Eur. Phys. J. B **27**, 135–146 (2002)
DOI: 10.1140/epjb/e20020138

THE EUROPEAN PHYSICAL JOURNAL B
EDP Sciences

Melting and unzipping of DNA

Y. Kafri[1,a], D. Mukamel[1], and L. Peliti[2]

[1] Department of Physics of Complex Systems, The Weizmann Institute of Science, Rehovot 76100, Israel
[2] Dipartimento di Scienze Fisiche and Unità INFM, Università "Federico II", Complesso Monte S. Angelo, 80126 Napoli, Italy

Received 21 August 2001 and Received in final form 26 January 2002

Abstract. Existing experimental studies of the thermal denaturation of DNA yield sharp steps in the melting curve suggesting that the melting transition is first order. This transition has been theoretically studied since the early sixties, mostly within an approach in which the microscopic configurations of a DNA molecule consist of an alternating sequence of non-interacting bound segments and denaturated loops. Studies of these models neglect the repulsive, self-avoiding, interaction between different loops and segments and have invariably yielded continuous denaturation transitions. In the present study we take into account in an approximate way the excluded-volume interaction between denaturated loops and the rest of the chain. This is done by exploiting recent results on scaling properties of polymer networks of arbitrary topology. We also ignore the heterogeneity of the polymer. We obtain a first-order melting transition in $d = 2$ dimensions and above, consistent with the experimental results. We also consider within our approach the unzipping transition, which takes place when the two DNA strands are pulled apart by an external force acting on one end. We find that the under equilibrium condition the unzipping transition is also first order. Although the denaturation and unzipping transitions are thermodynamically first order, they do exhibit critical fluctuations in some of their properties. For instance, the loop size distribution decays algebraically at the transition and the length of the denaturated end segment diverges as the transition is approached. We evaluate these critical properties within our approach.

PACS. 87.14.Gg DNA, RNA – 05.70.Fh Phase transitions: general studies – 64.10.+h General theory of equations of state and phase equilibria – 63.70.+h Statistical mechanics of lattice vibrations and displacive phase transitions

1 Introduction

The unbinding phase transition of the two complementary strands of the DNA molecule has been a subject of continual interest for over four decades [1–7]. In thermal denaturation this transition takes place when temperature is increased. In a typical experiment, a sample containing DNA molecules of specific length and sequence is prepared. The fraction of attached bound base pairs, θ, is measured through light absorption at a wavelength of 260 nm. At low temperatures all base pairs are attached to each other while at high temperature they are all unbound. Thus, θ decreases from one to zero as the temperature is increased. For heterogeneous DNA molecules, containing both AT and GC base pairs, θ does not decrease smoothly with temperature, but rather exhibits a multistep behavior. It consists of plateaus of various sizes separated by sharp jumps. This behavior is related to the fact that the GC bonds are stronger than AT ones. Thus, long domains with higher concentration of AT bonds will denaturate at lower temperatures. The resulting stepped structure of θ is therefore characteristic of the particular DNA sequence. It thus yields statistical information on the sequence of the molecule under study. The denaturation process has been verified through electron microscopy where denaturated loops and bound segments have been observed directly [8]. The sharpness of the jumps indicates that the unbinding transition is *first order*.

More recently, the introduction of new techniques such as optical tweezers and atomic force microscopy [9,10] has allowed the manipulation of single biological molecules. This made it possible to study a wider variety of physical properties of the DNA molecule. For example optical tweezers have been used to apply a force and pull apart the two strands at one end of the molecule. It is found that a phase transition takes place at a critical force where the molecule is unzipped and the two strands are separated [11].

Thermal denaturation has been studied theoretically since the early sixties. The early models, which we refer to as Poland-Scheraga (PS) type models [2,3], consider the molecule as being composed of an alternating sequence of

[a] e-mail: fekafri@wisemail.weizmann.ac.il

bound and denaturated segments. A bound segments is energetically favored over an unbound segment, while a denaturated segment (loop) is entropically favored over a bound one. Moreover, each base on a given strand can only bind to a specific matching base on the other strand and the binding energy is taken to the same for all base pairs. Thus, the heterogeneity of the chain and mismatch pairing (both within and between strands) are ignored. In the PS approach the interaction between a loop or a bound segment and the rest of the chain is ignored. This assumption simplifies the analysis considerably. The order of the transition is found to be determined by a parameter c which characterizes the statistical weight of a loop. The number of configurations of a loop of length ℓ behaves as s^ℓ/ℓ^c for large ℓ. Here s is a non-universal constant. It has been shown [12] that the phase transition is first order if $c > 2$ and second order if $1 < c \leq 2$, while for $c \leq 1$ no transition takes place and the strands are always bound. Using random-walk configurations to model a loop one finds that $c = d/2$ in d dimensions. The transition is thus predicted to be continuous in $d = 3$ dimensions [12]. This result is at variance with experimental observations. The model was later generalized to take into account the self-avoiding interactions *within* each loop [13]. It is found that the loop entropy takes the same general form as before [14]. However, the exponent c now takes the value $d\nu$, where ν the correlation length exponent of a self-avoiding random walk. Inserting the known values for ν one finds that although c is larger than that of a random-walk model, it is still smaller than 2 both in $d = 2$ and $d = 3$ dimensions, yielding a continuous transition. It was suggested [13] that self-avoiding interactions *between* the various parts of the chain (and not just within loops) would further sharpen the transition possibly making it first order. However, theoretical tools for carrying out this analysis have not been available at the time. More recently, excluded volume interactions have been fully taken into account in a numerical study of finite chains [15]. These simulations strongly suggest that the transition is indeed first order.

In a recent study [16], we have extended the PS model to take into account self-avoiding interactions both within a loop and between a loop and the rest of the chain. To carry out the analysis of this model one has to enumerate the configurations of a loop embedded in a chain with self-avoiding interactions. This has been done by taking advantage of recent results obtained by Duplantier *et al.* [17,18] for the number of configuration of a general polymer network. It is found that the statistical weight of a loop embedded in a chain has the same general form as before, namely s^ℓ/ℓ^c. However, the parameter c is now modified and becomes larger than two in $d \geq 2$ dimensions. In particular one finds $c \simeq 2.115$ in $d = 3$ and $c = 2 + 13/32$ in $d = 2$. Thus, self-avoiding interactions make the transition first order in two dimensions and above. Recently a different model in which excluded volume interactions were partially taken into account has been found to yield a first order transition [19]. In this model excluded volume interactions between the two strands of the chain were explicitly considered, but those within each strand were neglected.

In this work we present a detailed account of the results obtained when the interactions between various segments along the chain are taken into account using the scaling results of Duplantier. We show that the denaturation transition is first order. However, the transition is found to be accompanied by critical fluctuations in some the chain's properties. For example, the loop size distribution is found to decay algebraically at the transition. Indeed, the probability distribution for loops of length ℓ, $P(\ell)$, behaves as $P(\ell) \sim 1/\ell^c$ at the transition. This behavior was recently confirmed in numerical simulations of the model where the excluded volume interactions have been taken into account fully. The value of the measured exponent $c \sim 2.10 \pm 0.02$ agrees well with our predictions [20]. We also find that when the boundary conditions are such that the chain is open at one end, the end segment length, as given by number of unbound monomers at the chain end ξ, diverges as $1/|T - T_{\rm M}|$ when the melting temperature $T_{\rm M}$ is approached.

We have extended the model to consider the unzipping transition which takes place when a force of magnitude f is applied to separate the two strands. We find that under equilibrium conditions the transition is first order. We also find that the end segment length diverges as $\xi \sim 1/|f - f_{\rm U}|$ where $f_{\rm U}$ is the unzipping critical force. This behavior has been previously found in models where self avoiding interactions were not taken into account [21–25]. In calculating the critical force near the melting transition we find that $f_{\rm U} \sim |T - T_{\rm M}|^\nu$.

The paper is organized as follows. In Section 2 we consider the thermal denaturation transition in detail. In Section 2.1 we review the analysis of Poland and Scheraga and that of Fisher where self-avoiding interaction within a loop is taken into account. In Section 2.2 we analyze the model where self avoiding interactions are accounted for, not only within a loop but also between loops and the rest of the chain. The length distribution of the end segment is considered in Section 2.3 and a summary and overview of other approaches to thermal denaturation of DNA is given in Section 2.4. In Section 3 the unzipping transition is studied. A brief summary is given in Section 4.

2 Thermal denaturation of DNA

2.1 The model and basic analysis

The model considers two strands, each composed of monomers. Each monomer represents one persistence length of a single unbound strand. Typically this is about ∼40 Å [26], or roughly 8 bases. The persistence length of double stranded DNA is an order of magnitude longer [27]. We set the boundary conditions such that the monomers at one end of the molecule are bound. Such a boundary condition is necessary for a bound state between the two strands to exist. All other monomers on the chain can be either unbound or bound to a specific matching monomer on the second chain. The interactions between a

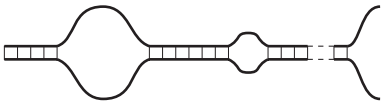

Fig. 1. Schematic representation of a microscopic configuration of the DNA molecule.

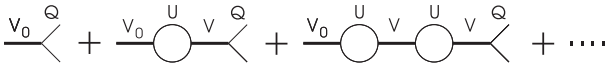

Fig. 2. Graphic illustration of the expansion of the partition function (2) in $U(z)V(z)$. Each segment of type U_0, U, V or Q represents a sum over all possible lengths of its type weighted properly with a fugacity.

monomer and other monomers on the second strand or on the same strand are ignored. The binding energy $E_0 < 0$ between matching monomers is taken to be the same for all monomer pairs.

A typical DNA configuration is shown in Figure 1. It is made of an alternating sequence of bound segments and denaturated loops. The configuration ends with two denaturated strands. For simplicity, the configurational entropy of a bound segment associated with its embedding in ambient space is neglected. It is easy to verify that this assumption does not affect the nature of the denaturation transition obtained within this model. The statistical weight of a bound sequence of length ℓ is then given by $w^\ell = \exp(-\ell E_0/T)$, where T is the temperature and the Boltzmann constant k_B is set to 1. Thus, w is a decreasing function of the temperature. On the other hand, a denaturated loop does not carry an energy and its statistical weight is derived from its degeneracy. In this model it is assumed that the loop is fully flexible, and thus it is described by a random walk which returns to the origin after 2ℓ steps. Considering all possible such walks the statistical weight for large ℓ has the form $\Omega(2\ell) = As^\ell/\ell^c$, where s is a non-universal constant and the exponent c is determined by the properties of the loop configurations. For simplicity, we set $A = 1$. Finally, the statistical weight of the end segment, which consists of two denaturated strands each of length ℓ, takes the form $\Lambda(2\ell) = Bs^\ell/\ell^{\bar{c}}$ for large ℓ, where $\bar{c}$ is in general not equal to c. Again, for simplicity we set $B = 1$. The values of the exponents c and $\bar{c}$ will be held arbitrary for the moment. We shall later discuss them in detail.

Using the weights assigned to each segment of the chain the total weight of any given configuration may be calculated. For example the weight of a chain which consists of a bound segment of length ℓ_1, a denaturated loop of length ℓ_2, a bound segment of length ℓ_3, and a pair of denaturated strands of length ℓ_4, is given by

$$w^{\ell_1}\Omega(2\ell_2)w^{\ell_3}\Lambda(2\ell_4). \tag{1}$$

The statistical weight of a more general chain configuration made of p alternating bound segments and denaturated loops will have the same form with a suitable number of factors of the form $\Omega(2\ell_i)w^{\ell_{i+1}}$ before the end-segment weight $\Lambda(2\ell_p)$.

The model is most easily studied within the grand canonical ensemble where the total chain length L is allowed to fluctuate. The grand canonical partition function, $\mathcal{Z}$, is given by

$$\mathcal{Z} = \sum_{L=0}^{\infty} Z(L)\, z^L = \frac{V_0(z)Q(z)}{1 - U(z)V(z)}, \tag{2}$$

where $Z(L)$ is the canonical partition function of a chain of length L, z is the fugacity, and the functions $U(z)$, $V(z)$ and $Q(z)$ are defined by

$$U(z) = \sum_{\ell=1}^{\infty} \Omega(2\ell) z^\ell = \sum_{\ell=1}^{\infty} \frac{s^\ell}{\ell^c}\, z^\ell = \Phi_c(zs), \tag{3}$$

$$V(z) = \sum_{\ell=1}^{\infty} w^\ell z^\ell, \tag{4}$$

$$\begin{aligned} Q(z) &= 1 + \sum_{\ell=1}^{\infty} \Lambda(2\ell) z^\ell = 1 + \sum_{\ell=1}^{\infty} \frac{s^\ell}{\ell^{\bar{c}}}\, z^\ell \\ &= 1 + \Phi_{\bar{c}}(zs), \end{aligned} \tag{5}$$

with $V_0(z) = 1 + V(z)$. In this equation, $\Phi_c(z)$ is the polylog function whose basic properties are summarized in Appendix A. Equation (2) can be verified by expanding the partition function as a series in $U(z)V(z)$. The factors $V_0(z)$ and $Q(z)$ properly account for the boundaries. A graphical illustration of the series expansion in $U(z)V(z)$ is given in Figure 2. To set the average chain length, L, one has to choose a fugacity such that

$$L = \partial \ln \mathcal{Z} / \partial \ln z. \tag{6}$$

This implies that the thermodynamic limit $L \to \infty$ is obtained by letting z approach the lowest fugacity z^* for which the partition function (2) diverges. This can arise either from the divergence of the numerator or from the vanishing of the denominator. The relevant situation, at low temperature, is the second one, which corresponds to z^* satisfying

$$U(z^*)V(z^*) = 1. \tag{7}$$

Since $V(z) = wz/(1 - wz)$ this reduces to

$$U(z^*) = 1/(wz^*) - 1. \tag{8}$$

We shall see that above the transition, namely in the denaturated phase, the numerator diverges. Moreover, when one considers the problem of DNA unzipping by applying an external force on the strands, a divergence arising from a boundary factor will play an important role.

The fraction of bound monomer pairs θ is the experimentally measured quantity and the order parameter of the transition. Its temperature dependence in the thermodynamic limit $\langle L\rangle \to \infty$ can be calculated from the behavior of $z^*(w)$. The average number of bound pairs in a chain is given by $\langle m\rangle = \partial \ln \mathcal{Z}/\partial \ln w$, so that

$$\theta = \lim_{L\to\infty} \frac{\langle m\rangle}{\langle L\rangle} = \frac{\partial \ln z^*}{\partial \ln w}\,. \tag{9}$$

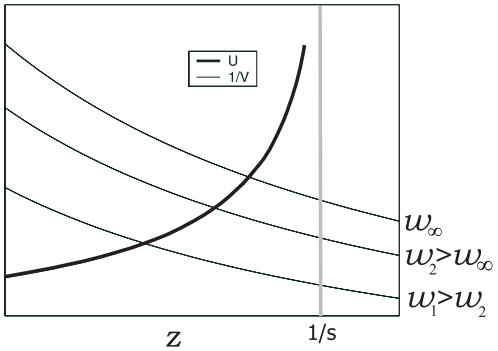

Fig. 3. A typical behavior of the functions U and $1/V$ for $c < 1$ (here $c = 0.5$). See text for explanation.

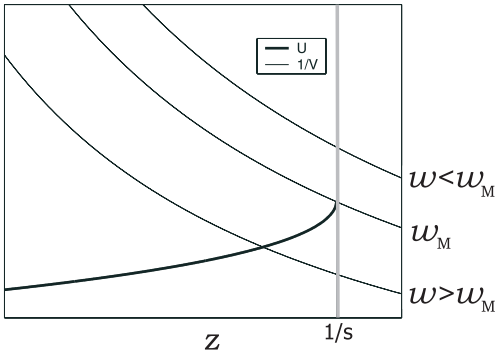

Fig. 4. A typical behavior of the functions U and $1/V$ for $1 < c \leq 2$ (here $c = 1.5$). See text for explanation.

Thus the nature of the denaturation transition is determined by the temperature dependence of the fugacity $z^*(w)$. This behavior can be classified into three distinct regimes depending on the value of the exponent c. These regimes are most easily understood through a graphical solution of (7).

Case (i): c ≤ 1. No phase transition

A schematic representation of the graphical solution of (7) in this case is given in Figure 3. The function $U(z)$ is finite for any $z < z_M = 1/s$. Since the sum

$$U(1/s) = \sum_{\ell=1}^{\infty} \frac{1}{\ell^c}, \tag{10}$$

diverges for $c \leq 1$, the function $U(z)$ increases smoothly to infinity as z approaches $1/s$. For a given value of z the function $1/V(z)$ increases as the temperature increases. In Figure 3, $1/V(z)$ is plotted *vs.* z for three values of w. One can see that as the temperature is increased the crossing point z^* of the two graphs increases smoothly until it saturates at $w = w_\infty$ ($T = \infty$). Therefore θ decreases smoothly as temperature is increased and no phase transition takes place. In this case the strands are always bound at all temperatures.

Case (ii): 1 < c ≤ 2. Continuous phase transition

A schematic representation of the graphical solution of equation (7) in this case is given in Figure 4. Here the sum (10) is finite at $z = z_M = 1/s$ since $c > 1$. The function $U(z) = \Phi_c(zs)$ increases smoothly to a finite value as z approaches z_M and becomes infinite for $z > z_M$. In Figure 4 the function $1/V(z)$ is plotted for three values of w. One can see that as temperature is increased z^* increases until it reaches $1/s$ at $w = w_M$. Above the transition, for $w < w_M$, z^* remains equal to $1/s$ in the thermodynamic limit. Here, the $\langle L \rangle \to \infty$ limit is obtained through the divergence of the factor $Q(z = 1/s)$ in the numerator. A more careful analysis of the denaturated regime is presented in Appendix B. Note that for a transition to take place one must have $1/V(z = 1/s, w = 1) \geq U(1/s)$. Otherwise there is no phase transition and the two strands are bound at all temperatures. Since $c \leq 2$ the derivative of $U(z)$ diverges at the transition. This implies that $\theta = \partial \ln z^* / \partial \ln w$ approaches zero continuously, yielding a continuous transition. Since the derivative $\partial \ln z^* / \partial \ln w$ decreases with increasing c, the closer c to two, the sharper the transition. In this case it can be shown [13] that near the transition the fraction of attached monomers behaves as $\theta \sim |T - T_M|^\beta$ with $\beta = (2 - c)/(c - 1)$.

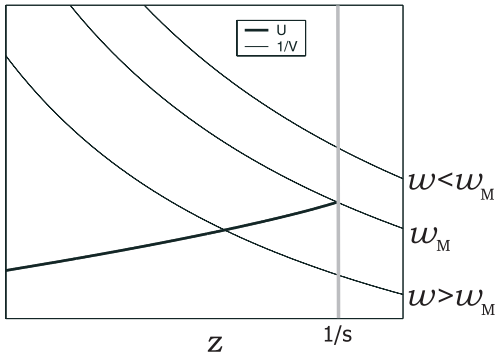

Fig. 5. A typical behavior of the functions U and $1/V$ for $c > 2$ (here $c = 2.5$). See text for explanation.

Case (iii): c > 2. First order phase transition

A schematic representation of the graphical solution of equation (7) in this case is given in Figure 5. Here both $U(z)$ and its derivative are finite at $z = z_M = 1/s$. As in the previous case there is a transition for $w = w_M$. However, since the derivative of $U(z)$ is finite, θ approaches a finite value as the transition is approached from below. Above the transition θ vanishes in the thermodynamic limit. The transition is therefore first-order. The discussion of the high-temperature phase is again deferred to Appendix B.

To summarize, three different scenarios exist depending only on the value of c. These are

$$\begin{aligned} c \leq 1: &\quad \text{no phase transition;} \\ 1 < c \leq 2: &\quad \text{continuous phase transition;} \\ c > 2: &\quad \text{first-order phase transition.} \end{aligned}$$

The nature of the phase transition is thus directly related to the number of configurations of long denaturated loops within the chain. In the early studies of this problem the exponent c was evaluated by counting all random walks with a given length which return to the origin [12]. It is easy to show that in d dimensions the model yields $c = d/2$. This implies that there is no transition for $d \leq 2$, a continuous transition for $2 < d \leq 4$ and a first-order phase transition for $d > 4$. The model was subsequently extended to include the repulsive short range interaction which exists between the strands constituting a loop. In this approach the loop is modeled as a self-avoiding walk [13]. This yields $c = d\nu$, where ν is the exponent associated with the radius of gyration $R_{\rm G}$ of a self-avoiding random walk. For a walk of length L one has $R_{\rm G} \sim L^{\nu}$, with $\nu = 3/4$ in $d = 2$ and $\nu \approx 0.588$ in $d = 3$. This yields $c = 3/2$ in $d = 2$ and $c = 1.766$ in $d = 3$. Thus the transition is continuous in both cases, although it is sharper than when the repulsive interaction is neglected altogether.

The two estimates of the exponent c described above treat the loop as an isolated object and thus neglect its interaction with the rest of the chain. This simplification seems essential, since the formalism of the model relies on the segments composing the chain as being independent. In the next section we show that the repulsive interaction between a loop and the rest of the chain may be accounted for. Although we treat these interactions only in an approximate way, we are able to give insight into the unbinding mechanism and on the nature of the transition.

2.2 Excluded-volume effects

To account for the excluded volume interactions between a loop and the rest of the chain we note that a microscopic configuration of the DNA molecule is composed of many bound and unbound segments of various length. In evaluating the number of available configurations of a loop, one has to take into account the interactions with all these bound and unbound segments. Here we simplify the problem and neglect the internal structure of the rest of the chain. We thus consider a loop embedded in a flexible chain (see Fig. 6) and study the number of configurations of a chain endowed with this topology, assuming that it is self avoiding. We will show that in the limit where the loop length, 2ℓ, is much smaller than the length of the rest of the chain, $2L$, the statistical weight of this topology can be written as a product of the statistical weight of the loop with that of the *rest of the chain*. The weight of the loop is found to be of the same form as that of a free loop but with a different exponent c. This exponent

Fig. 6. The topology of the loop embedded in a chain. The length of the chain from a vertex of type $V1$ to the nearest vertex of type $V3$ is L. The length of each of the two strands connected to the $V3$ vertices is ℓ.

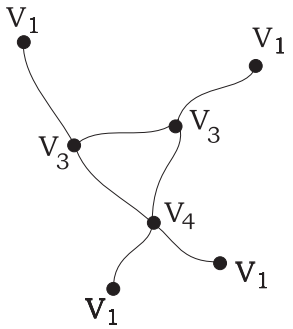

Fig. 7. An example of a polymer network. The network has four vertices V_1 of order 1, two vertices V_3 of order 3 and one vertex V_4 of order 4. It also has one loop. Thus in equation (13) $n_1 = 4$, $n_3 = 2$, $n_4 = 1$ and $\mathcal{L} = 1$.

is found to be larger than 2 in dimensions 2 and above, yielding a first order denaturation transition.

To carry out this analysis we use results obtained by Duplantier *et al.* [17,18] for the number of configurations of polymer networks of arbitrary topology. In order to make the paper self contained we first review in some detail these results. This represents an extension of the well known results for the number of configurations of a simple self-avoiding random walk [28]. In that case it is known that the number of configurations scales as

$$\Gamma_{\rm linear} \sim s^L L^{\gamma-1}, \tag{11}$$

where L is the length of the polymer, s is a non-universal geometrical constant and γ is a universal exponent. The exponent γ is known exactly in $d = 2$, numerically in $d = 3$ and via an ε expansion in $d = 4 - \varepsilon$. Above $d = 4$ self-avoiding interactions becomes irrelevant and thus the number of configurations of self-avoiding random walks scales as that of ordinary random walks, yielding $\gamma = 1$.

The generalization of this result to an arbitrary polymer network goes as follows: Consider a branched self-avoiding polymer $\mathcal{G}$ of arbitrary topology (see for example Fig. 7). The polymer is made of N chains of lengths $\ell_1, \ell_2, \ldots \ell_N$. These are tied together at vertices with different number of legs. A vertex with k legs is said to be of order k ($k \geq 1$). The number of vertices of order k is denoted by n_k. For large ℓ_i the number of configurations of the network, $\Gamma_{\mathcal{G}}$, is then given by

$$\Gamma_{\mathcal{G}} \sim s^L L^{\gamma_{\mathcal{G}}-1} g\left(\frac{\ell_1}{L}, \frac{\ell_2}{L}, \ldots, \frac{\ell_N}{L}\right), \tag{12}$$

where $L = \sum_i \ell_i$ is the total length of the network and g is a scaling function. The function g is smooth when its arguments are finite and may be singular when at least one of its arguments approaches zero (which amounts to

a crossover to a different topology). We note that the relation is valid also when the persistence length of each of the chains composing the graph is different. This variance can be absorbed through a rescaling of the chains lengths ℓ_i. Thus, when the thermodynamic limit is taken such that all length ℓ_i scale in the same way, the number of configurations is given by $\Gamma_{\mathcal{G}} \sim s^L L^{\gamma_{\mathcal{G}}-1}$. The exponent $\gamma_{\mathcal{G}}$ depends only on the topology of the network and is given by

$$\gamma_{\mathcal{G}} = 1 - \nu d\mathcal{L} + \sum_{k\geq 1} n_k \sigma_k. \tag{13}$$

Here $\mathcal{L}$ is the number of independent loops in the network, d the spatial dimension and ν is the exponent related to the radius of gyration of a self-avoiding random walk. Since vertices of order k appear n_k times we obtain the term $n_k\sigma_k$. The scaling dimensions σ_k, defined for $k \geq 1$, are known exactly in $d = 2$ from conformal invariance:

$$\sigma_k = (2-k)(9k+2)/64, \tag{14}$$

and to order ε^2 in $d = 4 - \varepsilon$:

$$\sigma_k = (\varepsilon/8)(2-k)k/2 + (\varepsilon/8)^2 k(k-2)(8k-21)/8 + O(\varepsilon^3). \tag{15}$$

Also, good estimates for the values of the exponents in $d = 3$ are available through Padé and Padé-Borel approximants. Clearly $\sigma_2 = 0$ as one would expect. Above $d = 4$, where the self-avoiding interaction is irrelevant, all the exponents σ_k are zero.

Consider now the topology depicted in Figure 6. We are interested in finding the number of configurations of the network in the limit $\ell \ll L$, when the loop size is much smaller than the length of the *rest of the chain*. Using the results by Duplantier (see Eq. (12)), the number of configurations can be written as

$$\Gamma \sim s^{L+\ell}(L+\ell)^{\gamma_{\rm loop}-1} g(\ell/L), \tag{16}$$

for large L and ℓ. Here $g(x)$ is a scaling function and $\gamma_{\rm loop}$ can be evaluated using equation (13). For the topology considered above of a loop embedded in two segments (Fig. 6) we have: one loop, $\mathcal{L} = 1$; two vertices of order one, $n_1 = 2$, corresponding to the two free ends of the chain (denoted by $V1$ in the figure); and two vertices of order three, $n_3 = 2$ (denoted in the figure by $V3$). Using equation (13) we obtain

$$\gamma_{\rm loop} = 1 - d\nu + 2\sigma_1 + 2\sigma_3. \tag{17}$$

The limit of interest is that of a loop size much smaller than the length of the chain, $\ell/L \ll 1$. Clearly, in the limit $\ell/L \to 0$, the number of configurations should reduce to that of a single self-avoiding open chain, which, to leading order in L, is given by $s^L L^{\gamma-1}$, where $\gamma = 1 + 2\sigma_1$. This implies that in the limit $x \ll 1$

$$g(x) \sim x^{\gamma_{\rm loop}-\gamma}. \tag{18}$$

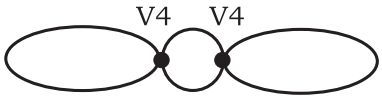

Fig. 8. An extreme topology where the loop of length $2l$ is embedded in two denaturated loops of size $2L$ each. The vertices of order 4 are denoted by $V4$.

Thus the number of configurations is given by

$$\Gamma \sim s^{\ell}\ell^{\gamma_{\rm loop}-\gamma} s^L L^{\gamma-1}. \tag{19}$$

It is therefore evident that, for large ℓ and L and in the limit $\ell/L \ll 1$, the partition sum is decomposed into a product of the partition sums of the loop and that of the rest of the chain. The excluded volume interaction between the loop and the rest of the chain is reflected in the value of the effective exponent c. This result is very helpful since it enables one to extend the Poland-Scheraga approach described in the previous section to the case of interacting loops. From equation (19) one sees that the appropriate effective exponent c is given by

$$c = \gamma - \gamma_{\rm loop} = d\nu - 2\sigma_3. \tag{20}$$

In $d = 2$, $\sigma_3 = -29/64$ [17] and $\nu = 3/4$, yielding

$$c = 2 + 13/32\,. \tag{21}$$

In $d = 4-\varepsilon$ to $O(\varepsilon^2)$, one has $\sigma_3 = -3\varepsilon/16 + 9\varepsilon^2/512$ and $\nu = 1/2(1 + \varepsilon/8 + 15/4(\varepsilon/8)^2)$, yielding

$$c = 2 + \varepsilon/8 + 5\varepsilon^2/256\,. \tag{22}$$

In $d = 3$, one may use Padé and Padé-Borel approximations to obtain $\sigma_3 \approx -0.175$ [18] which with the value $\nu \approx 0.588$ [18] yields

$$c \approx 2.115. \tag{23}$$

The value of the exponent c is unaffected by the different persistence length of a bound and unbound DNA segment. This is since as stated above equation (12) is valid also when the persistence length of different polymers composing the network are different. Equation (20) can be understood intuitively by remarking that taking the limit $\ell/L \to 0$ corresponds to shrinking the loop. By doing so one loop and two vertices of order 3 are eliminated. The exponent c is the difference between the exponent $\gamma_{\mathcal{G}}$ of the network after shrinking the loop and the same exponent before the loop has shrunk. Therefore c gets a contribution of $d\nu$ from the eliminated loop and $-2\sigma_3$ from the two vertices of order 3.

As stated above the rest of the chain is in fact composed of both bound and unbound segments. This structure has been neglected in the above analysis. To estimate the effect of the denaturated segments we consider the extreme case in which the rest of the chain is fully denaturated. That is, a loop embedded within two large loops each of size $2L$ (see Fig. 8). An analysis similar to the one

presented above yields for the value of c,

$$c = d\nu - \sigma_4 \,, \tag{24}$$
$$= 2 + 11/16 \,, \qquad \text{in } d = 2 \,, \tag{25}$$
$$= 2 + \varepsilon/4 - 15\varepsilon^2/128 \,, \qquad \text{in } d = 4 - \varepsilon, \tag{26}$$

where the values $\sigma_4 = -19/16$ in $d = 2$ and $\sigma_4 = -\varepsilon/2 + 11(\varepsilon^2/8)^2$ in $d = 4 - \varepsilon$ dimensions [17] are used along with those of ν. Using $\sigma_4 \approx -0.46$ obtained by Padé and Padé-Borel approximations gives in $d = 3$ the value $c \approx 2.22$. Therefore, the effect of the extra excluded volume interaction is to increase the value of c. It is easy to check that the exponent governing the ℓ-dependence remains unchanged even if the end points of one or both the outer loops are set to be unbound.

In both topologies considered above the value of the exponent c is larger than 2 in $d = 2$, $d = 4 - \varepsilon$ and $d = 3$. This strongly suggests that the transition is *first order* for any $d \geq 2$. Our analysis assumes that the size of the loops in the system is much smaller than the total chain length. The fact that the transition is first order implies that the loops size remain finite as the transition is approached. This makes the analysis self-consistent. Note that the loop size distribution, $P(\ell)$, is rather broad at the transition and behaves, for large ℓ, as

$$P(\ell) \sim \frac{1}{\ell^c} \,. \tag{27}$$

Thus high enough moments of the loop size distribution always diverge. Although the transition is first order for $c > 2$ it exhibits some critical properties. In particular since c is found to be between 2 and 3 already the variance of the loop size is predicted to diverge. A recent numerical study of the loop size distribution, where the excluded volume interactions have been fully taken into account, has verified this prediction with a measured value of $c = 2.10 \pm 0.02$ [20].

Consider now a configuration where "the rest of the chain" is composed of a particular sequence of bound segments and open loops. We note that for any such sequence, the effective exponent c associated with a loop is either given by (20) or by (24). In particular if the two bound segments connected to the loop under consideration are long, the value of the exponent c remains the same as that given by (20) (see Fig. 6). This is due to the fact that shrinking the loop eliminates one loop and two vertices of order 3. This result is also easy to verify using a scaling argument similar to the one presented above. The value of the exponent c is altered to (24) (see Fig. 8) when at least one of the segments connected to the loop is short. This may be verified by noticing that as the loop is shrunk a vertex of order 4 and one loop are lost. Note that this does not imply that these values of c can be used to give bounds on the true partition function of the model.

The analysis presented in this section is valid in the thermodynamic limit. For any finite chain the melting transition is clearly expected to be rounded. Finite size scaling of the model performed in [15] indicate that the width of the transition decreases as $1/L$, where L is the length of the chain. Numerical studies [15,20] of finite chains where self-avoiding interactions are fully taken into account seem to be consistent with this scaling behavior.

2.3 The end segment distribution

In the previous section it was argued that the excluded volume interactions between a denaturated loop and the rest of the chain cause the transition to be first order. As was noted this result is unaffected by the boundary terms $V_0(z)$ and $Q(z)$ (see Eq. (3)). However, when the DNA molecule is fully denaturated, its statistical properties are determined by the boundary term $Q(z)$. This is evident as the entropy of a denaturated molecule with free ends is much higher than that of one with the ends constraint to meet each other. Namely, the number of configurations of a self-avoiding random walk is much higher than that of a self-avoiding random walk which returns to the origin. Thus, one expects the average length of the end segment to diverge at the transition. This does not affect our previous results as long as the thermodynamic limit, $L \to \infty$, is taken before the temperature T approaches the melting temperature $T_{\rm M}$. In this case the size of the end segment is always much smaller than the rest of the chain.

The average length of the end segment, namely the number of unbound monomers at the chain end, is given by

$$\xi = z \frac{\partial \ln Q}{\partial z}\bigg|_{z=z^*} \,. \tag{28}$$

where z^* is evaluated using equation (7). The behavior of ξ near the transition takes one of three forms depending on the numerical value of $\bar{c}$:
(i) $\bar{c} \leq 1$. In this case $Q(sz) = 1 + \Phi_{\bar{c}}(sz)$ diverges like $|z - z_{\rm M}|^{\bar{c}-1}$ (see Appendix A) as $z \to z_{\rm M} = 1/s$. Clearly its derivative with respect to z diverges like $|z - z_{\rm M}|^{\bar{c}-2}$ and hence the average length of the end segment diverges like $\xi \sim |z^* - z_{\rm M}|^{-1}$.
(ii) $1 < \bar{c} \leq 2$. Here $Q(sz)$ is finite but its derivative diverges like $|z - z_{\rm M}|^{\bar{c}-2}$ as $z \to z_{\rm M} = 1/s$. Thus, the average length of the end segment diverges like $\xi \sim |z^* - z_{\rm M}|^{\bar{c}-2}$.
(iii) $\bar{c} > 2$. In this case both $Q(sz)$ and its derivative are finite at the transition. Hence the end segment length is finite at the transition.

We next use the temperature dependence of $z^*(w)$ in order to evaluate the behavior of the end segment length ξ as a function of the temperature near the transition. Here one finds two regimes depending on the value of the exponent c related to the entropy of the bulk loops [13]. To derive these relations we note that for z^* close to $z_{\rm M}$ equation (A3) yields for $c > 1$,

$$U(z_{\rm M}) - U(z^*) \sim |z_{\rm M} - z^*|^{\zeta}, \tag{29}$$

where $\zeta = \min(1, c - 1)$. Using $w(z) = [z(1 + U(z))]^{-1}$ from equation (8) we obtain

$$U(z_{\rm M}) - U(z^*) = \frac{1}{w_{\rm M} z_{\rm M}} - \frac{1}{w(z^*) z^*}, \tag{30}$$

where $w_{\rm M} = w(z_{\rm M})$. Rewriting this expression yields

$$|z_{\rm M} - z^*|^\zeta \sim a w_{\rm M}(z^* - z_{\rm M}) + a(w(z^*) - w_{\rm M})z^*, \quad (31)$$

where $a = 1/(w_{\rm M} z_{\rm M} w(z^*) z^*)$ is finite at the transition. Clearly $w(z^*) - w_{\rm M} \sim T_{\rm M} - T$ near the transition, where $T_{\rm M}$ is the melting temperature and T is the temperature corresponding to $w(z^*)$. Thus, equation (31) yields two regimes for the temperature dependence of z^*. For $1 < c \leq 2$ the linear term in $(z^* - z_{\rm M})$ in equation (31) is negligibly small compared to the left hand side and hence

$$z_{\rm M} - z^* \sim |T - T_{\rm M}|^{1/(c-1)}. \quad (32)$$

On the other hand for $c > 2$ $\zeta = 1$ and therefore

$$z_{\rm M} - z^* \sim |T - T_{\rm M}|. \quad (33)$$

Thus, the behavior of the end segment length ξ depends on both exponents c and $\bar{c}$. We have already shown that $c > 2$. We now turn to estimate the exponent $\bar{c}$. This exponent is associated with the degeneracy, $s^\ell/\ell^{\bar{c}}$, of an end segment of length ℓ, and its value may be deduced using the scaling argument presented above. One considers a Y-fork topology where the two denaturated strands are much smaller than the bound segment representing the rest of the chain. Using the terminology introduced in the preceding subsection one has to consider the difference of the exponent corresponding to the Y-fork topology and the one corresponding to linear chain. This yields

$$\bar{c} = -(\sigma_1 + \sigma_3). \quad (34)$$

In $d = 2$, $\sigma_1 = 11/64$ and $\sigma_3 = -29/64$, so that $\bar{c} = 9/32$. In $d = 3$, one may use the Padé and Padé-Borel approximation results $\sigma_1 \simeq 0.083$ and $\sigma_3 \simeq -0.175$ to obtain $\bar{c} \simeq 0.092$. Finally in $d = 4 - \varepsilon$, $\bar{c} = \varepsilon/8 + O(\varepsilon^2)$ while above four dimensions clearly $\bar{c} = 0$. This suggests that the exponent $\bar{c}$ is smaller than one for any $d \geq 2$.

Since the degeneracy exponents satisfy $c > 2$ and $\bar{c} < 1$ the analysis presented above suggests that near the melting transition the end segment length diverges like

$$\xi \sim \frac{1}{|T - T_{\rm M}|}\,. \quad (35)$$

This result is in agreement with numerical simulations [15] which measure the end to end distance which is expected to behave like ξ^ν.

2.4 Summary and overview of thermal denaturation

In the preceding subsections we reviewed the analysis of Poland and Scheraga who showed that the nature of the denaturation transition is governed by the exponent c. It has been found [12] that for $c \leq 1$ there is no transition, for $1 < c \leq 2$ there is a continuous transition, while for $c > 2$ the transition is first order. Using recent results for the number of configurations of a self avoiding polymer network of general topology we have demonstrated that $c > 2$ in $d = 2$ dimensions and above. While the treatment is approximate the results strongly suggest that the transition is indeed first order in these dimensions. The results are unchanged even if the different persistence length of the bound and denaturated configurations is taken into account.

The analysis of the length distribution of the end segment shows that the molecule melts from the unbound end. The boundary condition used in this analysis is such that the two strands are bound at one end. It is easy to show that when the boundary condition is modified such that the two strands are bound at the center rather than at one end, melting takes place from both ends.

It is interesting to consider the case of a homopolymer where all the bases on one strand are the same. For example a molecule in which one strand is made only of G bases and the other strand is made only of C bases. Here each base, and therefore a monomer in the model, can bind to any other base on the other strands and not just to a specific monomer, as is the case for a heteropolymer. The two arms of a loop need not be of the same length. Thus the number of configurations of a loop of length ℓ changes by a factor of ℓ. This amounts to using $\ell s^\ell/\ell^c$ for the weight of a loop, effectively reducing the exponent c to $c - 1$. Since we have shown that $2 < c < 3$ for $d \geq 2$, this implies for a homopolymer the effective c is smaller than 2 but greater than 1. The transition in this case is thus expected to be *continuous*.

Finally, we comment on a recent attempt [7] to account for the first-order nature of the denaturation transition without having to resort to excluded volume interactions. Within this approach the two strands are considered as directed polymers and thus they do not self intersect. Let $V(\boldsymbol{r})$ be the interaction potential between the corresponding monomers on the two strands. It is repulsive at short distances, has an attractive well at the characteristic pair bond distance and it tends to zero at large distances. Using a transfer matrix approach, the thermodynamic properties of the chain are obtained by the ground state $\psi_0(\boldsymbol{r})$ wavefunction of a Schrödinger equation which takes the form [28]

$$-\frac{1}{2m(\boldsymbol{r})}\nabla^2\psi_0(\boldsymbol{r}) + V(\boldsymbol{r})\psi_0(\boldsymbol{r}) = \epsilon_0\psi_0(\boldsymbol{r}). \quad (36)$$

Here ϵ_0 is the ground state eigenvalue associated with the wave function $\psi_0(\boldsymbol{r})$. The mass m represents the stiffness of the chain. The $\boldsymbol{r}$-dependence of the mass is taken to account for the change of stiffness between bound and unbound strands [7]. One usually assumes that the potential $V(\boldsymbol{r})$ is short ranged (exponentially decaying with $\boldsymbol{r}$ at large distances) and that $m(\boldsymbol{r})$ varies with $\boldsymbol{r}$ over the same distance. The probability density of finding the strands a distance $\boldsymbol{r}$ from each other is given by $|\psi_0(\boldsymbol{r})|^2$. In this language the order parameter θ is given by

$$\theta = \frac{\int_0^a |\psi_0(r)|^2 {\rm d}^d r}{\int_0^\infty |\psi_0(r)|^2 {\rm d}^d r}, \quad (37)$$

where a is the range of the binding potential $V(\boldsymbol{r})$. Here, for simplicity, we assume that neither $m(\boldsymbol{r})$ nor $V(\boldsymbol{r})$ depend on the orientation.

The denaturation transition occurs when ϵ_0 reaches zero so that there is no longer a bound state in the system. A first order transition occurs if θ has a non-zero value at that point. In this case one should have a bound state with energy $\epsilon_0 = 0$. Note that although θ might jump at the transition, the average distance between the strands, which is not the order parameter of the system, might diverge continuously. Thus, we are interested in the solution of the equation

$$-\nabla^2\psi_0(\boldsymbol{r}) + 2m(r)V(r)\psi_0(\boldsymbol{r}) = 0. \tag{38}$$

Since $V(r)$ decays exponentially with r, the asymptotic behavior of $\psi_0(\boldsymbol{r})$ at large distances is obtained by the equation

$$\nabla^2\psi_0(\boldsymbol{r}) = 0. \tag{39}$$

It is easy to show that this implies $\psi_0(\boldsymbol{r}) \sim 1/r^{d-2}$ for large r. That is, the wave function for large r is normalizable only for $d > 4$ so one has a localized zero energy ground state. Therefore using equation (37) one can see that for any dimension $d < 4$ the order parameter θ vanishes at the transition as is expected for a continuous transition. A first-order phase transition occurs only for $d > 4$. Note that this result is equivalent to using the Poland-Scheraga model with random walks modeling the loop entropy. Within this approach the transition may thus be altered to first order in lower dimensions *only* when $V(r)$ are long-range, in contrast to recent claims [7].

Recently, it has been argued that taking into account excluded-volume interactions *between* the two strands while neglecting these interactions within each strand effectively leads to a long range potential between the two strands [19]. This model, which takes into account the excluded volume interactions only partially also leads to a first order transition.

3 The unzipping transition

The introduction of new and powerful techniques, such as optical tweezers [9] and atomic force microscopes [10], has made possible the manipulation of single biological macromolecules. A number of experiments have investigated the response of double-stranded DNA to external forces and torques [29]. Recently, it has become possible to apply and measure a force pulling apart two strands of a DNA double helix [11]. Previous theoretical studies of the unzipping transition have been carried out using the directed polymer approach where self-avoiding interactions are not accounted for [21–25]. In most of these studies the unzipping of homopolymers has been analyzed. Heterogeneous chains have also been considered in some of these studies [22]. In this section we extend the analysis of the PS model to consider the unzipping of homopolymers with self-avoiding interactions. The treatment assumes that an equilibrium description of the transition is adequate. We show that the unzipping transition is first order. We also calculate the dependence of the critical unzipping force on the temperature at low forces, namely near the melting temperature.

We consider a configuration where the corresponding monomers at one end of the chain are bound together, while a force $\boldsymbol{f}$ is applied on the two monomers at the other end of the chain, pulling the two strands apart. In this setup, the grand canonical partition function takes the form

$$\mathcal{Z} = \frac{V_0(z)O(z)}{1 - U(z)V(z)}, \tag{40}$$

where the factor $O(z)$ is the grand partition function of the open tail under force. We have

$$O(z) = 1 + \sum_{\ell=1}^{\infty} \mathcal{Z}_{\rm end}(\ell) z^{\ell}, \tag{41}$$

where $\mathcal{Z}_{\rm end}(\ell)$ is the canonical partition function of an open end composed of two strands, each of length ℓ.

To evaluate $\mathcal{Z}_{\rm end}(\ell)$ we note that when no force is applied the partition sum takes the form

$$\mathcal{Z}_{\rm end}(\ell) = \Lambda(2\ell) \sim \frac{s^{\ell}}{\ell^{\bar{c}}}, \tag{42}$$

where, as discussed in Section 2.3, $\bar{c}$ is given by equation (34). When a force $\boldsymbol{f}$ is applied, we have

$$\mathcal{Z}_{\rm end}(\ell) = \Lambda(2\ell) \int \mathrm{d}\boldsymbol{r}\, p_{\ell}(r)\, \exp(\boldsymbol{f}\cdot\boldsymbol{r}/T), \tag{43}$$

where $p_{\ell}(r)$ is the probability distribution of the end-to-end distance in the absence of a force. Turning to angular coordinates we obtain

$$\mathcal{Z}_{\rm end}(\ell) = \Lambda(2\ell)\, \mathcal{I}_{\ell}(f/T), \tag{44}$$

where

$$\mathcal{I}_{\ell}(f/T) = S \int_0^{\pi} \mathrm{d}\phi\, \sin^{d-2}\phi \int_0^{\infty} \mathrm{d}r\, r^{d-1} \times p_{\ell}(r)\, \exp(fr\cos\phi/T), \tag{45}$$

in which S is a constant which depends on dimensionality. We assume that $p_{\ell}(r)$ has the same scaling form as that of linear polymers

$$p_{\ell}(r) = R^{-d}\hat{p}(r/R). \tag{46}$$

Here R is a scaling length related to ℓ by

$$R \simeq R_0\, \ell^{\nu}, \tag{47}$$

where ν is the correlation length exponent of a linear polymer.

We are interested in the behavior of $\hat{p}(x)$ at $x \gg 1$ (see below). We assume that in this limit $\hat{p}(x)$ takes a form similar to that corresponding to a linear polymer [30]

$$\hat{p}(x) = P\, x^{\mu}\, \exp(-Dx^{\lambda}). \tag{48}$$

Here P and D are constants, $\lambda = 1/(1-\nu)$ and the exponent μ is given by

$$\mu = (d/2 + \nu d - \bar{c})/(1-\nu). \tag{49}$$

This result can be obtained by applying the same reasoning used to derive the corresponding expression for linear polymers to the Y-fork configurations which are of interest here [30,31]. Substituting into expression (45) we obtain

$$\mathcal{I}_\ell(f/T) \propto \int_0^\pi \mathrm{d}\phi \, \sin^{d-2}\phi \int_0^\infty \mathrm{d}x \, x^{d-1+\mu} \times \exp\left(-Dx^{1/\lambda} + ux\cos\phi\right), \tag{50}$$

where $u = fR/T$. This expression is valid provided the integral is dominated by large values of x, which is the case for $u \gg 1$. This appears to be the relevant regime for piconewton forces. In this situation we can evaluate the integral by steepest descent and obtain the saddle-point equations

$$\phi^* = 0; \tag{51}$$

$$x^* = \left(\frac{\lambda u}{D}\right)^{1/(\lambda-1)}. \tag{52}$$

They correspond to the non-Hookean elongation regime, described by de Gennes [28, p. 47ff]. Note, that x^* scales as $u^{1/(\lambda-1)}$ and that $\lambda > 1$. Since we are interested in the limit of $u \gg 1$ this justifies our choice of using the tail of the distribution $\hat{p}(x)$. Therefore

$$\mathcal{I}_\ell(f/T) \propto \ell^{\mu(1-\nu)+\frac{d}{2}(1-2\nu)} \exp\left(A(fR_0/T)^{1/\nu}\,\ell\right), \tag{53}$$

where A is a constant. The elongation ($\rho = \langle r\cos\phi\rangle$)-force ($f$) curve of a polymer of length ℓ is readily deduced from this expression. Using equation (50) one notes that $\rho \sim \partial \ln \mathcal{I}/\partial u$ which yields $\rho \sim \ell f^{1/\nu - 1}$ as observed in [28]. Using (53) we have

$$O(z) \simeq 1 + \sum_{\ell=1}^{\infty} \frac{\left[zs\exp\left(A(fR_0/T)^{1/\nu}\right)\right]^\ell}{\ell^{\bar{\mu}}}, \tag{54}$$

where the exponent $\bar{\mu}$ is given by

$$\bar{\mu} = -\bar{c} - \mu(1-\nu) - \frac{d}{2}(1-2\nu). \tag{55}$$

Substituting μ from equation (49) into this expression we obtain $\bar{\mu} = 0$.

According to equation (54) at temperatures below the melting temperature T_M the end segment partition sum $O(z)$ diverges at a critical, unzipping force f_U, given by

$$\mathrm{e}^{-A(f_\mathrm{U}R_0/T)^{1/\nu}} = sz^*(w). \tag{56}$$

Here $z^*(w)$ is the solution of equation (7) (corresponding to an infinitely long polymer). At this point the average length of a loop in the bulk is finite. Hence the unzipping transition is first order.

Near the transition, the length ξ of the end segment diverges like $|z^* - z_\mathrm{U}|^{-1}$, where $z_\mathrm{U} = \mathrm{e}^{-\kappa(f_\mathrm{U}/T)^{1/\nu}}/s$. This is a result of the fact that the exponent $\bar{\mu}$ is smaller than 1 (in fact it vanishes, as we have seen). Since z^* is regular in f, we have $\xi \sim |f - f_\mathrm{U}|^{-1}$ or $\xi \sim |T - T_\mathrm{U}(f)|^{-1}$. Thus the two strands separate gradually from the end as the critical force is approached. Nonetheless the unzipping transition is first order. The reason is that the transition takes place at a temperature below the denaturation melting temperature T_M where the loop size distribution in the interior of the chain decays exponentially with the loop size. Thus at this point the average loop size in the interior of the chain is finite. On the other hand the length of the end segment is finite as long as $f < f_\mathrm{U}$ and its contribution to the order parameter θ and to the entropy is negligible. Therefore both the order parameter and the entropy exhibit a discontinuity to their values in the unzipped state at the transition. Let us remark that equation (56) implies that the critical force f_U behaves like

$$f_\mathrm{U} \sim |T - T_\mathrm{M}|^\nu \tag{57}$$

as $T \to T_\mathrm{M}$, at least as long as the forces are not too small (so that the $u \gg 1$ limit in (50) is valid).

4 Summary

In this paper we have extended the Poland-Scheraga type models introduced in the early sixties to take into account the effect of self-avoiding interactions on the DNA denaturation transition. We have shown that the model yields a first-order transition consistent with experiments and in agreement with numerical simulations. Although the transition is thermodynamically first order it exhibits critical behavior in some of its properties, such as its loop size distribution and the length of the end segment. It would be of interest to study these properties experimentally to test these predictions.

We have also studied the unzipping transition of DNA and found it to be first order. We have evaluated the behavior of the unzipping force near the melting point.

We are indebted to M.J.E. Richardson and H. Orland for many inspiring discussions and their involvement in the early stages of our study of the DNA denaturation transition. We also thank discussion and suggestions by B. Duplantier. The work of LP was partially supported by a Michael Visiting Professorship of the Weizmann Institute and has been performed within a joint cooperation agreement between Japan Science and Technology Corporation (JST) and Università di Napoli "Federico II".

Appendix A: Properties of the polylog function

We summarize here a few elementary properties of the polylog function that are used in the text [32, Sec. 1.11, p. 27ff.]. The polylog function $\Phi_c(z)$ is defined by the series

$$\Phi_c(z) = \sum_{\ell=1}^{\infty} \frac{z^\ell}{\ell^c}, \tag{A1}$$

which converges for $|z| < 1$. For $|z| < 1$ and $\Re\mathrm{e}\ c > 0$ it has the integral representation

$$\Phi_c(z) = \frac{1}{\Gamma(c)} \int_0^\infty \mathrm{d}t\, t^{c-1} \frac{z\mathrm{e}^{-t}}{1 - z\mathrm{e}^{-t}}, \tag{A2}$$

where $\Gamma(c)$ is Euler's gamma function. From equation (A2) it is easy to see that $\Phi_c(z)$ diverges like $|z-1|^{c-1}$ for $z \to 1$, if $c \leq 1$, and that, if $c > 1$ and $1 - z = \epsilon \ll 1$, one has

$$\Phi_c(1) - \Phi_c(1-\epsilon) \sim \epsilon^\zeta, \tag{A3}$$

where the exponent ζ is equal to $\min(1, c-1)$. From the series definition (A1) it is also evident that

$$z\frac{d\Phi_c(z)}{dz} = \Phi_{c-1}(z). \tag{A4}$$

Appendix B: High-temperature phase

Above the transition, *i.e.*, for $w < w_\mathrm{M}$, the fugacity z becomes $z_\mathrm{M} = 1/s$ in the thermodynamic limit. In this Appendix we discuss in some detail how the thermodynamic limit is taken.

The value of the fugacity for a chain of length L is obtained by solving equation (6) for the fugacity z. It reads

$$\langle L\rangle = z\frac{V_0'(z)}{V_0(z)} + z\frac{Q'(z)}{Q(z)} + z\frac{U'(z)V(z) + U(z)V'(z)}{1 - U(z)V(z)}\,. \tag{B1}$$

For $w < w_\mathrm{M}$, as z approaches z_M, the first term is regular, while the second diverges as $|z - z_\mathrm{M}|^{-1}$. In the third term the denominator does not vanish. The term may or may not diverge depending on whether $U'(z)$ diverges at z_M, namely, according to whether c is smaller or larger than 2. However, in any case this term is much smaller than the second. We therefore have

$$|z - z_\mathrm{M}| \sim L^{-1}. \tag{B2}$$

Since the second term in equation (B1) is equal to the number of units in the end segment, and since this is the dominant term we conclude that almost all units belong to it in the thermodynamic limit.

If both ends of the DNA chain are constrained to be bound, the partition function assumes the form

$$\mathcal{Z} = \frac{V_0^2(z)}{1 - U(z)V(z)}\,. \tag{B3}$$

Therefore the most singular term in equation (B1) is absent. If $c \leq 2$, $U'(z)$ diverges as $|z - z_\mathrm{M}|^{c-2}$, so that the equation corresponding to (B1) can still be solved for any finite L. This means that the typical size of the denaturated loops increases with increasing L.

In the case $c > 2$, on the other hand, the derivative of $\ln\mathcal{Z}$ with respect to $\ln z$ remains finite even for $z = z_\mathrm{M}$. The chain thus seems to have a maximum length, $L_0(w)$, corresponding to $z = z_\mathrm{M}$. Therefore, for longer chains one has to consider the total length constraint more explicitly. To do that we introduce a maximal length cutoff, $\bar{L}$, in the sums (3) and (4) defining $U(z)$ and $V(z)$, respectively. With this cutoff the number of terms in each of these series is finite, and thus they may be evaluated even for $zs > 1$. In this case the last term in equation (B1) diverges when $U(z)V(z) - 1$ vanishes. Therefore, for any $\langle L\rangle$, and particularly for $\langle L\rangle > L_0(w)$, one can find a fugacity which satisfies $1 - U(z)V(z) \sim 1/\langle L\rangle$. Since for large $\bar{L}$, $U(z)$ grows exponentially as $(sz)^{\bar{L}}$ this implies, to leading order in $\bar{L}$,

$$(sz)^{\bar{L}} \sim \left(1 - \frac{a}{\langle L\rangle}\right)\frac{1}{V}, \tag{B4}$$

where a is a constant. For large $\bar{L}$ the right hand side of this equation approaches a constant, independent of $\bar{L}$. This is due to the fact that for z larger but sufficiently close to z_M and at temperatures above the melting temperature $V(z)$ is finite even for $\bar{L}$ going to infinity. In the limit $\bar{L} \to \infty$ the fugacity behaves as

$$z = \frac{1}{s} + O\left(\frac{1}{\bar{L}}\right). \tag{B5}$$

Thus, in this case, z approaches its limiting value from above.

References

1. For a review see R.M. Wartell, A.S. Benight, Phys. Rep. **126**, 67 (1985); O. Gotoh, Adv. Biophys. **16**, 1 (1983).
2. For a review see *Theory of Helix-Coil Transitions in Biopolymers*, edited by D. Poland, H.A. Scheraga (Academic, New York, 1970); F.W. Wiegel, in *Phase Transitions and Critical Phenomena*, edited by C. Domb, J.L. Lebowitz, Vol. 7 (Academic, New York, 1983) p. 101.
3. T.L. Hill, J. Chem. Phys. **30**, 383 (1959); B.H. Zimm, J. Chem. Phys. **33**, 1349 (1960); S. Lifson, J. Chem. Phys. **40**, 3705 (1964); M. Ya. Azbel, Phys. Rev. A **20**, 1671 (1979).
4. M. Peyrard, A.R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989).
5. D. Cule, T. Hwa, Phys. Rev. Lett. **79**, 2375 (1997).
6. S. Cocco, R. Monasson, Phys. Rev. Lett. **84**, 5178 (1999).
7. N. Theodorakopoulos, T. Dauxois, M. Peyrard, Phys. Rev. Lett. **85**, 6 (2000).
8. V.M. Pavlov, J.L. Lyubchenko, A.S. Borovik, Y. Lazurkin, Nucl. Acids. Res. **4**, 4052 (1977); A.S. Borovik, Y.A. Kalambet, Y.L. Lyubchenko, V.T. Shitov, E. Golovanov, Nucl. Acids. Res. **8**, 4165 (1980).

9. K. Svoboda, S.M. Block, Ann. Rev. Biophys. Biomol. Structure **23**, 247 (1994); A. Ashkin, Proc. Natl. Acad. Sci. USA **94**, 4853 (1997).
10. H.G. Hansma, J. Vac. Sci. Technol. B **14**, 1390 (1995).
11. U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, Phys. Rev. Lett. **79**, 4489 (1997); B. Essevaz-Roulet, U. Bockelmann, F. Heslot, Proc. Natl. Acad. Sci. USA **94**, 11935 (1997).
12. D. Poland, H.A. Scheraga, J. Chem. Phys. **45**, 1456 (1966); J. Chem. Phys. **45**, 1464 (1966).
13. M.E. Fisher, J. Chem. Phys. **45**, 1469 (1966).
14. In this estimate of c all configurations, including the knotted ones are taken into account. However, it has been shown that in $d = 3$ the number of *unknotted* configurations is given by a formula of the same form but with a slightly smaller s and with an exponent c which is practically unmodified. See, *e.g.*, A.L. Kholodenko, T.A. Vilgis, Phys. Rep. **298**, 251 (1998).
15. M.S. Causo, B. Coluzzi, P. Grassberger, Phys. Rev. E **62**, 3958 (2000).
16. Y. Kafri, D. Mukamel, L. Peliti, Phys. Rev. Lett. **85**, 4988 (2000).
17. B. Duplantier, Phys. Rev. Lett. **57**, 941 (1986); J. Stat. Phys. **54**, 581 (1989).
18. L. Schäfer, C. von Ferber, U. Lehr, B. Duplantier, Nucl. Phys. B **374**, 473 (1992).
19. T. Garel, C. Monthus, H. Orland, Europhys. Lett. **55**, 132 (2001).
20. E. Carlon, E. Orlandini, L. Stella, `cond-mat/0108308`.
21. S.M. Bhattacharjee, J. Phys. A **33**, L423 (2000); Erratum: **33**, 9003 (2000).
22. D.K. Lubensky, D.R. Nelson, Phys. Rev. Lett. **85**, 1572 (2000); D.K. Lubensky, D.R. Nelson, Phys. Rev. E **65**, 031917 (2002).
23. N. Hatano, D.R. Nelson, Phys. Rev. Lett. **77**, 570 (1996); N. Hatano, D.R. Nelson, Phys. Rev. B **56**, 8651 (1997).
24. D. Marenduzzo, A. Trovato, A. Maritan, Phys. Rev. E **64**, 031901 (2001); A. Maritan, E. Orlandini, F. Seno, unpublished.
25. S. Cocco, R. Monasson, J.F. Marko, Proc. Natl. Acad. Sci. USA **98**, 8608 (2001).
26. M.T. Record, S.J. Mazur, P. Melanon, J.H. Roe, S.L. Shaner, L. Unger, Ann. Rev. Biochem. **50**, 997 (1981).
27. V.A. Bloomfield, D.M. Crothers, I. Tinoco *Nucleic Acids - Structure, Properties and Function* (University, California, 2000).
28. P.-G. de Gennes, *Scaling Concepts in Polymer Physics*, (Cornell, Ithaca, 1979).
29. See, *e.g.*, B. Smith, L. Finzi, C. Bustamante, Science **258**, 1122 (1992); T.R. Strick, J.-F. Allemand, D. Bensimon, A. Bensimon, V. Croquette, Science **271**, 1835 (1996); J.-F. Léger, G. Romano, A. Sarkar, J. Robert, L. Bourdieu, D. Chatenay, J.F. Marko, Phys. Rev. Lett. **83**, 1066 (1999).
30. J. des Cloiseaux, G. Jannink, *Les polymères en solution: leur modélisation et leur structure* (Les Éditions de Physique, Les Ulis, 1987).
31. D.S. McKenzie, M.A. Moore, J. Phys. A **4**, L82 (1971).
32. A. Erdélyi, *Higher Transcendental Functions*, Vol. I, (McGraw-Hill, New York).