting temperatures for end-domains as well as perfect NN two-state melting for long chains. Equal forking and capping energies have no effect on $\Theta_{\mathrm{int},AB}(T)$, but do change the association equilibrium Eq. (2) of oligonucleotides. Our example of a lattice model without forking energy (RHS of Fig. 3) shows a behavior close to the predictions of the standard PS model. Finally, forking energies which are much smaller than the capping energy lead to a situation where the lowest energy state has unbound terminal base pair steps. While this has again negligible consequences for end domain opening, such a model would not reduce to the NN model for oligomer melting in *any* limit. Thus, even if a model is parametrized on the basis of the NN model for oligonucleotide melting, it may fail to reproduce the targeted theory in the experimentally relevant range.

To obtain a unified theoretical description of the thermal denaturation of DNA, we now reverse the logic of Sec. III C and derive a PS-like description for our lattice model. By construction, Eq. (18) for open loop domains as well as the factor of 1 for double-stranded domains in the interior of the chain remain unaffected. However, Eq. (17) for a frayed chain end is replaced by

$$Z_{\mathrm{end}}^{\mathrm{PS}} = \exp\left(\sum_{i=1}^{N} \frac{\epsilon_{H,i} - T\epsilon_{S,i}}{k_BT}\right) Z_{\mathrm{end}}$$
$$= \left[4 \times 0.2 \times 2^{\gamma-1} \exp\left(-\frac{\gamma(T)}{k_BT}\right)\right] N^{\gamma-1}$$
$$\times \left[\mu^{2N} \exp\left(\sum_{i=1}^{N} \frac{\epsilon_{H,i} - T\epsilon_{S,i}}{k_BT}\right)\right].$$

This expression is based on Eq. (33) for the lattice model with energies calculated relative to the double stranded ground state. Using Eqs. (37), (36), and (40) this further reduces to

$$Z_{\mathrm{end}}^{\mathrm{PS}} = \sigma_E N^{\gamma-1} \exp\left(\sum_{i=1}^{N} \frac{\Delta h_i - T\Delta s_i}{k_BT}\right), \tag{49}$$

$$\sigma_E = 0.0043 \tag{50}$$

in the usual PS notation with the power law correction introduced in Ref. [33]. Furthermore, due to the capping energy Eqs. (38) and (39) in the lattice model, free ends of double stranded domains contribute an extra factor of

$$Z_{\mathrm{cap}}^{\mathrm{PS}} = 0.1 \exp\left(-\frac{\Delta h_{\mathrm{ini}}^0 - T\Delta s_{\mathrm{ini}}^0}{k_BT}\right) \tag{51}$$

to the partition function of the complex. By construction, this modified version of the PS model reproduces the results of the lattice model and hence the experimentally observed two-state melting of oligomers. Integrating this modification along with Eqs. (1), (2), and (4) into numerical solutions of the PS model [27,31,37] might open the possibility to describe the various experiments discussed in this article within a single theoretical framework.

C. Cooperativity parameter and loop exponent

Another, more fundamental problem of the standard approach is the determination of appropriate values for the cooperativity parameter σ and the loop exponent c. For example, one could argue [38] that the recent progress in determining an effective single-loop exponent of $c=2.15$ [10,11] calls for a reparametrization of the cooperativity parameter used in implementations of the Poland-Sheraga model. As shown in the Appendix, the shape of the melting curve fir medium-sized loop domains depends only on the combination σN^{-c} (i.e., the loop correction for the dominant loop size in the PS partition function). In good agreement with the observations in Ref. [38], this argument suggests for typical loop sizes an increase of σ by a factor of $400^{2.15-1.764} \approx 10$ as a consequences of changing the loop exponent c from its traditional value of $c=1.746$ to the current best estimate of $c=2.15$.

However, there are obvious problems in using $c=2.15$ indiscriminately. For example, it is hard to see why the "many-bubble" exponent should be applied to the (effectively two-state) opening of individual, medium sized domains or how experimental data for this situations could be used to determine the proper value of the cooperativity parameter for situations which are indeed described by $c=2.15$. Thus it is doubtful, if sequence-dependent stepwise melting can be properly described with a single pair of values for c and σ. In Ref. [30] Fisher concluded that "it seems very difficult to treat this problem theoretically in any realistic way." Borrowing these words, the the present paper presents an attempt to formulate a model which allows a quantitative, realistic, albeit *numerical* treatment of these issues. To parametrize the lattice model, we need the microscopic parameters of the Poland-Sheraga model. The analysis of previous simulations of the generic lattice model [16,17,33] has already provided evidence for the emergence of the renormalized PS behavior [10,11] in sufficiently long chains close to the melting temperature.

VI. SUMMARY AND CONCLUSION

To summarize, we have presented a minimal, semiquantitative lattice model for studying the (thermal) denaturation of double stranded DNA. The model describes the molecule on the same length scale as the standard nearest-neighbor model for oligomer melting and the Poland-Sheraga model of polymer melting. In the present paper we have focused on the correct parametrization of the model using experimentally well studied limiting cases where the assumptions of the standard approach are non-controversial. We have validated our results (i) by calculating melting curves for oligomer and domain melting from exact enumeration data of the lattice model and (ii) by subjecting these numerical data to the standard analysis for experimental data.

An interesting consequence of our effort to parameterize the lattice model is our proposal for the further unification of the theoretical description of oligomer and polymer melting. In a first step, we have expressed the initiation terms in the nearest-neighbor model for oligonucleotides as the sum of the entropy of mixing at standard conditions and a free energy penalty for the ends of double-helical chains. In a second step, we have interpreted the Poland-Sheraga cooperativity parameter σ for loop formation in terms of an interfacial (forking) free energy between coiled and helical sections of a double-stranded complex. Quite interestingly, both free energy penalties turn out to be of comparable magnitude. Their inclusion has negligible consequences for end domain melting on long chains, but is essential to suppress fluctuations in the case of oligomer melting.

The most serious shortcoming of the present lattice model is probably the suppression of the helical character of double stranded segments, i.e., it is not possible to study twist induced denaturation and (transient) effects of super-coiling on the melting transition [39–41]. Nevertheless, there is an obvious interest in applying the same logic to the problem of RNA [42,43] and DNA [44] folding as well as to the parameterization of more microscopic (lattice) models [18] which try to distinguish (hydrogen-bond) base pairing from intrastrand and interstrand stacking interactions. The latter should allow for the inclusion of residual stacking interactions of neighboring bases in single-stranded DNA [45] and could provide a microscopic interpretation for the forking energy [46]. A further challenge is the simultaneous description of experiments on kinked or nicked double-stranded DNA [47]. It remains to be seen, up to which point lattice models keep their intrinsic advantages over off-lattice representation along the lines of Refs. [19,48], if further and further microscopic details are included.

In conclusion, we have extended and unified the theoretical description of DNA melting by a simple, semiquantitative lattice model. The main difference between the lattice model and the formulation of Poland and Sheraga is the preservation of the embedding of the molecules in three dimensional space. Clearly, the lattice model is not worth the high computational price for the well-behaved limiting cases we have studied in the present paper. For future work, we nevertheless see a number of intrinsic advantages over the traditional formulation: The inclusion of chain stiffness and of the generic long-range excluded volume interactions in partially molten complexes [16,17], the straightforward extension to situations where external mechanical forces couple to the shape of the molecule [14,15], the absence of any *a priori* restrictions on the secondary structure of partially molten complexes, the inclusion of fluctuations in situations where ground states with different secondary structure compete, and the possibility to study simplified kinetics [18].

ACKNOWLEDGMENTS

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

APPENDIX: APPROXIMATE SOLUTIONS OF THE POLAND-SHERAGA MODEL

For short chains (i.e., in the absence of fluctuations) the PS model reduces to a nearest neighbor description. The long chain limit can, in general, only be treated numerically [27,31,37]. For parametrization purposes it is interesting to

consider "medium-sized" oligonucleotides or domains of a few dozen to a few hundred bp's, for which the PS model can be solved in the zero- and the one-bubble approximation, respectively. The results provide insight into the role of fluctuations and allow to determine the conditions under which a nearest-neighbor description is appropriate.

1. Nearest-neighbor model of domain opening

Ignoring fluctuations, domain opening is described by a two-state transition between the double stranded ground state and the fully opened domain. For the standard PS model the condition $\Delta G(T_m) \equiv 0$ yields for end domains

$$\Delta H = \Delta H_{\mathrm{NN}}, \tag{A1}$$

$$\Delta S = \Delta S_{\mathrm{NN}}, \tag{A2}$$

$$T_m = \frac{\Delta H_{\mathrm{NN}}}{\Delta S_{\mathrm{NN}}}. \tag{A3}$$

For internal domains the corresponding estimates are given by [4]

$$\Delta H = \Delta H_{\mathrm{NN}}, \tag{A4}$$

$$\Delta S = \Delta S_{\mathrm{NN}} - k_B \ln(\sigma N^{-c}), \tag{A5}$$

$$T_m = \frac{\Delta H_{\mathrm{NN}}}{\Delta S_{\mathrm{NN}} - k_B \ln(\sigma N^{-c})}. \tag{A6}$$

Numerical values obtained from these sets of equations for the MELTSIM parameters are shown as dotted (green) lines in Fig. 4. Agreement with the MELTSIM data is fair (in particular, for the melting temperatures) but not quantitative [29].

2. Zero-bubble approximation for end-domain opening

Obviously, fluctuations in the degree of domain opening are already relevant in the experimentally accessible range of domains of a few dozen to a few hundred bps. Nevertheless, some simplifications are possible due to the small value of the cooperativity parameter. For domains of a few hundred base-pairs DNA conformations with extra bubbles may be safely neglected. Here we refer to these domains as "medium-sized" in contrast to "small" (negligible fluctuations) and "large" (fluctuation dominated).

For a homogeneous, medium-sized end domain it is possible to calculate the resulting simplified expressions for the partition function and the degree of pairing

$$Z(T) = \sum_{n=0}^{N} a^n = \frac{a^{N+1} - 1}{a - 1}, \tag{A7}$$

$$\Theta_{\mathrm{int}}(T) = \frac{1}{NZ(T)} \sum_{n=0}^{N} (N-n) a^n = \frac{1}{N} \left(\frac{1}{a-1} - \frac{1+N}{a^{N+1}-1} \right), \tag{A8}$$

where $a = \exp\left(\frac{\Delta h - T\Delta s}{k_B T}\right)$. In a second step, one obtains from the position and height of the peak in the differential melting curve

$$\Delta H = \frac{N+2}{3N} \Delta H_{\mathrm{NN}}, \tag{A9}$$

$$\Delta S = \frac{N+2}{3N} \Delta S_{\mathrm{NN}}, \tag{A10}$$

$$T_m = \frac{\Delta H_{\mathrm{NN}}}{\Delta S_{\mathrm{NN}}}. \tag{A11}$$

Figure 4 shows excellent agreement between MELTSIM data and the combination of Eqs. (A9)–(A11) with Eq. (9) for a two-state melting transition.

3. One-bubble approximation for loop-domain opening

Using Eq. (18) for the single loop partition function, the corresponding expressions for homogeneous internal domains

$$Z(T) = 1 + \sum_{n=2}^{N} (N+1-n) Z_{\mathrm{loop}}(n), \tag{A12}$$

$$\Theta_{\mathrm{int}}(T) = \frac{N + \sum_{n=2}^{N} (N-n)(N+1-n) Z_{\mathrm{loop}}(n)}{NZ(T)} \tag{A13}$$

can still be summed exactly, but with a rather lengthy result. In this case, it is instructive to consider a further approximation: a two-state transition between the double stranded ground state and an *ensemble* of PS microstates where the domain is almost fully opened. The resulting refined estimate of the melting temperature

$$T_m = \frac{\Delta H_{\mathrm{NN}}}{\Delta S_{\mathrm{NN}} - k_B \ln(\tilde{\sigma} N^{-c})} \tag{A14}$$

can be expressed in terms of an *effective* cooperativity parameter

$$\tilde{\sigma} = \frac{\sigma}{[1 - (\sigma N^{-c})^{1/N}]^2}. \tag{A15}$$

Similarly, the apparent entropy and enthalpy differences extracted from a two-state melting analysis are given by

$$\frac{\Delta H}{\Delta H_{\mathrm{NN}}} = 1 - \frac{4}{N[(\tilde{\sigma} N^{-c})^{-1/N} - 1]}, \tag{A16}$$

$$\Delta S = \Delta H / T_m. \tag{A17}$$

Figure 4 shows that these expressions are again in excellent agreement with the results extracted from the MELTSIM data.

4. Zero-bubble approximation for Oligomer melting

Treating oligomer melting within the PS model one may again safely neglect bubble opening and restrict the partition function to states where the complex opens from the two ends

$$Z(T) = \sum_{n=0}^{N-1} (n+1)a^n = \frac{1 + a^N[N(a-1)-1]}{(a-1)^2}, \quad \text{(A18)}$$

$$\Theta_{\text{int}}(T) = \frac{1}{NZ(T)} \sum_{n=0}^{N-1} (N-n)(n+1)a^n = \frac{1}{NZ(T)} \frac{a\{2+N+a^N[N(a-1)-2]\}-N}{(a-1)^3} \quad \text{(A19)}$$

with $a = \exp\left(\frac{\Delta h - T\Delta s}{k_B T}\right)$ as in Eqs. (A7) and (A8). For the experimentally observed NN model to hold, excited states have to play a minor role. Limiting their contribution to the partition function Eq. (18) of the complex in the standard PS-model to 20%, we find $a_{\max} < 0.09$ or

$$c_{T,\max} = 4a_{\max}^N c_0 \approx 4 \times 10^{-N} \text{ M} \quad \text{(A20)}$$

as an upper limit for the total strand concentrations. Under MELTSIM standard conditions $a_{\max} < 0.09$ corresponds to melting temperatures below 285 K. Under the conditions where most experiments are carried out, Eqs. (A18) and (A19) combined with Eq. (1) predict a complicated scenario, where Θ_{ass} is *increased* and Θ_{int} is reduced compared to the predictions and assumptions of the NN model. Note that the inclusion of sequence heterogeneity further reduces the estimated validity range of the NN model.

[1] C. Calladine, H. Drew, B. Luisi, and A. Travers, *Understanding DNA; The Molecule and How it Works*, 3rd ed. (Elsevier Academic Press, Amsterdam, 2004).
[2] R. Wartell and A. Benight, Phys. Rep. **126**, 67 (1985).
[3] M. Frank-Kamenetskii, Biopolymers **10**, 2623 (1971).
[4] R. D. Blake and S. G. Delcourt, Nucleic Acids Res. **26**, 3323 (1998).
[5] J. SantaLucia, Proc. Natl. Acad. Sci. U.S.A. **95**, 1460 (1998).
[6] R. Owczarzy, P. M. Vallone, F. J. Gallo, T. M. Paner, M. J. Lane, and A. S. Benight, Biopolymers **44**, 217 (1998).
[7] D. Poland and H. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers; Statistical Mechanical Theory of Order-disorder Transitions in Biological Macromolecules* (Academic Press, New York, 1970).
[8] M. Peyrard and A. R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989).
[9] M. Fisher, J. Chem. Phys. **45**, 1469 (1966).
[10] Y. Kafri, D. Mukamel, and L. Peliti, Phys. Rev. Lett. **85**, 4988 (2000).
[11] Y. Kafri, D. Mukamel, and L. Peliti, Eur. Phys. J. B **27**, 135 (2002).
[12] A. Hanke and R. Metzler, Phys. Rev. Lett. **90**, 159801 (2003).
[13] S. Kumar, D. Giri, and Y. Singh, Europhys. Lett. **70**, 15 (2005).
[14] S. Kumar, D. Giri, and S. M. Bhattacharjee, Phys. Rev. E **71**, 051804 (2005).
[15] D. Giri and S. Kumar, Phys. Rev. E **73**, 050903(R) (2006).
[16] M. S. Causo, B. Coluzzi, and P. Grassberger, Phys. Rev. E **62**, 3958 (2000).
[17] E. Carlon, E. Orlandini, and A. L. Stella, Phys. Rev. Lett. **88**, 198101 (2002).
[18] M. Sales-Pardo, R. Guimera, A. A. Moreira, J. Widom, and L. A. N. Amaral, Phys. Rev. E **71**, 051902 (2005).
[19] C. Hyeon and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **102**, 6789 (2005).
[20] B. EssevazRoulet, U. Bockelmann, and F. Heslot, Proc. Natl. Acad. Sci. U.S.A. **94**, 11935 (1997).
[21] C. Danilowicz, V. W. Coljee, C. Bouzigues, D. K. Lubensky, D. R. Nelson, and M. Prentiss, Proc. Natl. Acad. Sci. U.S.A. **100**, 1694 (2003).
[22] A. Montrichok, G. Gruner, and G. Zocchi, Europhys. Lett. **62**, 452 (2003).
[23] Y. Zeng, A. Montrichok, and G. Zocchi, Phys. Rev. Lett. **91**, 148101 (2003).
[24] Y. Zeng, A. Montrichok, and G. Zocchi, J. Mol. Biol. **339**, 67 (2004).
[25] M. Rubinstein and R. Colby, *Polymer Physics* (Oxford University Press, Oxford, 2003).
[26] H. T. Allawi and J. SantaLucia, Biochemistry **36**, 10581 (1997).
[27] R. D. Blake, J. W. Bizzaro, J. D. Blake, G. R. Day, S. G. Delcourt, J. Knowles, K. A. Marx, and J. SantaLucia, Bioinformatics **15**, 370 (1999).
[28] A. Vologodskii, B. Amirikyan, Y. Lyubchenko, and M. Frank-Kamenetskii, J. Biomol. Struct. Dyn. **2**, 131 (1984).
[29] B. R. Amirikyan, A. V. Vologodskii, and Y. L. Lyubchenko, Nucleic Acids Res. **9**, 5469 (1981).
[30] D. Poland and H. Scheraga, J. Chem. Phys. **45**, 1456 (1966).
[31] M. Fixman and J. Freire, Biopolymers **16**, 2693 (1977).
[32] T. Garel and H. Orland, cond-mat/0304080.
[33] L. Schäfer, cond-mat/0502668.
[34] C. Richard and A. J. Guttmann, J. Stat. Phys. **115**, 925 (2004).
[35] C. Vanderzande, *Lattice Models of Polymers* (Cambridge University Press, Cambridge, 1998).
[36] P. Butera and M. Comi, Phys. Rev. B **60**, 6749 (1999).
[37] T. Garel and H. Orland, Biopolymers **75**, 453 (2004).
[38] R. Blossey and E. Carlon, Phys. Rev. E **68**, 061911 (2003).
[39] S. Cocco and R. Monasson, Phys. Rev. Lett. **83**, 5178 (1999).
[40] M. Barbi, S. Cocco, M. Peyrard, and S. Ruffo, J. Biol. Phys. **24**, 97 (1999).
[41] T. Garel, H. Orland, and E. Yeramian, cond-mat/0407036.
[42] D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner, J. Mol. Biol. **288**, 911 (1999).
[43] P. Schuster, Rep. Prog. Phys. **69**, 1419 (2006).
[44] J. SantaLucia and D. Hicks, Annu. Rev. Biophys. Biomol. Struct. **33**, 415 (2004).
[45] V. Ivanov, Y. Zeng, and G. Zocchi, Phys. Rev. E **70**, 051907 (2004).
[46] V. Ivanov, D. Piontkovski, and G. Zocchi, Phys. Rev. E **71**, 041909 (2005).
[47] P. Yakovchuk, E. Protozanova, and M. Frank-Kamenetskii, Nucleic Acids Res. **34**, 564 (2006).
[48] B. Mergell, M. R. Ejtehadi, and R. Everaers, Phys. Rev. E **68**, 021911 (2003).