Complete Phase Diagram of DNA Unzipping: Eye, *Y* Fork, and Triple Point

Rajeev Kapri,[1,*] Somendra M. Bhattacharjee,[1,†] and Flavio Seno[2,‡]

[1]*Institute of Physics, Bhubaneswar 751 005, India*

[2]*INFM, Dipartimento di Fisica, Universitá di Padova, Via Marzolo 8, 35131 Padova, Italy*

(Received 31 March 2004; published 7 December 2004)

We study the unzipping of double stranded DNA by applying a pulling force at a fraction s ($0 \leq s \leq 1$) from the anchored end. From exact analytical and numerical results, the complete phase diagram is presented. The phase diagram shows a strong ensemble dependence for various values of s. In addition, we show the existence of an eye phase and a triple point.

DOI: 10.1103/PhysRevLett.93.248102 PACS numbers: 87.14.Gg, 36.20.Ey, 64.70.–p, 82.35.Pq

The initial step in DNA replication and RNA transcription, as the enzyme associates with DNA, is to open a few base pairs near it. In the case of replication this opening takes place near one of the ends, whereas for transcription it can be anywhere on the DNA [1–3]. The ubiquity of the process calls for a mechanism that does not require high temperatures or extreme pH conditions, unlike melting to which it generally gets associated. One such possibility is a force induced unzipping transition [4]. This transition has now been well established both theoretically [4–13] and experimentally [14–16]. The focus of attention so far has been the geometry reminiscent of the DNA replication, pulling only the open end of dsDNA. However, transcription requires pulling DNA at an intermediate point, often with DNA getting anchored to the cytoplasmic membrane [3] *in vivo* . Similarly, end constraint is important in circular DNA as, e.g., in bacteria like *E. coli.* Anchoring of one end is also used in single molecule experiments [14,15]. The richer surprises in this type of geometry provide the primary motivation for working out the full unzipping phase diagram.

The extensive theoretical work [4–13] on the unzipping transition in various avatars of the basic Poland-Scheraga model [17] and the nature of the real phase diagram [14] recently obtained for lambda phage DNA indicate that the basic features are preserved in simpler exactly solvable models [6,8] *even in two dimensions.* These basic results include the first order nature of the unzipping transition and the existence of a reentrant region allowing unzipping by decreasing temperature. In this Letter our aim is to study the force induced transition on the lattice model used previously in Ref. [8] but with the pulling force applied on a base pair which is N_1 monomers away from the anchored end of the DNA molecule (of total length N). We call s the fixed fraction N_1/N; see Fig. 1. In single molecule experiments, the results may depend on the statistical ensemble used [18,19]. One may recall that instruments like atomic force microscopes [15] use the fixed distance ensemble while the magnetic bead method of Ref [14] uses the fixed force ensemble. Therefore, we studied the unzipping transition both in the fixed force and the fixed distance ensembles, by using analytical and exact transfer matrix methods, though we concentrate mostly on the fixed force case in this Letter.

Before describing the model, let us point out a few of the basic results we have obtained. The phase diagrams in both the fixed distance and fixed force ensembles are obtained. The qualitative features of the phase diagram, especially the nature of the phases, the s-dependence, and the ensemble dependence are generic as can be shown by general arguments. An "eye"-like configuration exists for all $s < 1$ in the fixed distance ensemble, either as a distinct phase (when it extends up to the anchored end) or as two "Y"'s joined together (see Fig. 2) where a Y is a coexistence of the unzipped and the zipped phases. These configurations resemble the Y fork in replication and the transcription bubble. In a fixed force ensemble, an eye exists only for low values of s ($s < 1/2$ in our model), and the phase diagram depends on s. For values of s where the eye phase exists, there is triple point at the intersection of the zipped-eye (Zp-Ey), eye-unzipped (Ey-Uz), and zipped-unzipped (Zp-Uz) phase boundaries.

The model is defined as follows. The two strands of a homopolymer DNA are represented by two directed random walks on a $d = 2$ dimensional square lattice. The walks, starting from the origin, are restricted to go towards the positive direction of the diagonal axis (z direction) *without crossing each other.* There is a gain of energy $-\epsilon$ ($\epsilon > 0$) for every contact (i.e., separation $x = 0$). The directional nature of the walks takes care of the

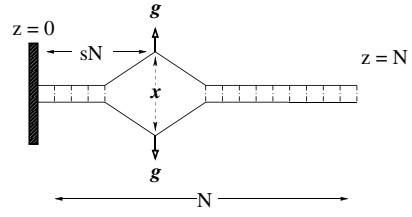

FIG. 1. Schematic diagram of DNA unzipping by a pulling force at a fraction s ($0 \leq s \leq 1$) from the anchored end. In the fixed force ensemble the force g is kept fixed while the separation x is kept fixed in the fixed distance ensemble.

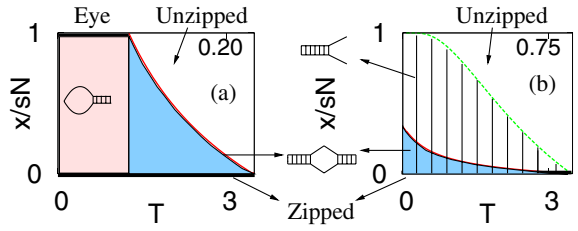

FIG. 2 (color online). Fixed distance ensemble T versus $x/(sN)$ phase diagram for (a) $s = 0.2$ and (b) $s = 0.75$. The zipped and the eye phases are shown by thick lines. The coexistence regions are marked by different shades or vertical lines. $X_c(T)$, defined after Eq. (2), is represented by the dotted line in (b).

correct base pairing of DNA. In addition to this bonding, a force g acts along the transverse direction (x direction) at a fixed fraction s ($0 \le s \le 1$) from the anchored end ($z = 0$). As is well known, this force, though it acts at a point, affects the bulk behavior [4]. The quantities of interest depend on the ensemble one is working with. For example, it is the average separation $\langle x \rangle$ at the point of application of the force in the fixed force ensemble, whereas, in the fixed distance ensemble, it is the average force $\langle g \rangle$ needed to maintain the distance x between the two strands. (g, x) constitute a thermodynamic conjugate pair.

A taste of the surprise for $s \neq 1$ can be gleaned from a simple analysis that is exact in the low temperature region. For $s = 1$, this argument gives the exact reentrant phase boundary [6–8]. If $s < 1/2$, at $T = 0$, the force opens the chain maximally so as to form an eye (extensive in length). The energy E and entropy S of the eye with respect to the completely zipped chain are $E = -gasN + 2\epsilon sN$ and $S = -2sN \ln\mu_b$, where μ_b is the connectivity constant of the bound phase and a is a geometric factor. Throughout this Letter we take the Boltzmann constant $k_B = 1$. For our lattice model, $\mu_b = 2$ and, we set $a = 1$ by choosing the elementary diagonal of the underlying square lattice as the unit of length. A transition from the zipped (Zp) state is therefore possible if $g > g_c(s, T) = 2(\epsilon + T \ln\mu_b)/a$, which is double the force required for the unzipping transition at $s = 1$. The situation is different for $s \ge 1/2$, where complete unzipping is possible at $T = 0$. In the completely unzipped (Uz) state, the energy and entropy with respect to the zipped state are $E = N\epsilon - gasN$ and $S = -N \ln\mu_b + 2(1 - s)N \ln\mu_1$, where μ_1 is the connectivity constant for a single chain. The low temperature phase boundary is given by $g_c(s, T) = (\epsilon + T \ln\mu_b/\mu_1^{2(1-s)})/(sa)$. For our lattice problem $\mu_1 = 2$. Hence, there will be no low temperature reentrance if $s = 1/2$. These s dependences match the exact results.

To trace out the exact phase boundary we use the recursion relation method. Let $D_t(x, x')$ be the partition function with separations x and x' at the two extreme ends of a double stranded DNA of length t. Then,

$$D_{t+1}(x, x') = [2D_t(x, x') + D_t(x, x' + 1) + D_t(x, x' - 1)](1 + P\delta_{x',0}), \quad (1)$$

with $P = e^\beta - 1$, $\beta = 1/T$, and initial conditions $D_0(x, x') = \delta_{x,x'}$. Mutual exclusion is ensured by $D(x, x') = 0$ whenever any (or both) of the two arguments $x, x' < 0$. One can construct two other partition functions, (i) $d_t(x) \equiv D_t(0, x)$ when one end is held fixed, and (ii) $\bar{d}_t(x) \equiv \sum_{x'} D_t(x, x')$ when one end is free. Of these, $d_t(x)$ has been used in previous studies of the force at the end case [8] where the phases and the transitions come from the singularities of the generating function, $\mathcal{G}(z, \beta, g) = \sum_t \sum_x z^t e^{\beta g x} d_t(x)$. These singularities are $z_1 = 1/4$, $z_2 = (2 + 2\cosh\beta g)^{-1}$ and $z_3 = \sqrt{-e^{-\beta} + 1} - 1 + e^{-\beta}$. The zero-force melting, coming from $z_1 = z_3$, is at $T_c = 1/\ln(4/3) = 3.476\,059\,497\ldots$. The unzipping phase boundary can be determined in the fixed distance ensemble by noting that the force required to maintain the separation x is $-T\partial \ln d_N(x)/\partial x$. By using $d_N(x) \approx \lambda(z_3)^x/z_3^{N+1}$ for large N with $\lambda(z) = (1 - 2z - \sqrt{1 - 4z})/(2z)$, one gets

$$g_c(T) \equiv g_c(s = 1, T) = -T \ln\lambda(z_3), \qquad (\mathrm{Zp} \Leftrightarrow \mathrm{Uz}). \quad (2)$$

This is the known $s = 1$ phase boundary [8] coming from $z_2 = z_3$ in the fixed force ensemble. On this boundary, the end separation is given by $X_c(T) \equiv x/N = \tanh[g_c(T)/T]$. The phase coexistence on this boundary gives the Y-fork structure, which we simply call a Y.

Fixed distance ensemble.—If the distance or separation of the two strands at $t = sN$ is kept fixed at x, while the DNA is anchored ($x = 0$) at $t = 0$ but free at the other end at $t = N$, the partition function is (z_3 dependence of λ suppressed)

$$Z_N(x, s) = d_{sN}(x)\bar{d}_{(1-s)N}(x) \approx \lambda^x z_3^{-sN}(4^{(1-s)N} + \lambda^x z_3^{-(1-s)N}). \quad (3)$$

In the limit $N \to \infty$ for a fixed s, the larger of the two terms (in the big parenthesis on the right hand side) contributes to the free energy. For

$$\frac{x}{sN} < X(s, T) \equiv \frac{1 - s}{s} \frac{\ln(4z_3)}{\ln(\lambda(z_3))}, \quad (4)$$

the larger term is the second one, otherwise it is the first one. Therefore, we get a phase boundary

$$g_c(s, T) = 2g_c(T), \qquad \text{if Eq.(4)} \quad (5a)$$

$$= g_c(T), \qquad \text{otherwise}, \qquad (\mathrm{Zp} \Leftrightarrow \mathrm{Uz}). \quad (5b)$$

With the increase of the separation x, the end point gets detached at the critical value $x = sNX(s, T)$, provided $X(s, T) < 1$. Once all the bonds are broken the two open tails behave like free independent chains. In such a situation, the force required to maintain the separation is just like the $s = 1$ end case (in the $sN \to \infty$ limit) as we see in

Eq. (5a). For this to happen we also require $X(s,T) < X_c(T)$, or else the DNA will be in the unzipped phase. Figure 2 shows the phase diagram on a temperature-distance plane for two values of s in the fixed distance ensemble, though in the force-temperature plane the phase boundaries are independent of s as follows from Eqs. (5a) and (5b). An eye of the type shown in Fig. 1 occurs in the coexistence region shown by solid vertical lines in Fig. 2 and is to be interpreted as two Y's joined together. This configuration is analogous to the transcription bubble produced, e.g., by RNA polymerase [2] a subunit of which keeps the two strands of DNA separated.

To supplement the exact results, the partition function for the two strands starting from the origin is obtained numerically by using the exact transfer matrix technique for the recursion relation in Eq. (1). In the fixed distance ensemble, the distance between the two strands x is varied at a step of one (length of the diagonal of the square lattice) from zero ("zipped") to sN ("completely stretched"). The quantity of interest, the average force required to maintain the distance x between two chains, is calculated by using finite differences in free energy. The scaled separation between the two strands of DNA, x/sN, versus the corresponding average force at $T = 1.0$ is shown in Fig. 3(a), for several values of s. Figure 3(b) shows the eye formation in the fixed force ensemble and is discussed next. When the end monomers of the chains ($s = 1.0$) are maintained at a fixed distance, there is only one plateau at the critical force given by Eq. (2) (the Zp-Uz phase boundary). When $s < 1.0$, the force-distance isotherm has two plateaus as per Eqs. (5a) and (5b) with a step that matches with the critical value $X(s,T)$. We do not go into further details of this phase diagram here because the subtleties are more prominent in the fixed force ensemble.

Fixed force ensemble.—In the fixed force ensemble, the generalization of the generating function defined for the end case below Eq. (1) is $\mathcal{G}_s(z,\beta,g) = \sum_{x>0} e^{\beta g x} \times \sum_t z^t Z_t(x,s)$. Using Eq. (3), this can be written as

$$\mathcal{G}_s(z,\beta,g) \approx \sum_x e^{\beta g x}\{[\lambda(4^{(1-s)/s} z^{1/s})]^x + [\lambda(z)]^{2x}\}, \quad (6)$$

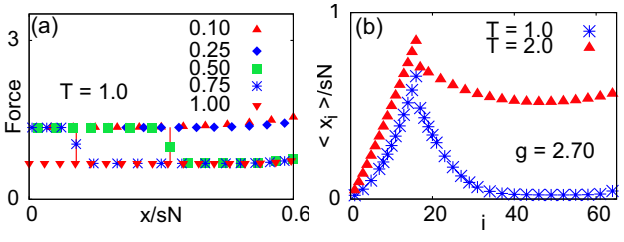

FIG. 3 (color online). (a) Scaled separation $x/(sN)$ versus force isotherm for different s at $T = 1.0$ with $N = 256$ in the fixed distance ensemble. The location of $X(s,T)$ (see text) is shown by the solid line. (b) Average separation (fixed force ensemble) between the ith monomer of the DNA of length $N = 64$ for $T = 1.0$ for a force $g = 2.7$ at $s = 0.25$ at two different temperatures.

where the first term on the right hand side in curly brackets represents the unzipped state while the second term is for the eye state. The additional singularities in z are then $z_2 = [4^{1-s}(2(1+\cosh\beta g))^s]^{-1}$, which goes over to the z_2 mentioned earlier for $s = 1$, and $z_4 = \{2[1+\cosh(\beta g/2)]\}^{-1}$. The zipped-to-unzipped phase transition occurs at $z_2 = z_3$ while a zipped-to-eye phase transition takes place at $z_3 = z_4$.

The s dependence of the singularities show that there cannot be an eye phase in the fixed force ensemble if $s \geq 1/2$, even though one may open an eye of the type of Fig. 1 in the fixed distance ensemble. In this situation, the only transition possible is unzipping with an s-dependent boundary given by

$$g_c(s,T) = T\ln\lambda(4^{(1-s)/s} z_3^{1/s}) \quad (\mathrm{Zp} \Leftrightarrow \mathrm{Uz}), \quad (7)$$

which matches with $g_c(T)$ for the end case. In addition, close to $T = 0$, we see $g_c(s,T) \approx \frac{\epsilon}{s} + \frac{2s-1}{s}\ln 2$, which corroborates the results from the simple argument, including the vanishing slope at $T = 0$ for $s = 1/2$ (the absence of a low temperature reentrance). The inset in Fig. 4(a) shows that there is still a small region in the intermediate temperature range where one does see a reentrance.

With a force at a point $s < 1/2$, a phase boundary comes from $z_3 = z_4$, which matches with the boundary we already derived in Eq. (5a). This is the transition to eye in presence of a force, as found in the fixed distance ensemble also. However the eye phase cannot continue for the whole temperature range, definitely not at the melting point where the unzipped phase should be recovered. The unzipped phase boundary comes from $z_2 = z_4$, which yields Eq. (7). The lower of the two curves would determine the thermodynamic phase boundary; see Fig. 4. These results also suggest the possibility of an eye to unzipped phase transition which however eludes the approximation done in Eq. (6) [based on Eq. (3)]. This phase boundary is determined numerically below. The intersection of the three boundaries, which occurs only for $s < 1/2$, is a *triple point.*

To extend the analysis for $s < 1/2$, and to verify the analytical results, we use the exact transfer matrix technique, mentioned already, but now with an applied force.

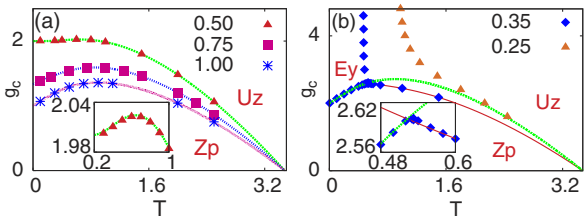

FIG. 4 (color online). The g_c versus T phase diagram in the fixed force ensemble. Lines are the exact results while the points are from numerics. (a) For $0.5 \leq s \leq 1$. The inset shows the reentrant region for $s = 0.5$. (b) For $s = 0.35$ and $s = 0.25$. The inset shows the triple point for $s = 0.35$.

The average separation $\langle x \rangle$ between the two chains at the site of application of force at a temperature T, is calculated by taking the finite difference of exact free energies as g is increased in steps of $\Delta g = 0.001$. The critical force in the thermodynamic limit is determined by $N \to \infty$ extrapolation of the crossing points of $\langle x \rangle$ versus g curves for pairs of length N. We take N from 400 to 1000. The results are shown in Fig. 4, with very good agreement with the analytical results.

The maximum separation between the chains at the point of application of the force is sN, and at this separation, the end monomers are always in an unzipped state for $0.5 \leq s \leq 1$ even for $T \to 0$. This is not so for $s < 0.5$. In this case, the end monomers will of course be unzipped at higher temperatures but that does not rule out the possibility of a zipped phase at low temperatures. To study how the end monomer separation behaves with T when $s < 0.5$, the force is fixed to a value which lies on the phase boundary obtained above and T is increased. At low temperatures, we find that the end monomers are in a zipped state (i.e., in contact) even when the separation between the two chains, where force is applied, is maximum. At a particular temperature, which depends on the fraction s, the end monomer separation becomes macroscopic, signaling an Ey-Uz phase transition. Figure 4 shows the phase boundary obtained by repeated application of this procedure.

The complete phase diagram for $s = 0.35$ and 0.25 is shown in Fig. 4(b). The three coexistence lines meet at a triple point T_p which shifts with s. The triple point moves towards the high temperature side when s is decreased and merges with the melting point for $s = 0.25$. The inset shows the details around the triple point. The limit $s \to 0$ is singular because the eye phase cannot exist and a force applied at the anchored point cannot open a chain. In that limit, the Ey-Uz boundary becomes vertical (parallel to the force axis) and is the only meaningful phase boundary.

The Ey-Uz transition can be seen in Fig. 3(b) which shows the average separation between the monomers of the DNA of length $N = 64$ with a force at $s = 0.25$. We calculate the base pair separations for two different temperatures both at $g = 2.7$, which lies on the phase boundary of the zipped and the eye phases at $T = 1$, and another for the same force but at $T = 2$ deep in the unzipped region (see Fig. 4). The monomers in the outer most part of the DNA are in the zipped state for $T = 1$ and $g = 2.7$, showing the formation of the eye, but for $T = 2$ the DNA is in the unzipped phase. In the limit of $N \to \infty$ the number of bound pairs at the end of the chain is extensive ($\propto N$) in the eye phase but not in the unzipped phase.

We may now summarize some of the results of this Letter. (i) The phase diagrams in the fixed distance and fixed force ensemble are shown in Figs. 2 and 4, which also show when the eye appears as a distinct phase or as two Y's. (ii) There is a low temperature reentrance in the fixed force ensemble for all values of s except for $s = 1/2$. In the latter case the reentrance is restricted to a small intermediate temperature range. (iii) In the fixed force ensemble, the phase boundary shifts with s as it is decreased from $s = 1$ to $s = 1/2$. In this range there is no triple point. (iv) For $s < 1/2$, the low temperature phase boundary representing the zipped-to-eye phase transition in the fixed force case is independent of s and it intersects the s-dependent zipped-to-unzipped boundary at a triple point. There is an additional eye-to-unzipped phase transition line in the large force regime. (v) The triple point shifts towards the zero-force melting point T_c as s is decreased. Although our approach is based on a coarse grained model, we believe that our results are robust to be observed by high precision measurements of DNA unzipping under a pulling force.

We thank Sanjay Kumar for suggesting the relevance of s in unzipping and acknowledge the support from PRIN-MURST 2003 and FISR 2001.

*Electronic address: rajeev@iopb.res.in
†Electronic address: somen@iopb.res.in
‡Electronic address: flavio.seno@pd.infn.it

[1] B. Alberts, Nature (London) **421**, 431 (2003).
[2] T. H. Tahirov *et al.* Nature (London) **420**, 43 (2002); G. M. T. Cheetham, D. Jeruzalmi, and T. A. Steitz, Nature (London) **399**, 80 (1999).
[3] P. R. Cook, Science **284**, 1790 (1999) and references therein.
[4] S. M. Bhattacharjee, J. Phys. A **33**, L423 (2000).
[5] K. L. Sebastian, Phys. Rev. E **62**, 1128 (2000).
[6] D. Marenduzzo, A. Trovato, and A. Maritan, Phys. Rev. E **64**, 031901 (2001).
[7] E. Orlandini *et al.*, J. Phys. A **34**, L751 (2001).
[8] D. Marenduzzo *et al.*, Phys. Rev. Lett. **88**, 028102 (2002).
[9] H. Zhou, cond-mat/0007015.
[10] Jeff Z. Y. Chen, Phys. Rev. E **66**, 031912 (2002).
[11] D. K. Lubensky and D. R. Nelson, Phys. Rev. Lett. **85**, 01 572 (2000); Phys. Rev. E **65**, 031917 (2002).
[12] Y. Kafri, D. Mukamel, and L. Peliti, Eur. Phys. J. B **27**, 135 (2002).
[13] Puri-Man Lam *et al.*, Biopolymers **73**, 293 (2004).
[14] C. Danilowicz *et al.*, Phys. Rev. Lett. **93**, 078101 (2004).
[15] D. Anselmetti *et al.*, Single Mol. **1**, 53 (2000).
[16] B. Essevaz-Roulet, U. Bockelmann, and F. Heslot, Proc. Natl. Acad. Sci. U.S.A. **94**, 11 935 (1997); U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, Phys. Rev. E **58**, 2386 (1998).
[17] D. Poland and H. Scheraga, J. Chem. Phys. **45**, 1464 (1966).
[18] S. M. Bhattacharjee and D. Marenduzzo, J. Phys. A **35**, L349 (2002).
[19] D. Keller *et al.*, Biophys. J. **84**, 733 (2003); R. M. Neumann, *ibid.* **85**, 3418 (2003).