Coupling between denaturation and chain conformations in DNA: stretching, bending, torsion and finite size effects



2009 J. Phys.: Condens. Matter 21 034104

(http://iopscience.iop.org/0953-8984/21/3/034104)

IOP PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **21** (2009) 034104 (18pp) doi:10.1088/0953-8984/21/3/034104

# Coupling between denaturation and chain conformations in DNA: stretching, bending, torsion and finite size effects

**Manoel Manghi, John Palmeri and Nicolas Destainville**

Laboratoire de Physique Théorique, Université de Toulouse, CNRS, 31062 Toulouse, France

E-mail: manghi@irsamc.ups-tlse.fr, palmeri@irsamc.ups-tlse.fr and destain@irsamc.ups-tlse.fr

Received 30 May 2008, in final form 7 August 2008
Published 17 December 2008
Online at stacks.iop.org/JPhysCM/21/034104

**Abstract**

We develop further a statistical model coupling denaturation and chain conformations in DNA (Palmeri *et al* 2007 *Phys. Rev. Lett.* **99** 088103). Our discrete helical wormlike chain model takes explicitly into account the three elastic degrees of freedom, namely stretching, bending and torsion of the polymer. By integrating out these external variables, the conformational entropy contributes to bubble nucleation (opening of base-pairs), which sheds light on the DNA melting mechanism. Because the values of monomer length, bending and torsional moduli differ significantly in dsDNA and ssDNA, these effects are important. Moreover, we explore in this context the role of an additional loop entropy and analyze finite size effects in an experimental context, where polydA–polydT is clamped by two G–C strands, as well as for free polymers.

(Some figures in this article are in colour only in the electronic version)

## 1. Introduction

The study of the physical properties of DNA is seeing intense activity from both a theoretical [1–18] and an experimental perspective [19–27]. The first theoretical and experimental studies were published several decades ago, but the recent development of experimental techniques enabling one to address DNA properties at the single molecule level has brought a significant renewal of interest in the field. These techniques provide not only average properties like their former bulk counterparts, but also the statistics of fluctuations around the average values. Single molecule setups range from magnetic and optical tweezers [28, 29] or tethered particle motion apparatus [30–33], to atomic force microscopy [24, 34]. They give access to huge amounts of data concerning DNA physical properties such as bending, stretching and twisting elasticities or conformational dynamics [30, 35–37, 31]. In parallel, the genomic revolution leads to the elucidation of a large number of biological functions involving nucleic acids. A pressing demand has followed for reliable and precise physical models, able to validate the many hypotheses emerging from molecular biology or microscopy experiments. This constitutes a double motivation for theoreticians to refine the existing microscopic DNA models: accounting for the new, accurate physics experiments, and validating (or invalidating) the physical assumptions underlying the proposed biological mechanisms.

Denaturation is one of the intimate DNA physical features that are supposed to be involved in many critical cellular functions, such as transcription, replication and protein binding, but are not fully understood. Even though DNA unwinding at the cellular level is generally an active process due to enzymes consuming energy, such as helicases [38], understanding the subtle statistical mechanics of this biopolymer is an essential first step towards the elucidation of more complex, active mechanisms. Furthermore, the spontaneous opening of base-pairs due to thermal activation is likely to play a direct role in several biological processes. Recently, Yan and Marko [12] have, for example, proposed that coupling the DNA elasticity to a minimal model of base-pair melting can account for the increased cyclization probability observed by Cloutier and Widom [39]: even if it is rare, local denaturation increases short-range flexibility because single strand DNA (ssDNA) is nearly two orders of magnitude more

 

flexible than double strand DNA (dsDNA). This increased flexibility should play a role everywhere the polymer must be bent or looped on length scales shorter than its persistence length (typically equal to 50 nm). In the nucleosome, it is twisted around histones, the diameters of which are about 11 nm [40].

In order to get more insight into this coupling between denaturation and elasticity, we recently proposed a more refined coupled, nonlinear model, where the internal states of base-pairs (open or closed) are described by a one-dimensional Ising model, whereas the chain configurations are encoded by a one-dimensional Heisenberg one, taking into account DNA bending [17, 18]. By solving exactly this model, we demonstrated that taking into account this coupling between internal and external degrees of freedom enables the prediction of the modifications of elastic properties when increasing the temperature: Ising parameters are renormalized by temperature in such a way that DNA denaturation is accompanied by a collapse of the chain persistence length. Following this route, we were able for the first time to write the melting temperature $T_m$ as a function of microscopic parameters only—when it was a fit parameter in previous models—and to give a new description of boundary and finite size effects.

However, our model was minimal in the sense that only bending was taken into account. Torsion is also known to play a role on elasticity because a strong flexion of an elastic rod is in general accompanied by a torsion [41] which decreases the energy cost of the deformation. Similarly, stretching of base-pairs ought to be included in a complete elastic model. In the present paper, we systematically explore these effects in detail, by proposing an exactly solvable discrete helical wormlike chain (DHWC) model, and predicting how Ising parameters are renormalized in this context (section 2).

In section 3, we investigate the influence of the chain length (or finite size effects) on melting profiles. At the experimental level, it has been shown in [1, 42] that they are measurable even for DNA made of several thousand base-pairs. These effects are usually measured for polydA–polydT flanked by more stable G–C rich strands. Hence we modify our model to account for such clamped boundary conditions. In other models of denaturation [2, 42, 43], chain configurations are partially incorporated via a so-called 'loop entropy' that takes into account the entropic cost of closing a denaturation bubble when it is not located at a polymer end. We investigate the role of loop entropy in finite clamped and free DNA chains.

## 2. Coupling between internal and external DNA degrees of freedom

In [17, 18], we showed that the denaturation melting temperature emerges naturally by taking into account the difference in bending rigidities of ssDNA sequences (bubbles) and dsDNA ones. Indeed, the ratio of both moduli, $\kappa_{ds}/\kappa_{ss}$ is of the order of 50. It is at the origin of an entropic barrier which stems from the fact that in the ssDNA state, the allowed spatial configurations for unit tangent vectors $\hat{\mathbf{t}}_i$, which describe the chain conformations, are much more numerous, leading to a significant increase in entropy. More precisely, it has been shown that the free energy (mostly of entropic nature) coming out by integrating the Hamiltonian part that depends on the external variables $\hat{\mathbf{t}}_i$, renormalizes the bare Ising parameters, $K$ and $J$, which are the energy costs of creating a domain wall and destacking two adjacent base-pairs respectively. The third Ising parameter, $\mu$, which corresponds to the energy required to break a base-pair (or 'magnetic field' in a magnetism analogy), is not renormalized. In particular, the full penalty of breaking one base-pair located in DNA's interior, $L = \mu + K$, becomes

$$L_0 = \mu + K - \frac{k_B T}{2}\left[G_0\left(\frac{\kappa_{ds}}{k_B T}\right) - G_0\left(\frac{\kappa_{ss}}{k_B T}\right)\right]$$
$$\simeq \mu + K - \frac{k_B T}{2}\ln\left(\frac{\kappa_{ds}}{\kappa_{ss}}\right) \qquad \text{for } \kappa \gg k_B T \tag{1}$$

where $k_B T$ is the thermal energy and $G_0(x) = x - \ln(\frac{\sinh x}{x})$. The approximation is valid in the temperature range of interest since $\kappa_{ss} \approx 6\, k_B T$.

In the infinitely long chain limit, the melting temperature $T_m$, defined as the temperature at which half of the base-pairs are broken, is simply given by $L_0(T_m) = 0$. The melting temperature thus naturally emerges in this model and is determined by the competition between the enthalpic cost of breaking base-pairs (mostly hydrogen bonds and $\pi$-overlap of carbon ring wavefunctions of adjacent nucleotides but also charge, dipolar and van der Waals interactions) and the entropic gain in nucleating bubbles made of very flexible single-stranded DNA chains.

However, other external variables than $\hat{\mathbf{t}}_i$, which also characterize the chain elasticity, may lead to a renormalization of the parameter $L$. Clearly, two other external degrees of freedom should also be taken into account:

- many force–extension experiments have shown that the monomer size $a$ is not the same in dsDNA and ssDNA (see the review [44] and references therein). Indeed, the monomer size in the B-form of double-stranded DNA is generally defined as the rise along the central axis per base-pair which is $a_{ds} = 0.34$ nm. The generally accepted value [12, 45] of the monomer size in ssDNA is $a_{ss} = 0.71$ nm and we choose in the following $a_{ss} \approx 2\, a_{ds}$.[1]
- the B-form of dsDNA is the famous double helix and a torsional energy has to be taken into account in a more refined model. Indeed, in the continuous helical wormlike chain model for DNA [47], the elastic energy of the chain has two contributions: a bending term already taken into account in [17, 18] and an energy of torsional deformations which in the continuum limit reads

$$\mathcal{E}_{twist} = \frac{C}{2}\int \Omega_3^2(s)\, ds \tag{2}$$

where $\Omega_3 = \boldsymbol{\Omega}\cdot\hat{\mathbf{e}}_3$. The Darboux vector $\boldsymbol{\Omega}$ characterizes the rotation of the material frame, $\hat{\mathbf{e}}_3$ is along the molecular axis and $s$ is the curvilinear index. The twist

[1] The rotation per base-pair (or equilibrium twist) is $\Phi_0 = 0.19\,\pi$ rad [46] (the pitch is then $p = 3.55$ nm) and the radius $r = 1.19$ nm, so one can estimate the monomer size in the fully extended ssDNA form as $a_{ss} = \Phi_0\sqrt{r^2 + (p/2\pi)^2} \approx 0.79$ nm which is larger than $a_{ds}$. It should be noticed, however, that ssDNA is usually not fully extended and has a helical shape too.

(or torsional) rigidity modulus $C$ has been measured in torsional experiments on dsDNA [48–50, 37], and is of the order of $C_{ds} \simeq 2.4$–$4.5 \times 10^{-19}$ J nm. The twist rigidity of ssDNA is lower because it loses its stiff helical structure and has been evaluated to be $C_{ss} \simeq 9 \times 10^{-20}$ J nm [51]. The ratio $C_{ds}/C_{ss}$ is of the same order as $\kappa_{ds}/\kappa_{ss}$ and will certainly modify the Ising parameters in a similar way as for the bending energy.

### 2.1. Discrete helical wormlike chain model

In the present work, the DNA is modeled as a fluctuating polymer chain in a space of three dimensions, characterized by the *external chain variables*, the set of $N$ bond vectors $\mathbf{t}_i$ and their orientation in space (it is thus implicitly assumed that the monomer has a three-dimensional structure) and an *internal Ising variable* $\sigma_i = \pm 1$ which models the internal state of dsDNA, unbroken (U) or broken (B) respectively. The modeling of the base-pair internal state by an Ising model was developed in the 1960s by Lehman, Montroll and Vedenov (see review [52] and references therein).

We focus on the coupling of the internal variables with the external variables which is included in the Hamiltonian part treating the fluctuating chain. A material coordinate frame is defined for each monomer $i$, $\{\hat{\mathbf{e}}_{\mu,i}\}_{\mu=1,2,3} = \{\hat{\mathbf{u}}_i, \hat{\mathbf{n}}_i, \hat{\mathbf{t}}_i\}$, where $\hat{\mathbf{t}}_i$ is the unit bond vector $\mathbf{t}_i = \mathbf{R}_{i+1} - \mathbf{R}_i = t_i\hat{\mathbf{t}}_i$ and the two other unit vectors are in the directions of the principal axes of inertia. This triad is defined with respect to a fixed referential $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$ through a rotation matrix $\mathbf{A}_i$ characterized by Euler angles $\boldsymbol{\omega}_i = (\alpha_i, \beta_i, \gamma_i)$. The evolution of the triad along the molecular chain from monomer $i$ to monomer $i+1$ is obtained by a rotation, also defined by Euler angles $(\phi_{i,i+1}, \theta_{i,i+1}, \psi_{i,i+1})$

$$\hat{\mathbf{e}}_{\mu,i+1} = \mathbf{\Lambda}_{\mu\nu}(\phi_{i,i+1}, \theta_{i,i+1}, \psi_{i,i+1})\hat{\mathbf{e}}_{\nu,i} \tag{3}$$

where the rotation matrix $\mathbf{\Lambda}$ is the product of three rotation matrices associated with each Euler angle, but can also be viewed as the product of two rotations of angles $\theta_{i,i+1}$ and $\phi_{i,i+1} + \psi_{i,i+1}$ [50]

$$\begin{aligned}\mathbf{\Lambda}&(\phi_{i,i+1}, \theta_{i,i+1}, \psi_{i,i+1})\\ &= R(\hat{\mathbf{t}}_i, \psi_{i,i+1})R(\hat{\mathbf{n}}_{i,i+1}, \theta_{i,i+1})R(\hat{\mathbf{t}}_i, \phi_{i,i+1})\\ &= R(\hat{\mathbf{t}}_i, \phi_{i,i+1} + \psi_{i,i+1})R(R(\hat{\mathbf{t}}_i, -\phi_{i,i+1})\hat{\mathbf{n}}_{i,i+1}, \theta_{i,i+1}).\end{aligned} \tag{4}$$

In the material coordinate frame $\{\hat{\mathbf{e}}_{\mu,i}\}$, the bond vector $\hat{\mathbf{t}}_{i+1}$ is thus defined by its spherical coordinates $(\theta_{i,i+1}, \phi_{i,i+1})$. Moreover, the Euler angles $(\phi_{i,i+1}, \theta_{i,i+1}, \psi_{i,i+1})$ which will appear in the Hamiltonian are completely determined by the two sets of Euler angles $\boldsymbol{\omega}_i$ and $\boldsymbol{\omega}_{i+1}$ through $\mathbf{\Lambda}_{i,i+1} = \mathbf{A}_{i+1} \cdot \mathbf{A}_i^{-1}$.

The configurational part of the Hamiltonian is defined as the sum of two terms

$$\mathcal{H}[\sigma, \mathbf{t}, \psi] = \mathcal{H}_{\text{Ising}}[\sigma] + \mathcal{H}_{\text{chain}}[\sigma, \mathbf{t}, \psi] \tag{5}$$

where $\mathcal{H}_{\text{Ising}}[\sigma]$ is the usual Ising Hamiltonian already defined in [17, 18] with three parameters $(\mu, J, K)$, and $\mathcal{H}_{\text{chain}}[\sigma, \mathbf{t}, \psi]$ is the DHWC Hamiltonian

$$\mathcal{H}_{\text{Ising}}[\sigma] = -\mu \sum_{i=1}^{N} \sigma_i - \sum_{i=1}^{N-1}\left[J\sigma_{i+1}\sigma_i + \frac{K}{2}(\sigma_{i+1} + \sigma_i)\right] \tag{6}$$

$$\begin{aligned}\mathcal{H}_{\text{chain}}[\sigma, \mathbf{t}, \psi] &= \frac{1}{2}\sum_{i=1}^{N}\frac{\epsilon_i}{2}\left(|\mathbf{t}_i|^2 - a_i^2\right)^2\\ &+ \tfrac{1}{2}\sum_{i=1}^{N-1}\left[\kappa_{i,i+1}(\hat{\mathbf{t}}_{i+1} - \hat{\mathbf{t}}_i)^2 + 2C_{i,i+1}(\cos\theta_{i,i+1}\right.\\ &\left. - \cos\lambda_{i,i+1})\right].\end{aligned} \tag{7}$$

The first term of (7) is a nonlinear stretching term dictated by rotational and translational invariances. The values of the Lamé coefficient $\epsilon_i$ and the monomer length $a_i$ depend on the state of the base-pair ($(\epsilon_U, a_U)$ for $\sigma_i = +1$ and $(\epsilon_B, a_B)$ for $\sigma_i = -1$). The second term corresponds to the bending and torsional energies. The latter can be written as $C_i(\text{tr}\,\mathbf{\Lambda}(0, \theta_{i,i+1}, 0) - \text{tr}\,\mathbf{\Lambda}(\phi_{i,i+1}, \theta_{i,i+1}, \psi_{i,i+1}))$ and accounts for the energy penalty associated with the twist defined by the angle $\phi_{i,i+1} + \psi_{i,i+1}$. Indeed, the angle $\lambda$ of the rotation defined in (3) is a function of $\phi + \psi$ and $\theta$ (indices $i$, $i+1$ are omitted):

$$\cos\lambda = \tfrac{1}{2}[\cos(\phi + \psi)(\cos\theta + 1) + \cos\theta - 1]. \tag{8}$$

The bending $\kappa_{i,i+1}$ and torsional $C_{i,i+1}$ moduli also vary locally with the state of nearest-neighbor links ($(\kappa_U, C_U)$ for type $U$–$U$, $(\kappa_B, C_B)$ for $B$–$B$ and $(\kappa_{UB}, C_{UB})$ for $U$–$B$). We assume in this model that all the parameters appearing in (5) are independent of the nucleotide type. Hence we focus on homopolynucleotides. The case of sequence dependent parameters could be handled numerically.

Equation (7) defines our discrete version of the continuous helical wormlike chain model first employed by Yamakawa for DNA [47] and extended in several articles in the literature [53, 50, 54, 55]. First, one observes that if there is no twist, i.e. no rotation around the tangent vector $\hat{\mathbf{t}}_i$, it imposes $\phi + \psi = 0$ and from (7) and (8), the torsional term vanishes. Hence if there is no twist along the chain (or if the DNA chain is modeled as linear), the DHWC becomes the classical discrete wormlike chain already developed in [17, 18]. Furthermore, the DHWC simplifies in the continuum limit, $x_{i+1} - x_i \to \frac{\partial x}{\partial s}\Delta s$ with $\Delta s \to 0$ where $s$ is the curvilinear index. Indeed it is straightforward to see that $\sum_{i=1}^{N-1}\kappa(\hat{\mathbf{t}}_{i+1} - \hat{\mathbf{t}}_i)^2 \to \int \kappa[\Omega_1^2(s) + \Omega_2^2(s)]\,\mathrm{d}s$ and with more algebra that $\sum_{i=1}^{N-1} C[\text{tr}\mathbf{\Lambda}(0, \theta_{i,i+1}, 0) - \text{tr}\mathbf{\Lambda}(\phi_{i,i+1}, \theta_{i,i+1}, \psi_{i,i+1})]$ simplifies into (2) where the Darboux vector is defined by $\hat{\mathbf{e}}_{\mu,i+1} - \hat{\mathbf{e}}_{\mu,i} \to \mathbf{\Omega} \times \hat{\mathbf{e}}_{\mu,i}$ and $\Omega_\mu(s) = \mathbf{\Omega}\cdot\hat{\mathbf{e}}_\mu(s)$. Finally, in the low temperature regime where the spin-wave approximation is valid ($\phi + \psi \ll 1$ and $\theta \ll 1$), bending and torsional contributions reduce to quadratic terms

$$\tfrac{1}{2}\sum_{i=1}^{N-1}\left[\kappa_{i,i+1}\theta_{i,i+1}^2 + C_{i,i+1}(\phi_{i,i+1} + \psi_{i,i+1})^2\right] + \mathcal{O}(\theta^4, \phi^4, \psi^4). \tag{9}$$

The discrete model defined by (9) has already been used in the context of DNA supercoiling [56].

### *2.2. Stretching contribution to the entropy of bubble nucleation*

The first stretching term in (7) is local without any coupling between the nearest neighbors. Therefore it can be integrated out easily. The Lamé elastic constant $\epsilon$ is very large for DNA molecules: $\epsilon a^3$ has been evaluated as 8.4 nN for ssDNA by fitting force–extension experimental curves using *ab initio* calculations [45][2], and one can expect the same order of magnitude for dsDNA. Therefore, $\epsilon a^3 \gg k_B T/a \simeq 4$ pN and the saddle-point approximation applied below is valid.

By expanding the first term of (7) and writing $|\mathbf{t}_i| = a_i + \delta_i$ we have

$$(|\mathbf{t}_i|^2 - a_i^2)^2 = (|\mathbf{t}_i| + a_i)^2 (|\mathbf{t}_i| - a_i)^2 \approx 2a_i^2 (|\mathbf{t}_i| - a_i)^2 + \mathcal{O}(\delta_i^3). \tag{10}$$

The elastic term of the Hamiltonian (5) simplifies into

$$\mathcal{H}_{\text{chain}}[\sigma, \mathbf{t}, \psi] \simeq \sum_{i=1}^{N-1} \frac{\epsilon_i a_i^2}{2} (|\mathbf{t}_i| - a_i)^2 + \kappa_{i,i+1}(1 - \cos\theta_{i,i+1}) + C_{i,i+1}(\cos\theta_{i,i+1} - \cos\lambda_{i,i+1}). \tag{11}$$

The configurational part of the partition function is

$$\mathcal{Z} = \sum_{\{\sigma_i\}} \mathrm{e}^{-\beta \mathcal{H}_{\text{Ising}}[\sigma]} \int \left( \prod_{i=1}^{N-1} \frac{\mathrm{d}^3\mathbf{t}_i \, \mathrm{d}\gamma_i}{8\pi^2 a_0^3} \right) \mathrm{e}^{-\beta \mathcal{H}_{\text{chain}}[\sigma, \mathbf{t}, \psi]} \tag{12}$$

where $\gamma_i$ is the second twist Euler angle of $\mathbf{t}_i$ with respect to the reference frame and $a_0$ is a normalization length. By using the decomposition of $\mathbf{t}_i$ in spherical coordinates, $(t_i, \alpha_i, \beta_i)$, one has $\mathrm{d}^3\mathbf{t}_i \, \mathrm{d}\gamma_i = t_i^2 \, \mathrm{d}t_i \sin\alpha_i \, \mathrm{d}\alpha_i \, \mathrm{d}\beta_i \, \mathrm{d}\gamma_i \equiv \mathrm{d}t_i \, \mathrm{d}^3\boldsymbol{\omega}_i$ and the partial partition function for the chain is

$$\mathcal{Z}_{\text{chain}}[\sigma] = \prod_{i=1}^{N} \int_0^\infty \frac{t_i^2 \, \mathrm{d}t_i}{a_0^3} \mathrm{e}^{-\frac{\beta\epsilon_i a_i^2}{2}(t_i - a_i)^2} \times \int \prod_{i=1}^{N} \left( \frac{\mathrm{d}^3\boldsymbol{\omega}_i}{8\pi^2} \right) \mathrm{e}^{-\beta \mathcal{H}_{\text{angle}}[\sigma, \boldsymbol{\omega}]} \tag{13}$$

where $\mathcal{H}_{\text{angle}}[\sigma, \boldsymbol{\omega}]$ is the bending and torsional Hamiltonian. Using the saddle-point approximation for the stretching integral, we get in the large stretching constant limit

$$\prod_{i=1}^{N} \int_0^\infty \frac{t_i^2 \, \mathrm{d}t_i}{a_0^3} \mathrm{e}^{-\frac{\beta\epsilon_i a_i^2}{2}(t_i - a_i)^2} \approx \prod_{i=1}^{N} \sqrt{\frac{2\pi}{\beta\epsilon_i}} \frac{a_i}{a_0^3} \equiv \mathrm{e}^{-\sum_i^N \ln \Lambda_i}. \tag{14}$$

As explained above, we assume that the stretching energy has two competitive minima for dsDNA and ssDNA. In our model it means that the elastic constant $\epsilon_i$ and the monomer size $a_i$ have two different values whether the monomer is in the unbroken ($\sigma_i = 1$) or broken state ($\sigma_i = -1$). Hence, once integrated over the local $t_i$ variables, the stretching energy part can be included in the Ising part of the Hamiltonian to get an effective Ising Hamiltonian with a renormalized $\mu$. Indeed, by defining $\ln \Lambda_i = \delta\mu \, \sigma_i + \Gamma$ where $\delta\mu = \ln(\frac{\Lambda_U}{\Lambda_B})$ and $\Gamma = \ln(\Lambda_U \Lambda_B)$, the *renormalized* temperature dependent chemical potential is

$$\mu_0 = \mu - k_B T \ln \left( \frac{a_B}{a_U} \sqrt{\frac{\epsilon_U}{\epsilon_B}} \right) \tag{15}$$

where the correction accounts for the entropic gain when the monomer state changes. It has two contributions:

(i) in the broken state, the monomer size is greater, $a_B \approx 2a_U$, which implies a larger volume in the phase space and thus an increase in entropy;
(ii) in the case of different elastic constants, $\epsilon_U \neq \epsilon_B$, since the stretching energy $\langle E \rangle = \frac{1}{2} k_B T$ is independent of these constants, the elastic free energy difference is purely of entropic origin, similarly to the simple Einstein model for solids.

In the present case, the elastic constants $\epsilon_U$ and $\epsilon_B$ are unknown. Although several experimental studies seem to show that the stretching constant of dsDNA is larger than for ssDNA [57, 58], we have not been able to find reliable values. If, for example, we assume them equal, then the chemical potential $\mu$ is lowered by 0.5–1 $k_B T$, which is non-negligible.

### *2.3. Bending and torsional contributions*

In this section, we focus on the partition function integrated over the angles ($\mathrm{d}^3\boldsymbol{\omega}_i = \sin\alpha_i \, \mathrm{d}\alpha_i \, \mathrm{d}\beta_i \, \mathrm{d}\gamma_i$). The full partition function (12) can be written as

$$\mathcal{Z} = \sum_{\{\sigma_i\}} \mathrm{e}^{-\beta \mathcal{H}_{\text{Ising},0}[\sigma]} \int \left( \prod_{i=1}^{N} \frac{\mathrm{d}^3\boldsymbol{\omega}_i}{8\pi^2} \right) \times \mathrm{e}^{-\beta \sum_{i=1}^{N-1} \kappa_{i,i+1}(1 - \cos\theta_{i,i+1}) + C_{i,i+1}(\cos\theta_{i,i+1} - \cos\lambda_{i,i+1})} \tag{16}$$

where $\mathcal{H}_{\text{Ising},0}$ is the same as (6) with $\mu$ replaced by $\mu_0$ given in (15). Similarly to the discrete wormlike chain model [17, 18], the partition function for the coupled system can be calculated using transfer matrix techniques. For example, we have

$$\mathcal{Z} = \sum_{\{\sigma_i\}} \prod_{i=1}^{N} \int \frac{\mathrm{d}^3\boldsymbol{\omega}_i}{8\pi^2} \langle V | \sigma_1 \rangle \langle \sigma_1 | \hat{P}(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) | \sigma_2 \rangle \cdots \langle \sigma_{N-1} | \times \hat{P}(\boldsymbol{\omega}_{N-1}, \boldsymbol{\omega}_N) | \sigma_N \rangle \langle \sigma_N | V \rangle, \tag{17}$$

where the matrix elements of the transfer kernel that appears $N - 1$ times in (17), are given by (the tilde means in units of $k_B T$)

$$\langle +1 | \hat{P}(\boldsymbol{\omega}_i, \boldsymbol{\omega}_{i+1}) | +1 \rangle = \mathrm{e}^{\tilde{\kappa}_U(\cos\theta_{i,i+1} - 1) + \tilde{C}_U(\cos\theta_{i,i+1} - \cos\lambda_{i,i+1}) + \tilde{J} + \tilde{K} + \tilde{\mu}_0} \tag{18a}$$

$$\langle -1 | \hat{P}(\boldsymbol{\omega}_i, \boldsymbol{\omega}_{i+1}) | -1 \rangle = \mathrm{e}^{\tilde{\kappa}_U(\cos\theta_{i,i+1} - 1) + \tilde{C}_U(\cos\theta_{i,i+1} - \cos\lambda_{i,i+1}) + \tilde{J} - \tilde{K} - \tilde{\mu}_0} \tag{18b}$$

$$\langle +1 | \hat{P}(\boldsymbol{\omega}_i, \boldsymbol{\omega}_{i+1}) | -1 \rangle = \mathrm{e}^{\tilde{\kappa}_{UB}(\cos\theta_{i,i+1} - 1) + \tilde{C}_{UB}(\cos\theta_{i,i+1} - \cos\lambda_{i,i+1}) - \tilde{J}} \tag{18c}$$

$$= \langle -1 | \hat{P}(\boldsymbol{\omega}_i, \boldsymbol{\omega}_{i+1}) | +1 \rangle. \tag{18d}$$

It is written in the canonical base $|U\rangle = |+1\rangle$ and $|B\rangle = |-1\rangle$ of the U and B states. The end vector

$$|V\rangle = \mathrm{e}^{\tilde{\mu}_0/2} |U\rangle + \mathrm{e}^{-\tilde{\mu}_0/2} |B\rangle \tag{19}$$

enters in order to take care of the *free chain* boundary conditions [18] (see also section 3).

The partition function can be rewritten by examining the effective Ising model obtained by integrating over the

[2] Classical calculations lead to smaller values by one order of magnitude [37].

chain conformational degrees of freedom $\boldsymbol{\omega}_i$ in (17). The problem reduces to that of an effective Ising model with an 'effective free energy' $H_{\mathrm{Ising,eff}}$ containing renormalized parameters. This method works because, for the coupled Ising-chain system, the rotational symmetry is not broken. Hence the matrix obtained by integrating the kernel $\hat{P}(\boldsymbol{\omega}_i, \boldsymbol{\omega}_{i+1})$ in (17) is the same for any site $i$.

We thus are able to carry out the angle integrations in sequential fashion by using the triad $\{\hat{\mathbf{e}}_{\mu,i-1}\}$ as the referential for the $i$th Euler angle integration. Since this corresponds for each integration to making a rotational transformation for the variables with the Jacobian equal to 1, the Euler angle integrated transfer matrix is

$$\hat{P}_{\mathrm{I,eff}} = \int \frac{\mathrm{d}^3\boldsymbol{\omega}_i}{8\pi^2}\hat{P}(\boldsymbol{\omega}_i, \boldsymbol{\omega}_{i+1}) = \begin{pmatrix} \mathrm{e}^{-G(\tilde{\kappa}_{\mathrm{U}},\tilde{C}_{\mathrm{U}})+\tilde{J}+\tilde{K}+\tilde{\mu}_0} & \mathrm{e}^{-G(\tilde{\kappa}_{\mathrm{UB}},\tilde{C}_{\mathrm{UB}})-\tilde{J}} \\ \mathrm{e}^{-G(\tilde{\kappa}_{\mathrm{UB}},\tilde{C}_{\mathrm{UB}})-\tilde{J}} & \mathrm{e}^{-G(\tilde{\kappa}_{\mathrm{B}},\tilde{C}_{\mathrm{B}})+\tilde{J}-\tilde{K}-\tilde{\mu}_0} \end{pmatrix} \tag{20}$$

where $G(\tilde{\kappa}, \tilde{C})$ is (in units of $k_{\mathrm{B}}T$) the free energy of a single joint (two-link) subsystem with bending and torsional rigidities $(\kappa, C)$ (either $U$–$U$, $B$–$B$, $U$–$B$):

$$G(\tilde{\kappa}, \tilde{C}) = -\ln\left[\int \frac{\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi\,\mathrm{d}\psi}{8\pi^2}\,\mathrm{e}^{\tilde{\kappa}(\cos\theta-1)+\tilde{C}(\cos\theta-\cos\lambda)}\right] \tag{21a}$$

$$= 2\tilde{\kappa} - \ln\left[\int_0^1 \mathrm{d}x\; I_0(\tilde{C}x)\,\mathrm{e}^{(2\tilde{\kappa}-\tilde{C})x}\right], \tag{21b}$$

where $I_0$ is the modified Bessel function of the first kind[3]. Two interesting cases are:

- $C = 0$ leading to $G(\tilde{\kappa}, 0) = G_0(\tilde{\kappa})$ already defined in (1) which is an increasing function of $\tilde{\kappa}$ (cf figure 1), and (21*b*) is a generalization of the previous result (1) [17, 18];
- $\kappa = 0$, $G(0, \tilde{C}) = \tilde{C} - \ln[I_0(\tilde{C}) + I_1(\tilde{C})]$ which is also an increasing function of $\tilde{C}$ (cf figure 1).

The function $G(\tilde{\kappa}, \tilde{C})$ is plotted in figure 1 which shows that it is a monotonically increasing function. In the spin-wave approximation, the integral (21*a*) is computed using the saddle-point approximation, and the asymptotic behavior of $G$ is

$$G(\tilde{\kappa}, \tilde{C}) \underset{\tilde{\kappa},\tilde{C}\gg 1}{\longrightarrow} \ln(2\tilde{\kappa}) + \frac{1}{2}\ln\left(2\sqrt{\frac{2}{\pi}}\tilde{C}\right). \tag{22}$$

We observe in figure 1 that the asymptotic limit is a very good approximation for $\tilde{\kappa}$ and $\tilde{C}$ larger than 2, and thus for real DNA.

The Hamiltonian of the model (5) then reduces to an effective Hamiltonian which is now of Ising-type

$$\mathcal{H}_{\mathrm{Ising,eff}}[\sigma] = -\mu_0\sum_{i=1}^{N}\sigma_i - \sum_{i=1}^{N-1}\left[J_0\sigma_{i+1}\sigma_i + \frac{K_0}{2}(\sigma_{i+1}+\sigma_i)\right] \tag{23}$$

[3] This result does not change if we include in the model the equilibrium twist $\Phi_0 = 0.19\,\pi$ rad for dsDNA in the Hamiltonian (7) by changing $\cos(\phi+\psi)$ in $\cos(\phi+\psi-\Phi_0)$ for unbroken nearest neighbors.

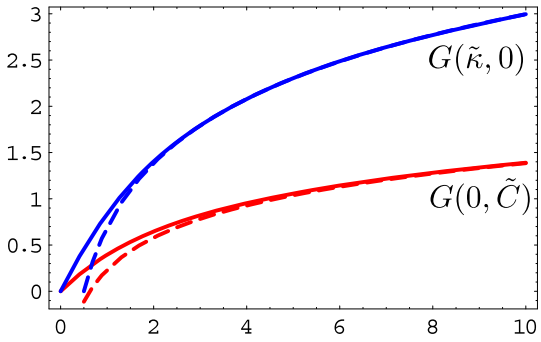


**Figure 1.** Plots of the function $G(\tilde{\kappa}, 0)$ and $G(0, \tilde{C})$ and the asymptotic expressions (22) as broken lines.

where the bare Ising parameters $K$ and $J$ are renormalized according to

$$K_0 = K - \frac{k_{\mathrm{B}}T}{2}[G(\tilde{\kappa}_{\mathrm{U}}, \tilde{C}_{\mathrm{U}}) - G(\tilde{\kappa}_{\mathrm{B}}, \tilde{C}_{\mathrm{B}})] \tag{24a}$$

$$J_0 = J - \frac{k_{\mathrm{B}}T}{4}[G(\tilde{\kappa}_{\mathrm{U}}, \tilde{C}_{\mathrm{U}}) + G(\tilde{\kappa}_{\mathrm{B}}, \tilde{C}_{\mathrm{B}}) - 2G(\tilde{\kappa}_{\mathrm{UB}}, \tilde{C}_{\mathrm{UB}})] \tag{24b}$$

and $\mu_0$ is defined in (15).

Usually, it is admitted that the torsional modulus is proportional to the bending modulus $C \simeq 1.6\,\kappa$ [50]. Taking the same values as in [17, 18] for a polydA–polydT homopolymer, $\tilde{\kappa}_{\mathrm{U}} = \tilde{\kappa}_{\mathrm{UB}} = 147$ and $\tilde{\kappa}_{\mathrm{B}} = 5.54$ at $T = T_{\mathrm{m}} = 326$ K, we get $G(\tilde{\kappa}_{\mathrm{U}}, \tilde{C}_{\mathrm{U}}) = 9.3$ and $G(\tilde{\kappa}_{\mathrm{B}}, \tilde{C}_{\mathrm{B}}) = 4.3$ which leads to a decrease of $K$ and $J$ by about 2–3 $k_{\mathrm{B}}T$ and 1–2 $k_{\mathrm{B}}T$ respectively in the temperature range of interest. We have found in [18] $\mu = 1.78\,k_{\mathrm{B}}T$, $J = 3.64\,k_{\mathrm{B}}T$ and $K$ was set to 0. Hence these entropic contributions are of the same order of magnitude as the bare values and must be taken into account.

Moreover, with these values, the spin-wave approximation applies and we can summarize (15) and (24*a*) as

$$L_0 = \mu_0 + K_0 \approx \mu + K - \frac{k_{\mathrm{B}}T}{2}\ln\left(\frac{a_{\mathrm{B}}^2\epsilon_{\mathrm{U}}\kappa_{\mathrm{U}}\sqrt{C_{\mathrm{U}}}}{a_{\mathrm{U}}^2\epsilon_{\mathrm{B}}\kappa_{\mathrm{B}}\sqrt{C_{\mathrm{B}}}}\right) \tag{25}$$

showing that the renormalization of the Ising parameters comes essentially from entropic effects, namely stretching, bending and torsional entropies.

This twist-induced melting might be important in the context of single molecule torque experiments [28, 35–37, 59] and in the context of superhelical stressed circular dsDNA [60]. For instance, within this model, applying a torque (or a twist) will locally modify the free energy cost $L$ to nucleate a bubble and will, in return, influence the mechanical response of the chain.

In the rest of the paper, we will be interested in expectation values depending only on the spin variables $\sigma_i$. Hence, everything can be computed using directly the effective Ising Hamiltonian (23) with renormalized parameters, $\mu_0$, $L_0$ and $J_0$. In principle, the DHWC model could be completely solved by transfer matrix techniques, thus requiring the diagonalization of the transfer operator $\hat{P}(\boldsymbol{\omega}_i, \boldsymbol{\omega}_{i+1})$ defined in (17). This is out of the scope of the present work.

### *2.4. End-to-end distance*

In this section, we compute the end-to-end distance of a dsDNA using the model presented in [17] where we neglect the torsional term. We show that the difference in monomer sizes in the unbroken and broken states modifies the end-to-end distance and should be taken into account. Therefore, we complete the findings of [18] where the monomer sizes were supposed to be equal.

The end-to-end distance of the chain is defined as $R = \sqrt{\mathbf{R}^2}$, where

$$\mathbf{R}^2 = \sum_{i,j=1}^{N} \langle (a_i \hat{\mathbf{t}}_i) \cdot (a_j \hat{\mathbf{t}}_j) \rangle$$
$$= \sum_{i,j=1}^{N} A^2 \langle \sigma_i \mathbf{t}_i \cdot \mathbf{t}_j \sigma_j \rangle + AB \left( \langle \sigma_i \mathbf{t}_i \cdot \mathbf{t}_j \rangle + \langle \mathbf{t}_i \cdot \mathbf{t}_j \sigma_j \rangle \right) + B^2 \langle \mathbf{t}_i \cdot \mathbf{t}_j \rangle. \quad (26)$$

The monomer size, which depends on the internal variable $\sigma_i$, is written as $a_i = A\sigma_i + B$ with $A = (a_\mathrm{U} - a_\mathrm{B})/2$ and $B = (a_\mathrm{U} + a_\mathrm{B})/2$. In the thermodynamic limit, $N \to \infty$, this expression simplifies to

$$\frac{\mathbf{R}^2}{N} \underset{N\to\infty}{\longrightarrow} \left( A^2 \langle \sigma_i^2 \rangle + 2AB \langle \sigma_i \rangle + B^2 \right) + 2 \sum_{r=1}^{\infty} \left[ A^2 \langle \sigma_i \mathbf{t}_i \cdot \mathbf{t}_{i+r} \sigma_{i+r} \rangle + AB (\langle \sigma_i \mathbf{t}_i \cdot \mathbf{t}_{i+r} \rangle + \langle \mathbf{t}_i \cdot \mathbf{t}_{i+r} \sigma_{i+r} \rangle) + B^2 \langle \mathbf{t}_i \cdot \mathbf{t}_{i+r} \rangle \right] \quad (27)$$

which is independent of $i$. By using the transfer matrix approach and the results already presented in [18], we find after some lengthy calculations (see [18] for detailed definitions)

$$\frac{\mathbf{R}^2}{N} \underset{N\to\infty}{\longrightarrow} A^2 + 2AB \langle \hat{\sigma}_z \rangle + 2B^2 \xi_\mathrm{eff}^p + 2 \sum_{\tau} (A^2 \langle 1, \tau | \hat{\sigma}_z | 0, + \rangle^2 + 2AB \langle 0, + | 1, \tau \rangle \langle 1, \tau | \hat{\sigma}_z | 0, + \rangle) \frac{\mathrm{e}^{-1/\xi_\tau}}{1 - \mathrm{e}^{-1/\xi_\tau}} \quad (28)$$

where the effective persistence length is defined as

$$\xi_\mathrm{eff}^p \equiv \frac{1}{2} \sum_{\tau} \langle 1, \tau | 0, + \rangle^2 \frac{1 + \mathrm{e}^{-1/\xi_\tau}}{1 - \mathrm{e}^{-1/\xi_\tau}}. \quad (29)$$

The Pauli matrix $\hat{\sigma}_z$ acts only on the second part of the basis that diagonalizes the transfer matrix operator $\hat{P}$: $|\Psi_{l,m,\tau}\rangle = |l, m\rangle \otimes |l, \tau\rangle$ (where $(l, m)$ are the quantum numbers associated to the spherical harmonics and $\tau = \pm$ labels the eigenstates of the Ising model). In the basis $|0, \pm\rangle$ we have

$$\hat{\sigma}_z = \begin{pmatrix} \langle c \rangle_\infty & \sqrt{1 - \langle c \rangle_\infty} \\ \sqrt{1 - \langle c \rangle_\infty} & -\langle c \rangle_\infty \end{pmatrix} \quad (30)$$

where $\langle c \rangle_\infty$ is the expectation value of the average spin variable (or 'magnetization') in the thermodynamic limit

$$\frac{1}{N} \sum_{1}^{N} \langle \sigma_i \rangle \underset{N\to\infty}{\longrightarrow} \langle c \rangle_\infty = \frac{\sinh(L_0)}{[\sinh^2(L_0) + \mathrm{e}^{-4J_0}]^{1/2}}. \quad (31)$$

The parameter $L_0$ is defined in (1) and $J_0$ in (24*b*), setting $\tilde{C} = 0$ for the three cases. The two orthonormal eigenvectors for a fixed $l$ are defined in [17, 18].

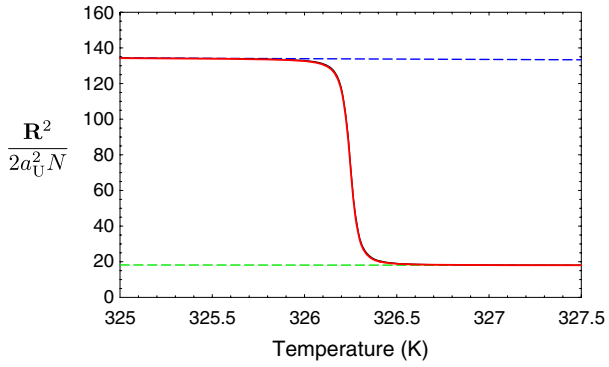


**Figure 2.** Normalized mean square end-to-end distance $R^2/(2a_\mathrm{U}^2 N)$ (in units of base-pair size) as a function of the temperature $T$ for the parameter values $\mu = 4.46$ kJ mol$^{-1}$, $J = 9.13$ kJ mol$^{-1}$, $K = 0$, corresponding to $T_\mathrm{m} = 326.4$ K and $\tilde{\kappa}_\mathrm{U} = \tilde{\kappa}_\mathrm{UB} = 147$, $\tilde{\kappa}_\mathrm{B} = 5.5$, $a_\mathrm{B} = 2a_\mathrm{U}$. The full calculation (28) (in red) and the interpolation formula (32) (in black) coincide. The upper (blue) and lower (green) broken lines correspond to the bare dsDNA and ssDNA values, respectively.

The result (28) is shown in figure 2 for $a_\mathrm{B} = 2a_\mathrm{U}$. An accurate interpolating formula is given by

$$\mathbf{R}^2_\mathrm{interpol} = 2N(\varphi_\mathrm{U} a_\mathrm{U}^2 \xi_\mathrm{U}^p + \varphi_\mathrm{B} a_\mathrm{B}^2 \xi_\mathrm{B}^p) = (1 - \varphi_\mathrm{B}) \mathbf{R}^2_\mathrm{ds} + \varphi_\mathrm{B} \mathbf{R}^2_\mathrm{ss} \quad (32)$$

thus generalizing a similar result given in [18] for the case $a_\mathrm{U} = a_\mathrm{B}$ ($\phi_\mathrm{B} = 1 - \phi_\mathrm{U}$ defined in (35)).

## 3. Finite size effects within the DHWC model

In this section, we study the behavior of the fraction of open base-pairs, $\varphi_\mathrm{B}(N, T)$, as a function of both temperature and chain length for homogeneous DNA with free and modified boundary conditions (necessary for DNA inserts). Despite early recognition [61] that a careful experimental study of such homogeneous DNA polymers of varying length would be of great help in advancing our theoretical understanding of DNA denaturation, unfortunately such a study has not yet been carried out. As a consequence, important questions concerning the competition between end unwinding and internal bubble formation for finite chains, as well as the correct form of the loop entropy factor (including the effect of chain rigidity) and the role of chain dissociation, remain open. Our goal here is to shed further light on the role of polymer length in the thermal denaturation homogeneous DNA (see [62] for a recent study of finite size effects within the framework of generalizations to the Peyrard–Bishop model).

The model we use here is a generalization of the one presented in [17, 18] and has been defined in section 2. The renormalized chemical potential is given by (15) and for purposes of illustration we use the simpler, but accurate, spin-wave approximations for the two other renormalized parameters, summarized here:

$$L_0(T) \approx L - \frac{k_\mathrm{B} T}{2} \ln \left( \frac{a_\mathrm{B}^2 \epsilon_\mathrm{U} \kappa_\mathrm{U} \sqrt{C_\mathrm{U}}}{a_\mathrm{U}^2 \epsilon_\mathrm{B} \kappa_\mathrm{B} \sqrt{C_\mathrm{B}}} \right) \quad (33)$$

$$J_0(T) \approx J - \frac{k_B T}{4} \ln\left(\frac{\kappa_U \kappa_B}{\kappa_{UB}^2} \frac{\sqrt{C_U C_B}}{C_{UB}}\right), \qquad (34)$$

where $L = \mu + K$. We also adopt the following physically reasonable set of model parameters: $\kappa_U/\kappa_B = 147/5.5 = 26.7$, $\kappa_{UB} = \kappa_U$, $a_B/a_U = \epsilon_U/\epsilon_B = 2$, and $C_B/\kappa_B = C_U/\kappa_U = C_{UB}/\kappa_{UB} = 1.6$. When loop entropy is not included in the model, we use a value for $J$ obtained previously by fitting experimental melting data for a homopolynucleotide polydA–polydT: $J = 9.13$ kJ mol$^{-1}$ [17, 18] (we recall that the renormalized value $J_0$ is a key parameter in determining the transition width). When the effect of loop entropy on the thermal denaturation of free chains is studied, we will use a smaller value of $J$, half of the larger one, as it is well known that loop entropy tends to sharpen the transition [42, 52, 61]. If the model prediction (without loop entropy) for the melting temperature for polydA–polydT of length $N = 30\,000$ base-pairs is chosen to agree with the experimental results in [1] ($T_m^{expt} = 338.70$ K), then we obtain $L = \mu + K = 9.87$ kJ mol$^{-1}$, close to the value obtained by setting $L_0(T_m^{expt}) = 0$ (which gives the model result without loop entropy for the infinite chain melting temperature, see equation (33)). Using $J = 9.13$ kJ mol$^{-1}$, we find that $J_0 \simeq 12.3$ kJ mol$^{-1}$ at $T = 339$ K, which implies that the entropic contribution is greater than 25% near the melting temperature ($\tilde{J}_0 = \beta J = 3.23$ for $\beta = 1/(k_B T_m^{expt})$). Using $J = 4.57$ kJ mol$^{-1}$, we find that $J_0 \simeq 7.70$ kJ mol$^{-1}$ at $T = 339$ K, which gives an entropic contribution of 41%.

In our previous work [17, 18] we assumed that the difference in bare stacking energy, $K$, between the U and B states was zero. This choice was based on evidence that near room temperature single-stranded polyrA remains stacked [63]. It seems, however, that near the dsDNA melting temperature dT single strands are probably completely and dA ones partially unstacked [1], with an unstacking fraction close to 75% near $T_m$ [1]. We can conclude that the single dT and dA strands in polydA–polydT bubbles may have much less stacking energy than the helical segments and incorporate this effect into the model by introducing a weighting parameter, $f$, that measures the contribution of $K$ to $L$ at fixed $L$: $K = fL$ and $\mu = (1 - f)L$. Although the two unbound single dT and dA strands in a polydA–polydT bubble may not behave exactly like two free single dT and dA strands, the above discussion does suggest that $f$ may be large near the melting temperature. Indeed, if we accept the putative experimental value for the bare enthalpy needed to open one A–T base-pair as a measure of $\mu$, then we find $\mu \simeq 5.25$ kJ mol$^{-1}$ [17, 18, 64]. Using this result and the above value for $L$ then yields $f \simeq 0.5$. When $f$ is taken to be zero there is no loss in stacking energy when a bubble opens and we recover the case previously studied in [17, 18].

An important question is how to incorporate bubble loop entropy into statistical models of fluctuating DNA. This loop contribution arises from the extra cost in free energy (with respect to two single unbound end chains) needed to form a closed loop of bases making up a bubble [42, 61, 65, 66]. When loop entropy is neglected Poland–Scheraga (PS) type models reduce to effective Ising ones, albeit without the end-interior asymmetry that naturally arises within our approach from the difference between $L_0$ and $\mu_0$ (see equation (20) of [18]). This can arise both from a dissimilarity between $\mu$ and $K$ and from the renormalizations coming from integrating out the conformational degrees of freedom. If, without justification, we formally set $\mu_0$ equal to $L_0$, we recover previous Ising/PS type models without loop entropy.

For finite DNA polymers, end effects may have a strong influence on both the thermal denaturation transition and chain conformational properties. As already discussed in [18] the coupled DNA model that we have developed is extremely useful for investigating the dependence of various system properties on chain length, $N$. For DNA homopolymers two types of situations can be envisaged: (i) finite homopolymers with free end boundary conditions, and (ii) finite polydA–polydT inserts between more stable G–C rich domains with much higher melting temperatures.

Case (i) has already been extensively studied theoretically in [18] when the loop entropy associated with bubbles is neglected. Although for very long chains end effects are unimportant and $f$ plays no role (only the value of $L$ is important), for not too long finite chains $f$ has a strong influence on the melting curves. Within the scope of our model with $f = 0$ it was found previously that for finite DNA chains thermal denaturation takes place in an inhomogeneous fashion with the probability of base-pair opening being higher at chain ends for temperatures $T < T^*$. At the temperature $T^*$ the fraction of broken base-pairs becomes independent of chain length and the probability of base-pair opening becomes independent of position on the chain (see figures 6 and 7 of [18]). For $f = 0$ it was also found that the melting temperature obeys $T^* < T_m^\infty < T_m(N)$ (where $T_m^\infty = T_m(N \to \infty)$) and, along with the transition width, decreases with increasing $N$. For $T < T^*$ the fraction of open base-pairs, $\varphi_B(N)$, decreases with increasing $N$, whereas for $T > T^*$, it increases with increasing $N$. We further this previous theoretical study here by investigating the influence of the weighting factor $f$.

Without loop entropy, previous Ising/PS type models predict $T_m$ independent of $N$ and therefore $T_m = T^*$. When loop entropy is added to these models $T_m(N)$ becomes an increasing function of $N$, which appears to agree with experiment (in [1] it was found that the melting transition for a free homopolymer of length $N = 30\,000$ takes place at a temperature 1 K higher than that of chains of length $N \approx 500$ and is much sharper). We also examine in detail the validity of the one-sequence approximation for free boundary conditions and investigate the influence of loop entropy in situations where the accuracy of this approximation can be gauged [42, 61]. For free boundary conditions this approximation involves keeping only the base-pair states forming one interior bubble or one helix section of variable length $0 \leqslant n \leqslant N$. Unfortunately, there do not appear to be any detailed experimental studies of the thermal denaturation of DNA homopolymers with free ends as a function of chain length (see, however [52, 67]) that can be used to test the model predictions and clarify the role and importance of both end effects and bubble loop entropy.

For the case (ii) of an A–T insert of length $N$ in larger more stable DNA polymers, detailed experiments [1] have

already been carried out for $60 < N < 140$ and also interpreted using both a simple two-state approximation for the A–T insert and the Poland–Scheraga model [42] (including loop entropy) for the entire polymer [1]. For inserts, the boundary conditions are fixed mainly by the exterior G–C rich domains and only $L$ enters (and not $f$, i.e. the individual values of $\mu$ and $K$). For inserts the one-sequence approximation involves keeping only the base-pair states forming one bubble of variable length $0 \leqslant n \leqslant N$. The two-state approximation accounts only for the completely closed and the completely open chain states in the partition function [42] and is a special case of the more general one-sequence approximation. The validity of these types of approximations relies intimately on the relatively large cost in free energy for creating a bubble (or base-pair domain walls) compared with the cost of changing the length of an already existing bubble (i.e. $|L_0| \ll J_0$). The upshot is that a one-bubble state can have a variable length (and in dynamics undergoes breathing) and such states should dominate the free energy for not too long chains (and for longer chains, temperatures not too close to the melting one).

We reexamine this problem by analyzing the same experimental results [1] using our coupled model for a finite chain with modified boundary conditions, because in such situations the nature of end monomers becomes extremely important. In doing so, we study the validity of both the two-state and one-sequence approximations without loop entropy by comparing the predictions of these simplified approaches to those obtained from the exact solution to our model. By incorporating the loop entropy into the one-sequence approximation, we also examine the role and importance of this effect for homopolymer inserts. In order to compare the predictions of the model with experiments on A–T inserts we have fitted the DNA melting data presented in figures 6 and 7 of [1] using simple fitting functions, the goal being to get a smooth approximation to the data (see the appendix) that will be useful in this section.

### *3.1. Exact results for general chain boundary conditions (without loop entropy)*

Using transfer matrix techniques we have shown that it is possible to obtain a compact expression for the average fraction of open base-pairs in a finite chain of length $N$ for arbitrary boundary conditions [18] (with neither loop entropy, nor chain sliding):

$$\varphi_{\mathrm{B}}(N, T; \tilde{\mu}') = \tfrac{1}{2}\left[1 - \langle c\rangle(N, T; \tilde{\mu}')\right] \tag{35}$$

where $\langle c\rangle(N, T; \tilde{\mu}') \equiv 1/N \sum_{i=1}^{N} \langle \sigma_i \rangle$ is given by

$$\langle c\rangle(N, T; \tilde{\mu}') = \langle c\rangle_\infty \left[1 - \frac{2R_V^2}{R_V^2 + \mathrm{e}^{(N-1)/\xi_I}}\right] + \frac{2R_V\sqrt{1-\langle c\rangle_\infty^2}\left(1 - \mathrm{e}^{-N/\xi_I}\right)}{N\left[1 + R_V^2\,\mathrm{e}^{-(N-1)/\xi_I}\right]\left(1 - \mathrm{e}^{-1/\xi_I}\right)} \tag{36}$$

$\xi_{\mathrm{I}}$ is the Ising correlation length, and

$$R_V(\tilde{\mu}') \equiv \frac{\langle V'|0,-\rangle}{\langle V'|0,+\rangle} \tag{37}$$

with the normalized end vector

$$|V'(\tilde{\mu}')\rangle = \left[2\cosh(\tilde{\mu}')\right]^{-1/2}\left(\mathrm{e}^{\tilde{\mu}'/2}|U\rangle + \mathrm{e}^{-\tilde{\mu}'/2}|B\rangle\right) \tag{38}$$

enforcing the chain boundary conditions. The quantities $\langle c\rangle(N, T; \tilde{\mu}')$, $\langle c\rangle_\infty$ (as given in (31)), $R_V(\tilde{\mu}')$, and $\xi_{\mathrm{I}}$ are all functions of $L_0$ and $J_0$ [18]. For free ends $\tilde{\mu}' = \tilde{\mu}_0$, whereas for closed (open) ends, $|V'\rangle = |U\rangle$ ($|B\rangle$), which can be seen by taking the $\tilde{\mu}' \to \pm\infty$ limits of (38). When $\tilde{\mu}'$ is formally set equal to $\tilde{L}_0$ there is no longer any end-interior asymmetry and the model reduces to older Ising/PS type [52] models without loop entropy.

A simple expression can be obtained for $R_V$ by setting $N = 1$ in (36) and solving for $R_V$:

$$R_V(\tilde{\mu}') = \frac{\langle c\rangle_1 - \langle c\rangle_\infty}{\sqrt{1-\langle c\rangle_\infty^2} + \sqrt{1-\langle c\rangle_1^2}} \tag{39}$$

where $\langle c\rangle_1 = \tanh(\tilde{\mu}')$ is a function of $\tilde{\mu}'$ and therefore reflects the boundary conditions.

*DNA chains with free boundary conditions.* When $\tilde{\mu}' = \tilde{\mu}_0$ (free boundary conditions), $R_{V,\mathrm{free}} = R_V(\tilde{\mu}_0)$ gets simplified in the following way for special values of $T$ [18]:

$$R_{V,\mathrm{free}} = \begin{cases} -\mathrm{e}^{-\tilde{\mu}_0}, & T < T^* \\ 0, & T = T^* \\ \tanh(\tilde{\mu}_0/2), & T = T_{\mathrm{m}}^{\infty} \\ \mathrm{e}^{\tilde{\mu}_0}, & T > T_{\mathrm{m}}^{\infty} \end{cases} \tag{40}$$

which shows that $R_{V,\mathrm{free}}$ is a monotonically increasing function of $T$ and vanishes at $T = T^*$.

In figure 3 we present model results (with neither loop entropy, nor chain sliding) based on equation (36) for free chains of different lengths and different values of $f$. We observe that $T^*$ increases with increasing $f$; for $f < 0.7$, $T_{\mathrm{m}}(N)$ decreases with increasing $N$, whereas for $f > 0.7$, $T_{\mathrm{m}}(N)$ increases with increasing $N$. When $f \approx 0.7$, the melting curves are nearly identical with the results obtained from older Ising/PS type models ($\mu' = L_0$) without loop entropy. When loop entropy is added to the model, the melting temperatures for the longer chains will be shifted to the right, amplifying the effect of finite $f$ (see below).

*DNA inserts with closed boundary conditions.* For an A–T insert of length $N$ in more stable G–C domains a simple starting approximation is to apply closed boundary conditions (i.e. base-pairs $i = 1$ and $N$ are considered to be held closed due to their coupling to the adjacent G–C domains). For closed boundary conditions, $\tilde{\mu}' \to \infty$, leading to

$$R_{V,\mathrm{cl}} = \frac{\sqrt{1-\langle c\rangle_\infty}}{\sqrt{1+\langle c\rangle_\infty}} \tag{41}$$

which is non-zero for all $T > 0$, implying that in this case $T^* = 0$.

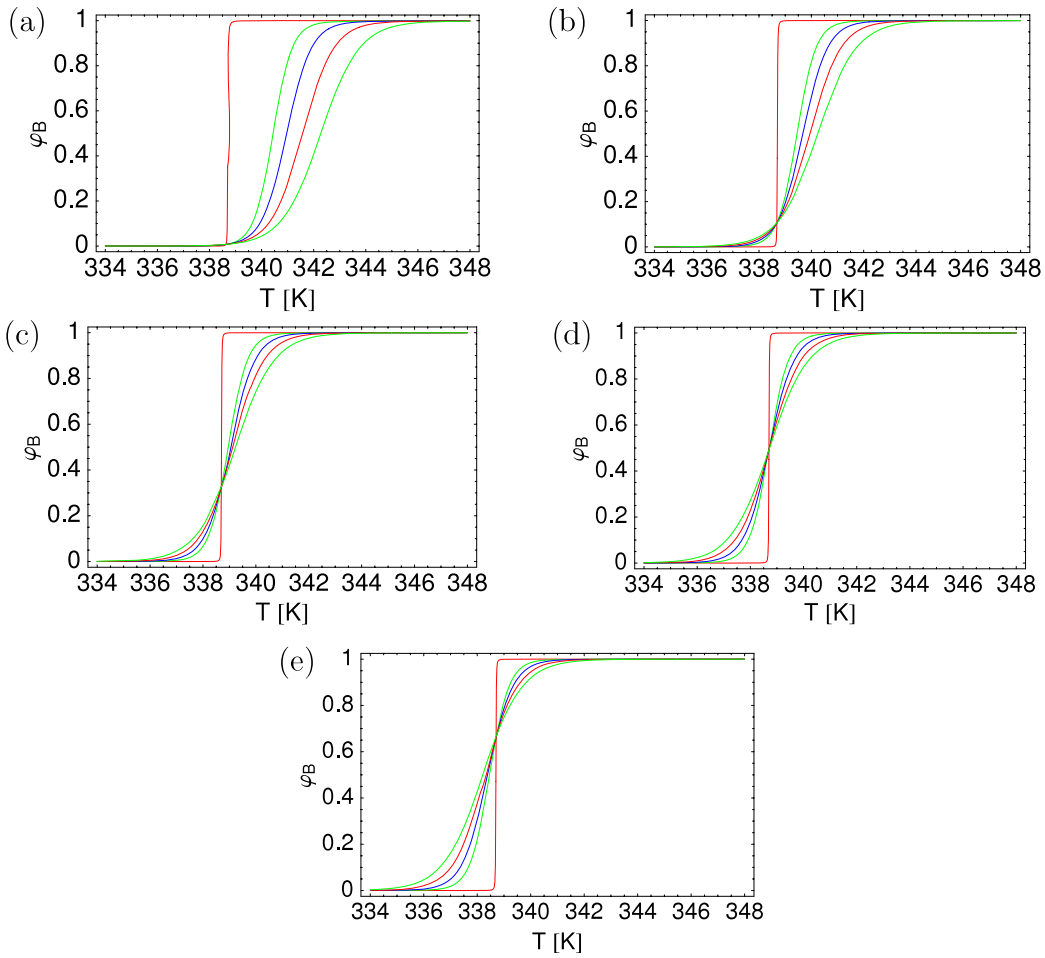


**Figure 3.** Fraction of broken base-pairs (35) versus temperature for free boundary conditions (without loop entropy) and chain lengths of $N = 30\,000$, 136, 105, 83 and 67 (from left to right, above the temperature of intersection, $T^*$) (a) $f = 0$, (b) 0.4, (c) 0.6, (d) 0.7, (e) 0.8 (other model parameters used are listed at the beginning of section 3).

Unfortunately in this case only $N - 2$ base-pairs can open. A better approach involves artificially extending the insert length from $N$ to $N + 2$ and using closed boundary conditions on the extended chain. In this case the 'fictitious' ($i = 1$ and $N + 2$) base-pairs are held closed by the boundary conditions in order to simulate the influence of the adjacent more stable G–C rich domains and the remaining $N$ base-pairs can fluctuate. Since the $i = 2$ and $N + 1$ base-pairs are adjacent to closed base-pairs, their probability of opening will be lower than that of interior ones. It is clear that in this case melting will begin near the center of the insert. If $\varphi_B^{cl}(N, T)$ is the fraction of open base-pairs for a chain of length $N$ with closed boundary conditions, then simple counting shows that the average fraction of open base-pairs in the extended model is given by

$$\varphi_B^{ext}(N, T) = \frac{N+2}{N}\varphi_B^{cl}(N+2, T). \tag{42}$$

A more sophisticated approach is to keep the physical insert length of $N$ and account for the coupling to the more stable G–C rich domains via a mean-field-type approximation by taking $\mu_0 < \mu' < \infty$. The approaches presented above are obviously valid only when the temperature is sufficiently far below the melting temperature of the G–C rich domains so that the experimental ultraviolet (UV) absorbance used to measure $\varphi_B(N, T)$ comes primarily from the A–T inserts in the temperature range of interest.

In figure 4 we show how $\varphi_B(N, T, \tilde{\mu}')$ varies as a function of $\tilde{\mu}'$ for $N = 136$. The melting temperature as a function of $\tilde{\mu}'$ interpolates smoothly between the results for free ($\tilde{\mu}' = \tilde{\mu}_0$) and closed boundary conditions over a temperature range of ~5 K and the width of the transition increases slightly with increasing $\tilde{\mu}'$.

In figures 5 we compare the experimental results for A–T inserts (figure A.2(a)) with the model predictions for $\varphi_B(T, N)$ for $f = 0$ and three different model boundary conditions: (i) free boundary conditions, (ii) optimized $\mu'$, (iii) extended model, closed boundary conditions. The value of $L = 9.87$ kJ mol$^{-1}$ is held fixed to reproduce the experimental melting temperature for $N = 30\,000$ and the model predictions for the optimized (for $T_m(N)$) $\mu'$ case are practically insensitive to changes in $f$ and $J$. For closed boundary conditions $\varphi_B(N)$ increases with increasing $N$ at fixed $T$ simply because the end effects become attenuated for long chains, as illustrated in figure 5. We conclude that the model in its present form can reproduce the qualitative

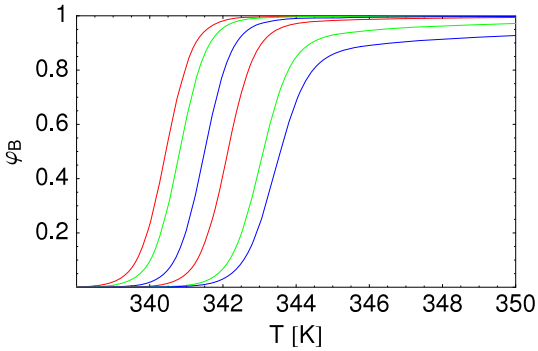


**Figure 4.** Fraction of broken base-pairs (35) versus temperature for $N = 136$ and $f = 0$ ($\mu = L = 9.87$ kJ mol$^{-1}$) as a function of $\mu'$: from left to right, free boundary conditions ($\mu' = \mu_0$); $\mu'/L = 0.86$; 1.14, 1.43; 2.00; 8.56; for closed boundary conditions the result is superimposed on the right-hand curve ($\mu'/L = +\infty$).

tendencies, but not the quantitative details, of experiments on short A–T inserts (for such short chains including loop entropy into the model will not lead to better fits, see below). The results presented here do allow us, however, to gauge the importance of chain boundary conditions on the melting curves. One difficulty in applying the present model arises because the simplified approach presented here does not account for the increased probability of opening for G–C base-pairs adjacent to the A–T inserts. The complete solution of our model for the full heterogeneous chain is in principal possible using known numerical methods, as is discussed in the conclusion.

The exact result for $\varphi_B(N, T; \mu')$ (35) does not reveal in a physically transparent way which states contribute the most for a given chain length $N$ and temperature $T$ and, as already mentioned, includes neither the effects of loop entropy, nor of chain sliding. In order to include such effects in a straightforward way we now study the one-sequence approximation to the exact partition function for our model, an approximation that should be valid for sufficiently short chains.

### 3.2. One-sequence approximation

*One-sequence approximation for closed boundary conditions: DNA inserts.* We start by examining the one-sequence approximation for homopolymer inserts of length $N$ for closed boundary conditions without loop entropy. The effective free energy of creating an interior $n$-bubble with two base-pair domain walls is [18],

$$\beta \Delta G_{\rm int}^{(n)} = 4\tilde{J}_0 + 2n\tilde{L}_0 \tag{43}$$

and therefore the restricted partition function, $Z_{\rm 1seq}^{\rm cl}$, including only $n$-bubbles varying in size between $n = 0$ (helical insert) and $N$ (bubble insert) is given by:

$$Z_{\rm 1seq}^{\rm cl} = 1 + \sum_{m=0}^{N-1} (m+1) \exp\left[-\beta \Delta G_{\rm int}^{(N-m)}\right] \tag{44}$$

where the first term equal to one comes from a completely closed chain state and for an $n$-bubble $m = N - n$ is the number of remaining intact base-pairs in the insert. The factor of $(m + 1) = N - n + 1$ in the sum is entropic in nature and equal to the number of ways of placing an $n$-bubble inside an insert of length $N$. We recall that $L_0$ becomes negative for $T > T_{\rm m}^{\infty}$ and therefore in the high temperature range of interest for inserts the term depending on $\Delta G_{\rm int}^{(n)}$ in (44) favors large bubbles. The entropic factor, on the other hand, favors

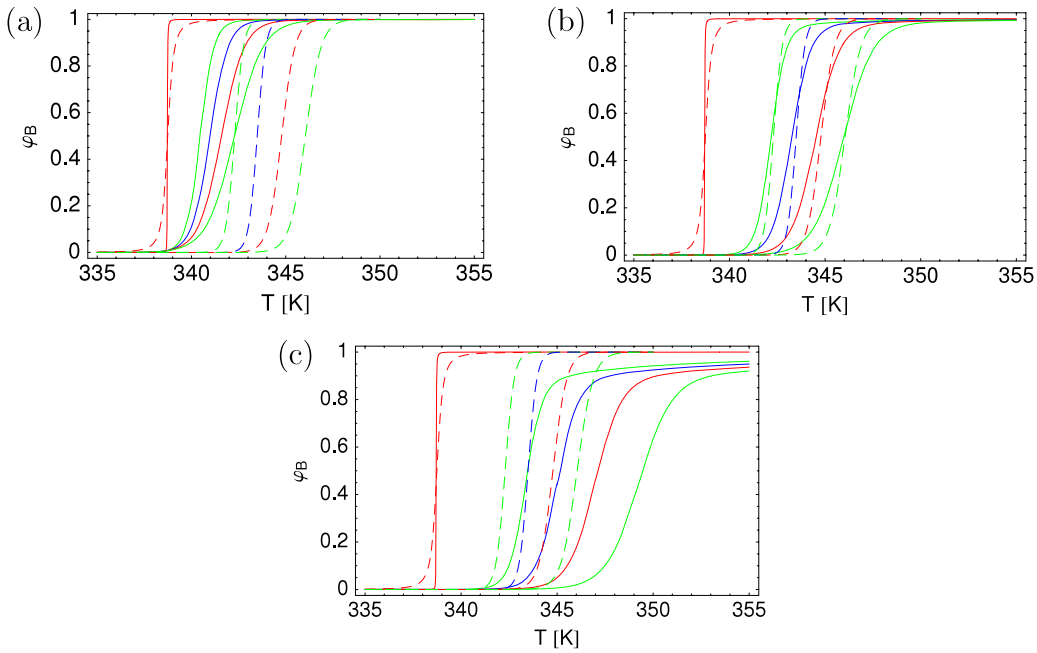


**Figure 5.** Fraction of broken base-pairs (35) versus temperature for (from left to right) $N = 30\,000$, 136, 105, 83 and 67: fitted experimental results from figure A.2(a) (dashed curves) and (a) model predictions using free boundary conditions, (b) $\mu'$ optimized to fit $T_{\rm m}(N)$ and (c) closed boundary conditions.

small bubbles. The one-sequence approximation incorporates the first two terms (of order 0 and 1) in an expansion in powers of the loop initiation factor,

$$\sigma_{\rm LI} = {\rm e}^{-4\tilde{J}_0}, \tag{45}$$

which counts the number of bubbles [42, 61].

Within the one-sequence approximation, the average fraction of broken base-pairs can be obtained from $Z^{\rm cl}_{\rm 1seq}$:

$$\varphi^{\rm cl}_{\rm B,1seq}(N) = -\frac{1}{2N}\frac{\partial\left(\ln Z^{\rm cl}_{\rm 1seq}\right)}{\partial L_0}. \tag{46}$$

The sums in (44) can be carried out to find the following compact expression:

$$Z^{\rm cl}_{\rm 1seq} = 1 + {\rm e}^{-4\tilde{J}_0}\mathcal{C}({\rm e}^{2\tilde{L}_0}) \tag{47}$$

where

$$\mathcal{C}(x) \equiv x^{-N}\left(xp'(x) + p(x)\right) \tag{48}$$

with

$$p(x) \equiv \frac{x^N - 1}{x - 1}. \tag{49}$$

By using (47) the following expression can be obtained for $\varphi^{\rm cl}_{\rm B,1seq}(N)$:

$$\varphi^{\rm cl}_{\rm B,1seq}(N) = -\frac{{\rm e}^{-4\tilde{J}_0}}{N}\left[\left(\frac{\partial\mathcal{C}}{\partial x}\right)x\right]_{x={\rm e}^{2\tilde{L}_0}}. \tag{50}$$

For sufficiently short chains the one-sequence approach without loop entropy defined above will be an accurate approximation to the exact result for the extended model ($N+2$ base-pairs with closed boundary conditions, $\varphi^{\rm ext}_{\rm B}$ (42)). When this approximation is valid, multi-bubble states are extremely rare (the range of validity in $N$ of the one-sequence approximation depends on the value of $J_0$ via $\sigma_{\rm LI}$ (45)).

Although it is difficult to incorporate bubble loop entropy into our model in a general way because of mathematical complications arising from the 'long-range' nature of the loop entropy factor, it is easy to do so within the one-sequence approximation. Including the loop entropy lowers the probability of $n$-bubble opening. We adopt a common simplified form for the loop entropy factor associated with $n$ broken base-pairs [65, 43, 66],

$$g_{\rm LE}(n) = (n_0 + 2 + 2n)^{-k} \tag{51}$$

that depends on the bubble loop length, $\ell_{\rm B} = 2 + 2n$, and is parametrized by a constant $n_0$ and an exponent $k$. The loop entropy exponent $k$ is thought to be in the range $3/2 \leqslant k \leqslant 2.1$, depending on the extent to which chain self-avoidance is taken into account [7]. The term $n_0$ accounts for the enhanced difficulty of forming small closed bubbles arising from DNA chain stiffness. Including the loop entropy leads to a modified one-sequence partition function, given by

$$Z^{\rm cl,LE}_{\rm 1seq} = 1 + \sum_{m=0}^{N-1}(m+1)g_{\rm LE}(N-m)\exp\left[-\beta\Delta G^{(N-m)}_{\rm int}\right]. \tag{52}$$

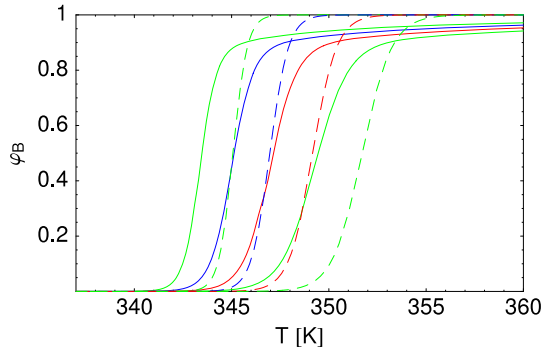


**Figure 6.** Comparison of the two-state approximation (dashed curves) without loop entropy (55) with the full result (42) (solid curves) for closed boundary conditions; from left to right $N = 136$, 105, 83 and 67 (same parameters and colors as figure 3).

The introduction of loop entropy ($k > 0$) in $Z^{\rm cl,LE}_{\rm 1seq}$ can have an exaggerated effect on the calculated melting curves if the loop initiation factor, $\sigma_{\rm LI}$ (45), is not readjusted at the same time. If we define $D = (n_0 + 2)/2$ and use $\tilde{J}_0 \to \tilde{J}_0 + (k/4)\ln(2D)$ in (52) then $Z^{\rm cl,LE}_{\rm 1seq}$ can be rewritten as

$$Z^{\rm cl,LE}_{\rm 1seq} = 1 + \sum_{m=0}^{N-1}(m+1)[1 + (N-m)/D]^{-k} \times \exp\left[-\beta\Delta G^{(N-m)}_{\rm int}\right], \tag{53}$$

with $G_{\rm int}$ still given by (43) (in the fitting of experimental data, the value of $D$ has been taken to be as large as 96 [1] and even 450 [65]). The above readjustment of $J_0$ means that only long $n$-bubbles ($n = N - m > D$) 'feel' the effect of loop entropy (the suppression of short bubble formation due to increased chain stiffness being incorporated directly into the readjusted $J_0$). We will compare the predictions of the one-sequence approximation with ($k$, $D > 0$) and without ($k = 0$) loop entropy using (53). Although the sums in (53) apparently cannot be carried out analytically, once they are performed numerically, the analog of (46) can be used to obtain $\varphi^{\rm cl,LE}_{\rm B,1seq}$.

If in evaluating the one-sequence partition function, $Z^{\rm cl,LE}_{\rm 1seq}$, we retain only the completely closed ($m = N$) and completely open ($m = 0$) states, we obtain the *two-state approximation*:

$$\varphi^{\rm cl,LE}_{\rm B,2st} = \frac{1}{1 + \left\{(1 + N/D)^{-k}\exp\left[-\beta\Delta G^{(N)}_{\rm int}\right]\right\}^{-1}}. \tag{54}$$

A more general $s$-state approximation can be defined by including the $m = 0, \ldots, s-2$ terms in the sum (52). Without loop entropy ($k = 0$), (54) simplifies to

$$\varphi^{\rm cl}_{\rm B,2st} = \tfrac{1}{2}\left\{1 - \tanh\left[-\beta\Delta G^{(N)}_{\rm int}/2\right]\right\}. \tag{55}$$

In figure 6, the two-state approximation without loop entropy is compared to the exact result for the extended case (42). We observe a cross-over temperature (at which the two-state approximation begins to overestimate $\varphi^{\rm cl}_{\rm B}$) roughly given by the temperature at which $\Delta G^{(N)}_{\rm int}$ goes from positive

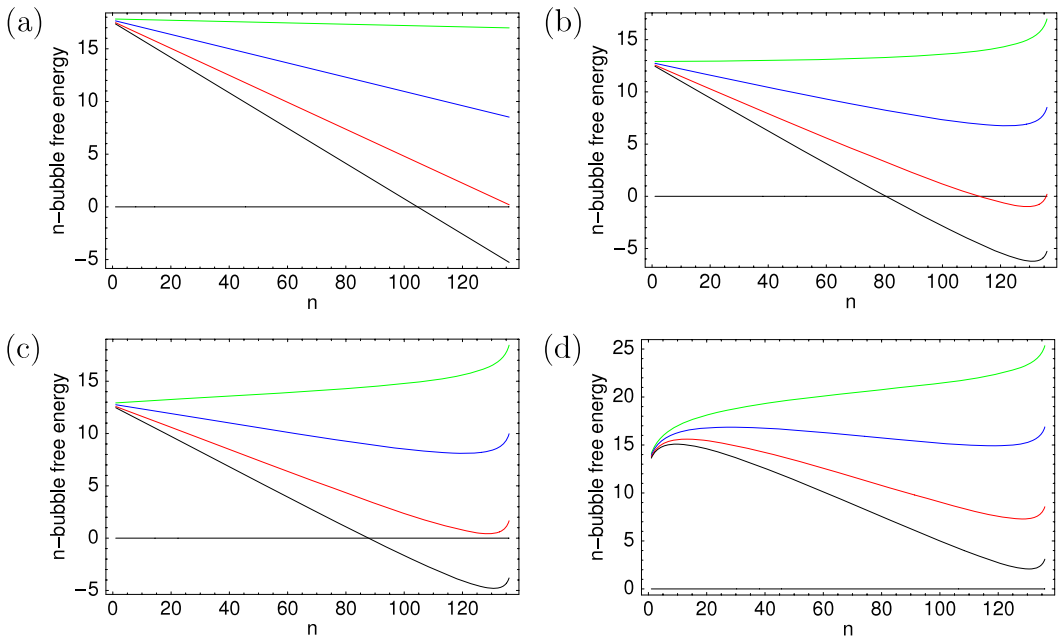


**Figure 7.** Free energy of bubble formation for $N = 136$ and $T = 339, 342, 345, 347$ K (from top to bottom): (a) intrinsic free energy, $\beta\Delta G_{\rm int}^{(n)}$; total free energy, $\beta\Delta F_{\rm int}^{(n)}$: (b) $k = 0$ (without loop entropy); (c) $k = 1.7$, $D = 100$; (d) $k = 1.7$, $D = 1$ ($f = 0$, other model parameters as in figure 3).

to negative (signaling a vanishing 'nucleation barrier' for the completely open insert). Contrary to previous claims [1], in the present case the two-state approximation overestimates $T_{\rm m}(N)$ by more than 2 K and underestimates the transition width.

The form (53) suggests defining an effective total $n$-bubble free energy

$$\beta\Delta F_{\rm int}^{(n)} = \beta\Delta G_{\rm int}^{(n)} - \ln(N - n + 1) + k\ln(1 + n/D) \quad (56)$$

that accounts for the intrinsic free energy of bubble formation (first term), as well as positional (second term) and loop entropy (third term). Following (43), $\beta\Delta G_{\rm int}^{(n)}$ decreases with increasing $n$ for $T > T_{\rm m}^{\infty}$ ($L_0 < 0$) and increases for $T < T_{\rm m}^{\infty}$ ($L_0 > 0$). The positional and loop entropy contributions increase the effective free energy cost of bubble creation as the bubble size $n$ increases.

In figure 7 we plot bubble free energies for $N = 136$ and increasingly important loop entropy effects. The intrinsic part, $\Delta G_{\rm int}^{(n)}$ is a linearly decreasing function of $n$ ($T_{\rm m}^{\infty} \approx 338$ K see figure A.2) and vanishes at $n = N$ for $T = 345$ K, close to the temperature at which the two-state approximation becomes an overestimation (see figure 6). We observe that (i) inclusion of the positional entropy alone (figure 7(b)) leads to a minimum in (56) near $n = N$ for sufficiently high temperatures and (ii) the loop entropy rigidity parameter $D$ plays a minor role when it is close to 100 (figure 7(c)) and an important one when it is close to 1 (figure 7(d)) (the value commonly used in the modeling of infinite chains). In the latter case (56) remains positive over the whole temperature range studied and has a maximum for small $n$ and a minimum near $n = N$ for sufficiently high temperatures (figure 7(d)).

In figure 8, we compare the one-sequence approximations with and without loop entropy (55) for short inserts obeying closed boundary conditions. For $J = 9.13$ kJ mol$^{-1}$ we find that for inserts without loop entropy the one-sequence approximation is practically indistinguishable from the exact result (42) for $N < 10\,000$. Because loop entropy further reduces the probability of bubbles, we therefore believe that the one-sequence approximation with loop entropy should be an excellent approximation in most cases of practical interest (i.e. inserts with lengths less than a few thousand base-pairs). We observe in figure 8 that for such inserts and fixed $L$ the net result of including the loop entropy is to shift the melting curves to the right by about 5 K for $D = 1$ and about 1 K for $D = 100$ without much change in the transition width. It therefore seems as if the addition of loop entropy will not enable us to improve the fits to experiment shown in figure 5(b).

Although it is possible to work out the details of the one-sequence approximation when the end base-pairs in an insert of length $N$ experience a chemical potential $\mu' < \infty$, we will not present these results here.

*One-sequence approximation for free boundary conditions.* We now examine the one-sequence approximation with and without loop entropy for DNA homopolymers of length $N$ with free boundary conditions. Because most synthetic DNA homopolymers are less than a few thousand base-pairs long [42, 61, 1, 52], the one-sequence approximation may be a useful and accurate simplified approach in such cases. For free boundary conditions, besides single interior bubbles, we must include the possibility of single helical sequences. The effective free energy of creating an interior $n$-bubble with two base-pair domain walls is given in (43); the effective free energy of creating a single end unzipped sequence of length

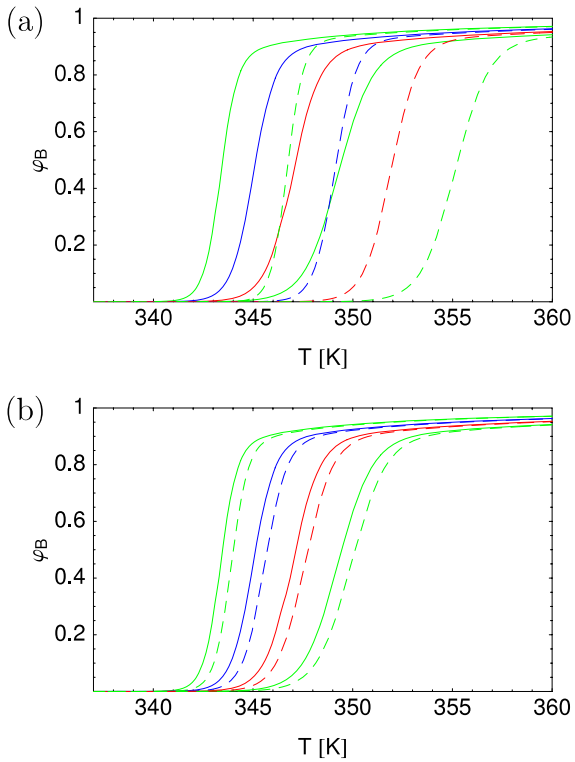


**Figure 8.** Comparison of the one-sequence approximations with loop entropy (dashed curves) and without (solid curves) (53), from left to right $N = 136$, 105, 83 and 67 for closed boundary conditions (same parameters as in figure 3) and $k = 1.7$ (a) $D = 1$, (b) $D = 100$.

$n$ starting at $i = 1$ or $N$ (with only one base-pair domain wall) is [18]:

$$\beta\Delta G^{(n)}_{\rm end} = 2\tilde{J}_0 - \tilde{K}_0 + 2n\tilde{L}_0. \tag{57}$$

The effective free energy for creating a single interior helical sequence of length $m = N - m$ (including neither the $i = 1$ nor $N$ base-pair) with two domain walls is [18]:

$$\beta\Delta G^{(m)}_{\rm helix} = 4\tilde{J}_0 - 2\tilde{K}_0 + 2(N-m)\tilde{L}_0. \tag{58}$$

The effective free energy needed to completely denature the DNA chain of length $N$ is $\beta\Delta G^{(N)}_{\rm open} = 2\tilde{L}_0 N - 2\tilde{K}_0$. The restricted one-sequence partition function for free boundary conditions, $Z^{\rm free}_{\rm 1seq}$, includes contributions from (i) the completely closed state (dsDNA), normalized to a weight of one, (ii) interior $n$-bubbles inserted in a domain of length $N-2$ varying in size between $n = 1$ and $N - 2$,

$$Z^{\rm B_{int}}_{\rm 1seq} = \sum_{m=0}^{N-3}(m+1)\exp\left[-\beta\Delta G^{(N-2-m)}_{\rm int}\right] \tag{59}$$

(iii) one unzipped end sequence of length $n$, $Z^{\rm end}_{\rm 1seq}$ with two-fold degeneracy

$$Z^{\rm end}_{\rm 1seq} = 2\sum_{n=1}^{N-1}\exp\left[-\beta\Delta G^{(n)}_{\rm end}\right] \tag{60}$$

(iv) a single interior helical sequence

$$Z^{\rm H_{int}}_{\rm 1seq} = \sum_{m=1}^{N-2}(N-1-m)^{\alpha}\exp\left[-\beta\Delta G^{(N-2-m)}_{\rm helix}\right], \tag{61}$$

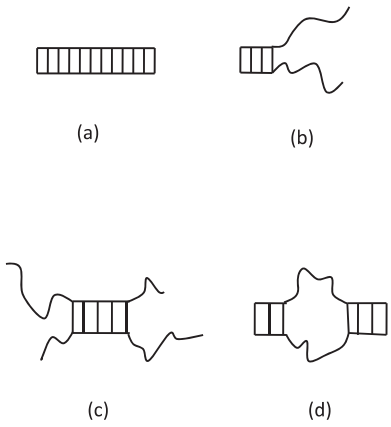


**Figure 9.** The four DNA states accounted for in the one-sequence approximation for free polymers (aside from the dissociated chains): (a) closed chain, (b) end unwinding, (c) internal helix and (d) internal bubble, corresponding, respectively to the first four terms in (63).

where $\alpha = 1$ without chain sliding (for heteropolymers using average parameter values) and 2 with (for homopolymers like polydA–polydT) [42, 61, 52], (v) the completely open state (op),

$$Z^{\rm op}_{\rm 1seq} = \exp\left[-\beta\Delta G^{(N)}_{\rm open}\right]. \tag{62}$$

$Z^{\rm free}_{\rm 1seq}$ can therefore be written as

$$Z^{\rm free}_{\rm 1seq} = 1 + Z^{\rm end}_{\rm 1seq} + Z^{\rm H_{int}}_{\rm 1seq} + Z^{\rm B_{int}}_{\rm 1seq} + Z^{\rm op}_{\rm 1seq}. \tag{63}$$

The four DNA states accounted for in the one-sequence approximation (aside from the dissociated chains) are shown in figure 9. It is now easy to include loop entropy by inserting the loop entropy factor $g_{\rm LE}$ into the fourth term of (63):

$$Z^{\rm B_{int},LE}_{\rm 1seq} = \sum_{m=0}^{N-3}(m+1)[n_0+2+2(N-m)]^{-k} \times \exp\left[-\beta\Delta G^{(N-2-m)}_{\rm int}\right]. \tag{64}$$

It is not possible now to simply readjust $J_0$ as was done for inserts, because unzipped end sequences 'see' the un-readjusted $J_0$. Unzipped end sequences are composed of two unbound chains joined at one end and therefore there is no loop entropy factor in $Z^{\rm end}_{\rm 1seq}$ or $Z^{\rm H_{int}}_{\rm 1seq}$ (a small correction term for two such self-avoiding chains, however, has been neglected, see [68]). We can, however, rewrite (64) as

$$Z^{\rm B_{int},LE}_{\rm 1seq} = \sum_{m=0}^{N-3}(m+1)[1+(N-m)/D]^{-k} \times \exp\left[-\beta\Delta\hat{G}^{(N-2-m)}_{\rm int}\right], \tag{65}$$

$\beta\Delta\hat{G}^{(n)}_{\rm int}$ is the same as $\beta\Delta G^{(n)}_{\rm int}$ with $\tilde{J}$ replaced by

$$\hat{J} \equiv \tilde{J} + (k/4)\ln(2D) > \tilde{J}. \tag{66}$$

It is then possible to define an effective loop initiation factor, $\hat{\sigma}_{\rm LI} \equiv {\rm e}^{-4\hat{J}_0} < \sigma_{\rm LI}$, that controls the probability of bubble formation in the presence of loop entropy and depends on the

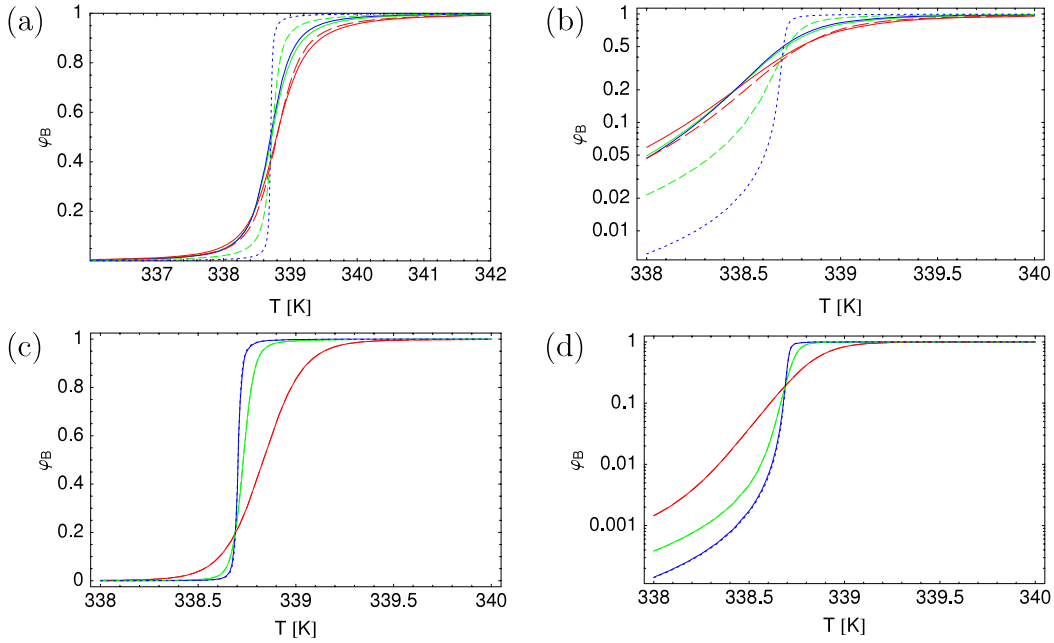


**Figure 10.** Melting curves: comparison of the exact result with the one-sequence approximation (including completely open state) for a free chain (no loop entropy, no sliding): exact results (solid curves), from right to left near the upper part of the curves ($T > 339$ K), $N = 500$, 2000, 10 000; one-sequence approximation, $N = 500$ (long dashed curve), 2000 (intermediate dashed curve), 10 000 (short dashed curve), $f = 0.5$, other parameters as in figure 3 (a) $J = 4.57$ kJ mol$^{-1}$; (b) as in (a) but now a linear-log plot; (c) $J = 9.13$ kJ mol$^{-1}$; (d) as in (c) but now a linear-log plot. In (c) and (d) the dashed and solid curves are practically superimposed.

readjusted value $\hat{J}$ (although it is still $\sigma_{\rm LI}$ that controls the probability of end unwinding and one internal helical section).

Within the free boundary condition one-sequence approximation the average fraction of broken base-pairs can be obtained from $Z^{\rm op}_{\rm 1seq}$ via

$$\varphi^{\rm free}_{\rm B,1seq}(N) = -\frac{1}{2N}\frac{\partial\left(\ln Z^{\rm free}_{\rm 1seq}\right)}{\partial L_0}. \qquad (67)$$

When chain dissociation is taken into account the contribution from the completely open chain, $Z^{\rm op}_{\rm 1seq}$, is dropped from $Z^{\rm free}_{\rm 1seq}$, which then becomes the internal partition function for associated chains:

$$Z^{\rm free}_{\rm 1seq} = 1 + Z^{\rm end}_{\rm 1seq} + Z^{\rm H_{int}}_{\rm 1seq} + Z^{\rm B_{int}}_{\rm 1seq} \quad \text{(associated chains)}. \qquad (68)$$

The corresponding $\varphi^{\rm free}_{\rm B,1seq}$ is the fraction of broken base-pairs in associated chains (clearly a lower bound for the experimentally measured total fraction of broken base-pairs, because the contribution of dissociated chains is neglected). In this case the one-sequence approximation incorporates the first four terms (of order 0, 1/2 and two terms of order 1) in an expansion in powers of the loop initiation factor, $\sigma_{\rm LI}$ (the so-called zipper model neglects the last (bubble) contribution) [42, 61]. The next higher term, neglected in (63) and of order 3/2, accounts for one internal bubble with chain sliding. In most cases of practical interest there is little difference between using (63) and (68).

The above one-sequence approximation should be valid for sufficiently short chains. After determining its range of validity when loop entropy is neglected, we can then use it with confidence within this range to examine the influence of loop entropy on DNA denaturation.

In figure 10 we test the validity of the one-sequence approximation with neither loop entropy nor chain sliding by comparing it with the exact result (35) for which the partition function includes the completely open state. From now on we fix the weighting factor $f$ at 0.5, which, as explained earlier, is close to the one estimated from experiment. We observe that the one-sequence approximation is accurate when $N \leqslant 500$ for $J = 4.57$ kJ mol$^{-1}$ (figure 10(a)) and is also accurate beyond $N \leqslant 10\,000$ for $J = 9.13$ kJ mol$^{-1}$ (figure 10(c)); in both cases studied the melting temperature is well reproduced, although the transition width is underestimated for $J = 4.57$ kJ mol$^{-1}$ when $N = 500$ (with the discrepancy increasing with increasing $N$). The one-sequence approximation also somewhat overestimates the temperature $T^*$ at which the melting curves intersect. We conclude that the limiting value of $N$ for which the one-sequence approximation is accurate depends critically on the value of $J_0$ via the loop initiation factor (45).

Because we are now interested in studying the effects of loop entropy on thermal denaturation, we employ the smaller value for $J$ ($J = 4.57$ kJ mol$^{-1}$). Despite the smaller value of $J$, the inclusion of loop entropy reinforces the validity of the one-sequence approximation. For $J = 4.57$ kJ mol$^{-1}$, $k = 1.7$,and $D = 100$ ($n_0 = 198$), the readjusted value $\hat{J}$ (66) is greater than 9.13 kJ mol$^{-1}$, implying that in this case bubbles are even more highly suppressed for $J = 4.57$ kJ mol$^{-1}$ with loop entropy than for $J = 9.13$ kJ mol$^{-1}$ without loop entropy. In figure 11 we observe that at low temperature the chain-sliding-only model gives the highest melting and the

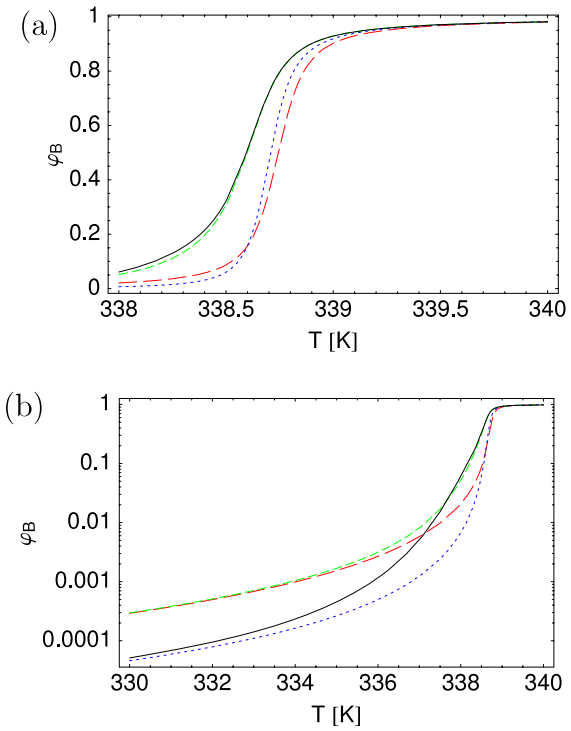


**Figure 11.** Internal melting curves (associated chains): a comparison of various one-sequence approximations for a free chain with $N = 20\,00$. (a) Linear plot, (b) linear-log plot: neither loop entropy nor sliding (long dashed curve), sliding only (intermediate dashed curve), loop entropy only (short dashed curve), both loop entropy and sliding (solid curve), ($k = 1.7$, $n_0 = 198$, $f = 0.5$, other parameters as in figure 3).

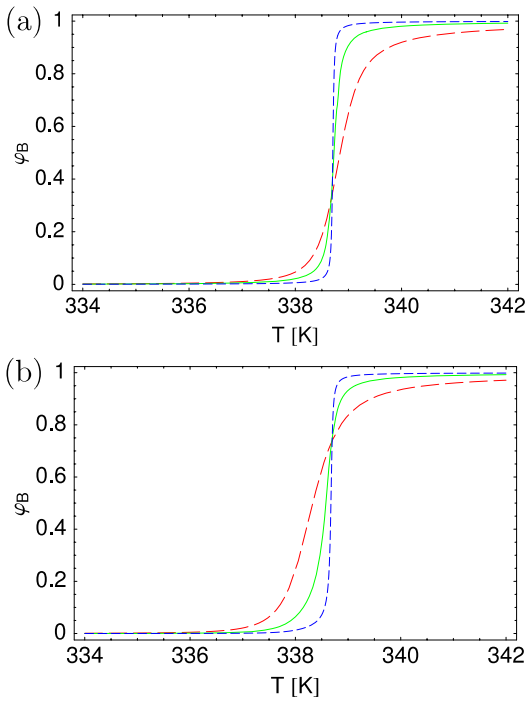


**Figure 12.** Internal melting curves (associated chains) obtained with and without loop entropy and sliding for free chains of three different lengths: $N = 500$ (long dashed curve); 2000 (solid curve); 10 000 (short dashed curve) with $J = 4.57$ kJ mol$^{-1}$, $k = 1.7$, $n_0 = 198$, $f = 0.5$ (other parameters as in figure 3): (a) with neither loop entropy, nor chain sliding; (b) with loop entropy and chain sliding.

loop entropy one the lowest. At higher temperature the sliding-loop entropy model gives the highest melting. For the case considered in figure 11, we therefore expect the accuracy of the one-sequence approximation to be comparable to that seen in figures 10(c) and (d) (and not figures 10(a) and (b)).

In figure 12 we plot the melting curves using the *loop entropy-sliding* model for free chains of three different lengths ($N = 500, 200, 10\,000$) and compare the results obtained without loop entropy and sliding. We note that due to the combined effects of sliding and loop entropy the melting temperature increases with increasing $N$ and the width of the transition decreases (figure 12(b)), in agreement with experiment [1] (for $f = 0.5$ the temperature $T^*$ at which the melting curves intersect is now greater than $T_m(N)$, the opposite of what occurs when loop entropy and sliding are neglected, see figure 12(a)). The model prediction for the difference between the melting temperatures for $N = 500$ and 10 000 is about 0.5 K (the results for $N > 10\,000$ should be very close to the $N = 10\,000$ one). When chain dissociation is added to the model, one can reasonably expect that the melting temperature for $N = 500$ will decrease by about 0.5 K [42, 61] and that for $N \geqslant 10\,000$ it will hardly change. This result suggests that once chain dissociation is incorporated into the current model, it should be possible to account for the experimental results of [1] ($T_m(30\,000) - T_m(500) \simeq 1$ K and decreasing transition width as $N$ increases).

## 4. Concluding remarks

This paper presents an extension of a theoretical model of DNA denaturation [17, 18] that couples the base-pair states, unbroken or broken, and the chain configurational degrees of freedom. The elastic contributions are taken into account, arising from chain bending, torsional and stretching rigidities, the values of which depend on the neighboring base-pair states. The difference of bond lengths in ssDNA (0.34 nm) and dsDNA (0.71 nm) is also included in the Hamiltonian. This model, tackled by analytical means, provides new insight into the dependence of the effective Ising parameters, used in previous Ising-like models, on microscopic elastic moduli. The main conclusion is that all these features lead to a renormalization of the bare Ising parameters of the same order of magnitude as the thermal energy. Hence, they cannot be ignored when relating microscopic properties, extracted for example from *ab initio* calculations or experiments on DNA fragments, to the collective properties of the whole chain, measured, for instance, in single DNA molecule experiments (atomic force microscopy, optical and magnetic tweezers, tethered particle motion). As an illustration, without considering the effects of stretching elasticity and base-pair length, the energy cost to open a base-pair, $2\mu$, would be directly related to the same quantity measured with a force apparatus [64][4]. But $\mu$ is renormalized by these effects and

[4] Even though the (screened) electrostatic repulsion of the phosphate atoms of the sugar DNA backbone is not taken into account in those experiments and is also likely to decrease the real $\mu$ [69].

is lowered by 0.5–1 $k_BT$ when the bare value is close to 2 $k_BT$. The same conclusion holds for the destacking, $J$, or stacking, $K$, parameters.

In this work, we also analyze finite size effects. In particular the role of closed boundary conditions on melting curves for finite lengths is investigated in order to model a clamped polydA–polydT DNA insert. Two approximations are considered: (i) the one-sequence approximation amounts to neglecting configurations with several bubbles and (ii) the two-state one keeps only the contributions from the completely closed and open chains [1]. In the range of parameters studied, the agreement with the exact result is excellent in case (i), whereas it is much less satisfactory in case (ii). We also undertake the integration of loop entropy in case (i), which leads to an increase in $T_m$ that is associated with the loop entropy cost and depends on the value of the loop entropy chain stiffness parameter $D$ (for $N \sim 100$ there is a shift of 1 K for $D = 100$ and of 5 K for $D = 1$). Finally, we study free polymer chains using exact results with neither loop entropy nor chain sliding and the one-sequence approximation with loop entropy and chain sliding. Our major conclusion is that the experimentally observed increase in $T_m$ with increasing chain length for homopolymers can be accounted for by incorporating both loop entropy and chain sliding into our model. The simplicity of our method of incorporating loop entropy into the one-sequence approximation paves the way to a deeper study of the role of chain stiffness in the loop entropy factor, $g_{LE}$. We underline that careful experiments on free and clamped homopolymers of different lengths (in solution or in single molecule experiments) would be extremely useful in elucidating the role of DNA finite size effects.

From an experimental perspective, our findings are relevant for free DNA in dilute solutions, without any constraint on chain configurations, nor any applied force or torque. An ingredient that we have not considered so far is the gain in translational entropy due to strand separation in the case of dissociation [66]. A correct treatment of this mechanism consists in writing a chemical equilibrium between completely denatured single strands and partially bound ones (work in progress).

The case of constrained DNA is more involved. If a force or a torque is applied, for instance in tweezer experiments, rotational symmetry is lost in the Hamiltonian, which prevents an analytical solution of the problem. Numerical or approximate schemes, such as variational principles, may be used. Another interesting constraint concerns polymer looping [60]. Circular DNAs appear in the case of transposons or insertion sequences [40, 31]. Writing down the polymer closure (e.g. for the determination of the $J$-factor) is a formidable task because it corresponds to the global constraint $\sum \mathbf{t}_i = 0$, formally equivalent to an applied force [12]. We can, however, partially take into account looping in our framework by imposing periodic boundary conditions on the vectors $\hat{\mathbf{e}}_{\mu,i}$ and/or on $\sigma_i$, instead of the end condition $|V\rangle$. This can be handled using the transfer matrix method. In the case of superhelical twist, the polymer winds one or several times around its tangent vectors $\mathbf{t}_i$. This condition can also be enforced via the boundary conditions, by requiring that the appropriate combination of Euler angles acquires a phase multiple of $2\pi$ when going from $i = N$ to 1. This topological constraint should lead to an increased fraction of denatured base-pairs, in order to release the torsional energy cost, and consequently to an increased flexibility, thereby facilitating cyclization. Our predictions for the end-to-end distance can also be compared to experiment, because $R$ is proportional to the radius of gyration, which can be measured in viscosity experiments.

All the results presented in this paper concern homopolynucleotides and the numerical applications focused on polydA–dT. This work can, however, be generalized to heteropolymers, although a minimal amount of numerical work is necessary to handle the reduction of the transfer matrices. Nonetheless, a numerical study of heteropolymers would require a knowledge of the microscopic elastic moduli, which are far from being known with any certainty for any pair of the four nucleotides A, T, G and C.

## Appendix

In this appendix we extract smooth melting curves from the experimental data in [1]. For the poly dA–dT DNA polymer with free ends and 30 000 base-pairs we have used the temperature derivative of

$$\varphi_{\rm fit} = \frac{c_f}{2}\left[1 - \frac{\sinh(-a_f + \beta\mu_f)}{\sqrt{e^{-4\beta J_f} + \sinh^2(-a_f + \beta\mu_f)}}\right], \qquad (69)$$

where $c_f$, $a_f$ and $\mu_f$ are fitting parameters (simplified $N = \infty$ Ising form); this functional form arises in simple Ising models of DNA denaturation [52].

For A–T inserts we have used the temperature derivative of

$$\varphi_{\rm fit} = \frac{c_f}{2h_f}\{1 - \tanh[h_f(T_f - T)]\}, \qquad (70)$$

where $c_f$, $h_f$ and $T_f$ are fitting parameters; this functional form arises in a two-state treatment of simple Ising models of DNA denaturation [42] (the use of the two-state form here to extract smooth experimental melting curves does not imply that the two-state approximation is a valid one, see figure 6). As shown in figure A.1 the areas under the fitted $d\varphi_B/dT$ functions are not normalized to one. We thus assume that the normalized fitted $\varphi_B$ functions (figure A.2) represent a good approximation to the fraction of open base-pairs for the A–T segments. By examining figure A.1 we see that this assumption is well borne out for the A–T inserts, but less so for the $N = 30\,000$ base-pair chain because of difficulties in reading the data off the experimental curve and the asymmetry of this curve. Our choice of fitting functions gives symmetric curves about the melting temperature and thus cannot account for the observed asymmetry for $N = 30\,000$. The observed asymmetry probably cannot be explained by loop entropy and chain sliding because (for infinite chains at least) when they are included in the model, the melting curves become flatter to the left of the melting temperature and steeper to the right (the opposite of what is observed in figure A.1). For finite chains, however, the combined effects of loop entropy and

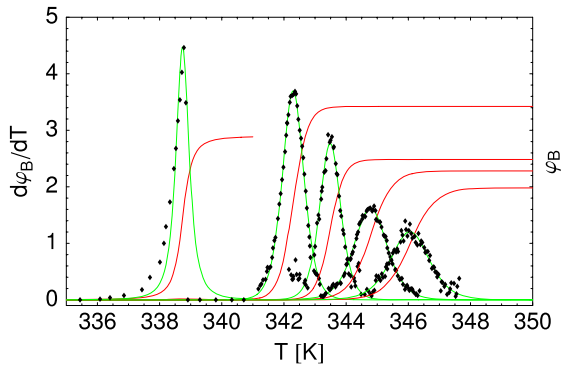


**Figure A.1.** Absorbance temperature derivative (unnormalized $d\varphi_B/dT$) versus temperature: experimental data points [1] and unnormalized fitted functions (green non-monotonic curves, left $y$-axis); UV absorbance (unnormalized fraction of broken base-pairs, $\varphi_B$) versus temperature (red monotonic curves, right $y$-axis) (from left to right: $N = 30\,000$, 136, 105, 83 and 67).

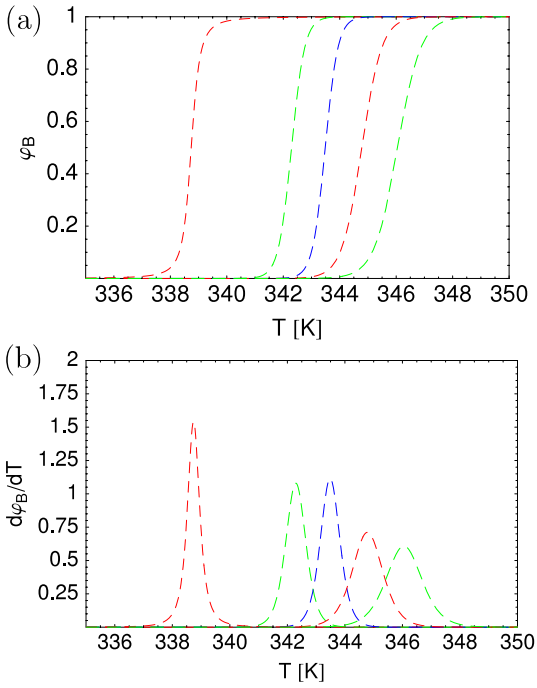


**Figure A.2.** Normalized functions fitted to the experimental data [1]: (a) fraction of broken base-pairs versus temperature, $\varphi_B$; (b) $d\varphi_B/dT$ versus temperature (from left to right, $N = 30\,000$, 136, 105, 83 and 67).

chain sliding can be different, see figure 11. Although the $N = 30\,000$ base-pair chain melting temperature ∼339 K is well reproduced, the width of the transition appears to be overestimated. The general trend is for both the melting temperature and transition width to decrease with increasing $N$. As the length of the insert increases the melting should tend to the infinite free chain result.

## References


[1] Blake R D and Delcourt S G 1998 *Nucleic Acids Res.* **26** 3323
[2] Blake R D *et al* 1999 *Bioinformatics* **15** 370
[3] Peyrard M and Bishop A R 1989 *Phys. Rev. Lett.* **62** 2755
[4] Dauxois T, Peyrard M and Bishop A R 1993 *Phys. Rev.* E **47** 684
[5] Peyrard M 2004 *Nonlinearity* **17** R1
[6] Cule D and Hwa T 1997 *Phys. Rev. Lett.* **79** 2375
[7] Kafri Y, Mukamel D and Peliti L 2000 *Phys. Rev. Lett.* **85** 4988
[8] Carlon E, Orlandini E and Stella A L 2002 *Phys. Rev. Lett.* **88** 198101
[9] Storm C and Nelson P C 2003 *Europhys. Lett.* **62** 760
[10] Blossey R and Carlon E 2003 *Phys. Rev.* E **68** 061911
[11] Jeon J-H, Sung W and Ree F H 2006 *J. Chem. Phys.* **124** 164905
[12] Yan J and Marko J F 2004 *Phys. Rev. Lett.* **93** 108108
[13] Mazur A K 2007 *Phys. Rev. Lett.* **98** 218102
[14] Hanke A, Ochoa M G and Metzler R 2008 *Phys. Rev. Lett.* **100** 018106
[15] Everaers R, Kumar S and Simm C 2007 *Phys. Rev.* E **75** 041918
[16] Joyeux M, Buyukdagli S and Sanrey M 2007 *Phys. Rev.* E **75** 061914
[17] Palmeri J, Manghi M and Destainville N 2007 *Phys. Rev. Lett.* **99** 088103
[18] Palmeri J, Manghi M and Destainville N 2008 *Phys. Rev.* E **77** 011913
[19] Altan-Bonnet G, Libchaber A and Krichevsky O 2003 *Phys. Rev. Lett.* **90** 138101
[20] Zeng Y, Montrichok A and Zocchi G 2003 *Phys. Rev. Lett.* **91** 148101
[21] Ivanov V, Zeng Y and Zocchi G 2004 *Phys. Rev.* E **70** 051907
[22] Pouget N *et al* 2004 *Nucleic Acids Res.* **32** 73 (erratum)
[23] Du Q, Simth C, Shiffeldrim N, Vologodskaia M and Vologodskii A 2005 *Proc. Natl Acad. Sci. USA* **102** 5397
[24] Wiggins P A *et al* 2006 *Nat. Nanotechnol.* **1** 137
[25] Linna R P and Kasti K 2008 *Phys. Rev. Lett.* **100** 168104
[26] Moukhtar J, Fontaine E, Faivre-Moskalenko C and Arneodo A 2007 *Phys. Rev. Lett.* **98** 178101
[27] Yuan C, Chen H, Lou X W and Archer L A 2008 *Phys. Rev. Lett.* **100** 018102
[28] Smith S B, Finzi L and Bustamante C 1992 *Science* **258** 1122
[29] Svoboda K and Block S M 1994 *Annu. Rev. Biophys. Biomol. Struct.* **23** 247
[30] Finzi L and Gelles J 1995 *Science* **267** 378
[31] Pouget N *et al* 2006 *Nucleic Acids Res.* **34** 4313
[32] Segall D E, Nelson P C and Phillips R 2006 *Phys. Rev. Lett.* **96** 088306
[33] Nelson P C *et al* 2006 *J. Phys. Chem.* B **110** 17260
[34] Ke C, Humeniuk M, Hanna S-Gracz H and Piotr Marszalek E 2007 *Phys. Rev. Lett.* **99** 018302
[35] Cluzel P, Lebrun A, Heller C, Lavery R, Viovy J-L, Chatenay D and Caron F 1996 *Science* **271** 792
[36] Strick T R, Bensimon D and Croquette V 1999 *Genetica* **106** 57
[37] Bryant Z, Stone M D, Gore J, Smith S B, Cozzarelli N R and Bustamante C 2003 *Nature* **424** 338
[38] Maiorano D, Lutzmann M and Méchali M 2006 *Curr. Opin. Cell Biol.* **18** 130
[39] Cloutier T E and Widom J 2005 *Proc. Natl Acad. Sci. USA* **102** 3645
[40] Alberts B *et al* 2002 *Molecular Biology of the Cell* 4th edn (New York: Garland Science)
[41] Landau L D and Lifshitz E M 1986 *Theory of Elasticity* 3rd edn (Oxford: Butterworth–Heinemann)
[42] Poland D and Scheraga H A 1970 *Theory of Helix Coil Transitionin Biopolymers* (New York: Academic)
[43] Fixman M and Freire J J 1977 *Biopolymers* **16** 2693
[44] Bustamante C, Smith S B, Liphardt J and Smith D 2000 *Curr. Opin. Struct. Biol.* **10** 279
[45] Hugel T, Rief M, Seitz M, Gaub H E and Netz R R 2005 *Phys. Rev. Lett.* **94** 048301

[46] Peck L J and Wang J C 1981 *Nature* **292** 375
[47] Yamakawa H 1997 *Helical Wormlike Chains in Polymer Solutions* (Berlin: Springer) chapter 4
[48] Benham C J 1992 *J. Mol. Biol.* **225** 835
[49] Moroz J D and Nelson P 1998 *Macromolecules* **31** 6333
[50] Bouchiat C and Mézard M 2000 *Eur. Phys. J.* E **2** 377
[51] Bloomfield V, Crothers D and Tinico I 1974 *Physical Chemistry of Nucleic Acids* (New York: Harper and Row)
[52] Wartell R M and Montroll E W 1972 *Adv. Chem. Phys.* **22** 129
[53] Shi Y, He S and Hearst J E 1996 *J. Chem. Phys.* **105** 714
[54] Kamien R D, Lubensky T C, Nelson P and O'Hern C S 1997 *Europhys. Lett.* **38** 237
[55] O'Hern C S, Kamien R D, Lubensky T C and Nelson P 1998 *Eur. Phys. J.* B **1** 95
[56] Jian H, Schlick T and Vologodskii A 1998 *J. Mol. Biol.* **284** 287
[57] Smith S B, Cui Y and Bustamante C 1996 *Science* **271** 795
[58] Hegner M, Smith S B, Cui Y and Bustamante C 1999 *Proc. Natl Acad. Sci. USA* **96** 10109
[59] Cocco S, Marko J F, Monasson R, Sarkar A and Yan J 2003 *Eur. Phys. J.* E **10** 249
[60] Fye R M and Benham C J 1999 *Phys. Rev.* E **59** 3408
[61] Poland D and Scheraga H A 1969 *Physiol. Chem. Phys.* **1** 389
[62] Buyukdagli S and Joyeux M 2007 *Phys. Rev.* E **76** 021917
[63] Krueger A, Protozanova E and Frank-Kamenetskii M D 2006 *Biophys. J.* **90** 3091
[64] Pincet P, Perez E, Bryant G, Lebeau L and Mioskowski C 1994 *Phys. Rev. Lett.* **73** 2780
[65] Gotoh O 1983 *Adv. Biophys.* **16** 1
[66] Wartell R M and Benight A S 1985 *Phys. Rep.* **126** 67
[67] Oliver A L, Wartell R M and Ratliff R L 1977 *Biopolymers* **16** 115
[68] Garel T, Monthus C and Orland H 2001 *Europhys. Lett.* **55** 132
[69] Perez E 2008 private communication