Unzipping DNA by force: thermodynamics and finite size behaviour

2006 J. Phys.: Condens. Matter 18 S215

(http://iopscience.iop.org/0953-8984/18/14/S06)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **18** (2006) S215–S223 doi:10.1088/0953-8984/18/14/S06

Unzipping DNA by force: thermodynamics and finite size behaviour

Rajeev Kapri[1] **and Somendra M Bhattacharjee**[1,2]

[1] Institute of Physics, Bhubaneswar 751 005, India
[2] TCMP, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata 700 064, India

E-mail: rajeev@iopb.res.in and somen@iopb.res.in

Received 5 October 2005
Published 24 March 2006
Online at stacks.iop.org/JPhysCM/18/S215

Abstract

We discuss the thermodynamic behaviour near the force induced unzipping transition of double-stranded DNA in two different ensembles. The Y-fork is identified as the coexisting phases in the fixed distance ensemble. From finite size scaling of thermodynamic quantities like the extensibility and the length of the unzipped segment of a Y-fork, the phase diagram can be recovered. We suggest that such procedures could be used to obtain the thermodynamic phase diagram from experiments on finite length DNA.

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

1. Introduction

The force induced unzipping transition [1] provides a mechanism for separating the two strands of double-stranded DNA without a drastic change of environment like changing the temperature or pH. In the last few years, the transition has been studied theoretically [1–15] and experimentally [16, 17]. The motion of a helicase has also been studied with phase coexistence at unzipping as the underlying mechanism [18, 19].

With the advent of single-molecule experiments, it is now possible to explore the behaviour of DNA, be it single or with other complexes [16, 20, 21]. These experiments, and also biological molecules, work effectively in different ensembles. For example the atomic force microscope (AFM) works in the fixed distance ensemble while several magnetic bead methods use the fixed force ensemble. Helicases like dnaB work in the fixed distance ensemble, but PcrA type helicases have both the fixed distance and the fixed force character (see, e.g., the references in [18, 19]). Consequently, these explore or take advantage of the various features of the phase diagrams. With this in mind, we would like to discuss the thermodynamic behaviour near the unzipping transition under either an applied fixed pulling force (fixed force ensemble) or for a fixed distance between the strands (fixed distance ensemble) at a preassigned fraction s from the anchored end. After a brief review of the known phase diagrams in sections 2 and 3,

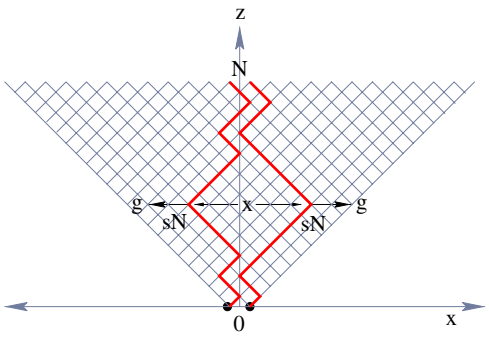

Figure 1. Schematic diagram of DNA unzipping by a pulling force at a fraction s ($0 \leqslant s \leqslant 1$) from the anchored end. In the fixed force ensemble the force g is kept fixed while the separation x is kept fixed in the fixed distance ensemble.

we show how the results from finite length DNA can be used to extract the phase diagram and also to determine other thermodynamic quantities. An interesting, and surely realizable, situation would be a closed cycle in the thermodynamic plane with the two ensembles revealing features of two types. These features are verifiable in experiments.

2. Model and ensembles

The two strands of a homopolymer DNA are mimicked by two directed random walks on a $d = 2$-dimensional square lattice. The walks start from the origin and are allowed to go in the positive direction of the diagonal axis (z-direction) *without crossing each other*. The non-crossing constraint is important for real DNA. The base pairing between the two strands of DNA is modelled by a contact energy $-\epsilon$ ($\epsilon > 0$) for every contact (i.e. separation $x = 1$; see figure 1). The directional nature of the walks takes care of the correct base pairing of DNA. In addition to this bonding, either a constant force g acts along the transverse direction (x-direction) or the separation between the strands x is kept constant at a fixed fraction s ($0 \leqslant s \leqslant 1$) from the anchored end ($z = 0$). The schematic diagram of our model is shown in figure 1.

The fixed distance and the fixed force ensembles at temperature T and chain length N are characterized by the partition functions

$$Z_N(x, T) = \sum \exp(n\beta\epsilon), \qquad \text{and} \qquad \mathcal{Z}_N(g, T) = \sum_x Z_N(x, T) \exp(\beta g x) \tag{1}$$

respectively, where, in Z_N, the sum is over only those configurations which have the separation x at the point in question (fixed distance), while in $\mathcal{Z}_N$, the sum is over all configurations with n base pairs and separation x with summations over all allowed values of n and x, for a given force g (fixed force). Here $\beta = 1/k_B T$ is the inverse temperature, k_B being the Boltzmann constant. We may choose units with $k_B = 1$ and $\epsilon = 1$. The free energies are given by

$$F_N(x, T) = -k_B T \ln Z_N(x, T) \qquad \text{and} \qquad \mathcal{F}_N(g, T) = -k_B T \ln \mathcal{Z}_N(g, T), \tag{2}$$

respectively.

In the fixed force ensemble, the separation fluctuates, and one defines the average separation $\langle x \rangle$, but in the fixed distance ensemble, one needs the average force to maintain the separation. These in the corresponding ensembles are given by

$$\langle x \rangle = -\frac{\partial \mathcal{F}_N(g, T)}{\partial g}, \qquad \text{and} \qquad \langle g \rangle = \frac{\partial F_N(x, T)}{\partial x} \tag{3}$$

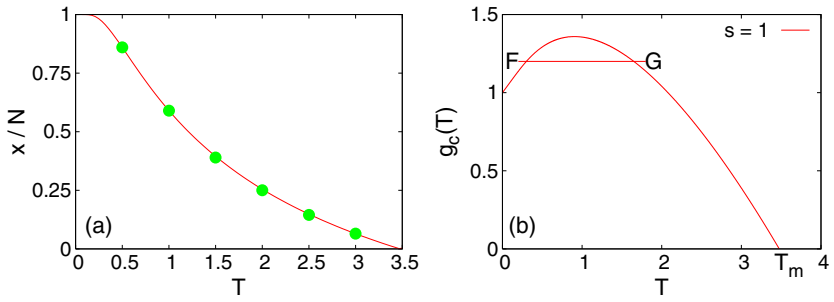

Figure 2. (a) The region of coexistence of the zipped and unzipped phases when the separation between the last monomers ($s = 1$) of the DNA is kept constant. The points are from the transfer matrix based numerical approach as discussed in the text and the solid line is the analytical result. (b) Phase diagram in the g–T plane in the fixed force ensemble. The path FG is used for the specific heat.

at a fixed T, with the angular brackets $\langle \cdots \rangle$ denoting thermal averaging in the appropriate ensemble.

In all these cases, the phase transition is obtained only in the limit $N \to \infty$. We show how finite size scaling can be used to extract the $N \to \infty$ phase diagram and other information from finite N results. Since experiments are done on short DNA, we suggest that a similar approach should be adopted to infer the thermodynamic behaviour.

3. Phase diagrams

Let us recapitulate the phase diagrams for this model for a few typical values of s in two different ensembles [5, 6].

3.1. $s = 1$: pulling at the end

In the case where the force is applied at the free end, there are only two possible phases, namely the zipped and the unzipped phases. The free energies per base pair of these two phases are given by

$$f_{\rm zp}(T, s = 1) = -k_{\rm B}T \ln z_3(T), \qquad \text{and} \qquad f_{\rm uz}(T, s = 1) = -k_{\rm B}T \ln z_2(T), \tag{4}$$

where

$$z_3(T) = \sqrt{1 - {\rm e}^{-\beta\epsilon}} - 1 + {\rm e}^{-\beta\epsilon} \qquad \text{and} \qquad z_2 = (2 + 2\cosh \beta g)^{-1}. \tag{5}$$

The force–temperature phase boundary between the two phases in the fixed force ensemble is given by

$$\begin{aligned} g_{\rm c}(T) &= k_{\rm B}T \cosh^{-1}\left[\frac{1}{2}\frac{1}{\sqrt{-{\rm e}^{-\beta\epsilon} + 1} - 1 + {\rm e}^{-\beta\epsilon}} - 1\right] \\ &= -k_{\rm B}T \ln \lambda(z_3), \end{aligned} \tag{6}$$

where $\lambda(z) = (1 - 2z - \sqrt{1 - 4z})/(2z)$. This boundary is shown in figure 2(b).

In the absence of any force, the DNA can be denatured simply by increasing the temperature. The denaturation (melting) temperature, $T_{\rm m} = 1/\ln(4/3) = 3.476\,059\,497\ldots$ corresponds to a pure second-order phase transition with a discontinuity in the specific heat. *The unzipping transition is however first order.* The end point separation of the two strands

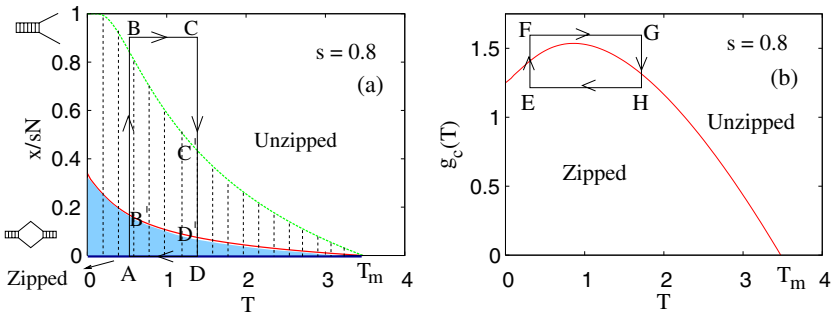

Figure 3. The phase diagrams for $s = 0.8$. (a) Fixed distance ensemble and (b) fixed force ensemble.

on the phase boundary jumps from a small value in the zipped phase to an extensive number (i.e., proportional to the length)

$$X_c(T) \equiv x/N = \tanh[g_c(T)/T]. \tag{7}$$

In the fixed distance ensemble, the phase diagram shows the coexistence region demarcated by the line $X_c(T)$ of equation (7), as shown in figure 2(a). *This coexistence region resembles the Y-fork structure* formed during DNA replication [22].

3.2. $s > 1/2$

Figure 3 shows the phase diagrams for $s = 0.8$ in both the fixed distance and the fixed force ensemble. These are representatives of all $s > 1/2$, for which the double-stranded DNA can be fully unzipped at low temperatures either by increasing the separation between the two strands in the fixed distance ensemble or by increasing the force in the fixed force ensemble. In the fixed force ensemble, the DNA is either in the zipped or in the unzipped phase depending on the applied force and the working temperature. The free energy per base pair for the unzipped phase is now

$$f_{uz}(T, s) = -k_B T \ln z_2(T, s), \qquad z_2(T, s) = [4^{1-s}(2 + 2\cosh \beta g)^s]^{-1}. \tag{8}$$

The s-dependent phase boundary is given by

$$g_c(s, T) = k_B T \ln \lambda(4^{(1-s)/s} z_3(T)^{1/s}). \tag{9}$$

The difference (from equation (6)) owes its origin to the extra entropy of the unzipped chain between the point of application of the force and the free end.

However an eye of the type of figure 1 is possible in the fixed distance ensemble. This is one of the ensemble differences. The phase diagram in the fixed distance ensemble shows two possible coexistence regions: (i) a region (shown by vertical lines) in which one end of the DNA is unzipped into two single strands and the other is a double strand—such configurations resemble the Y-fork as in the $s = 1$ case; (ii) a region (shown by shading) having the configurations of the type shown in figure 1 —these configurations resemble the transcription bubble formed, e.g. by RNA polymerase, during RNA transcription.

Quantitatively, for

$$\frac{x}{sN} < X(s, T) \equiv \frac{1-s}{s} \frac{\ln(4z_3)}{\ln(\lambda(z_3))}, \tag{10}$$

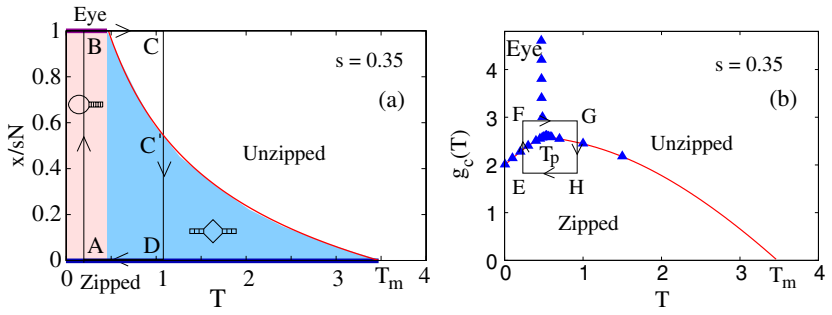

Figure 4. The phase diagrams for $s = 0.35$. (a) Fixed distance ensemble and (b) fixed force ensemble.

one finds bubbles beyond which the free end unzips yielding the Y-type configurations, though with free wings. In terms of the force required to maintain the distance, there will be two situations:

$$g_c(s, T) = 2g_c(T), \qquad \text{if equation (10) is satisfied} \tag{11}$$

$$= g_c(T), \qquad \text{otherwise.} \tag{12}$$

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 that in the $s = 1$ end case (in the $sN \to \infty$ limit) in equation (11). For this to happen we also require $x/sN < X_c(T)$; otherwise we have the unzipped phase.

3.3. $s < 1/2$

If a force is applied at $s < 1/2$, a new phase occurs, namely the 'eye' phase which involves the anchored point at $z = 0$ as one of its extremities. The phase diagrams in the two ensembles are shown in figure 4. The boundary separating the zipped from the unzipped phase matches with the boundary that we already derived in equation (11). This transition to the eye phase in the presence of a force is also found in the fixed distance ensemble. The zipped–unzipped phase boundary remains the same as equation (9). The bubbles are formed in the zipped–unzipped transition. Because of the presence of a third phase, a triple point appears in the phase diagram, easily identifiable in the fixed force ensemble phase diagram for $s = 0.35$ (figure 4(b)).

4. Finite size scaling

The response of the DNA to the pulling force can be defined by the extensibility χ as

$$\chi_T = \left.\frac{\partial x}{\partial g}\right|_T, \tag{13}$$

which in the fixed force ensemble can be related to the fluctuation in x,

$$\chi_T = \frac{1}{k_B T}(\langle x^2\rangle - \langle x\rangle^2). \tag{14}$$

This extensibility is an extensive quantity, proportional to the length of the chain. For convenience, we shall omit the subscript T unless required.

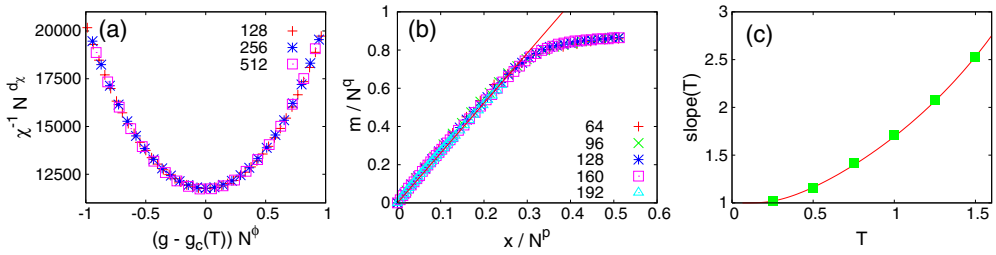

Figure 5. (a) Data collapse of the inverse extensibility at $T = 0.5$ for various lengths and forces. (b) Data collapse at $T = 1.5$ for the length m of the Y-fork with separation x between the end monomers of two strands. The straight line is the best fit to the linear region of the collapse and different symbols represent different chain lengths. (c) The slope of the linear region of the collapse as a function of temperature. The points are from numerics and the solid line is the analytical result.

A phase coexistence is characterized by a flat $g-x$ isotherm. This means a finite change in x can be induced without change in g yielding an infinite χ. A diverging χ is however not possible for finite chains. Instead a finite size scaling behaviour is expected. Such a scaling form would allow one to determine the phase boundary from finite size data and would be useful in cases where exact solutions are not known.

The free energy at a given value of x (with say $s = 1$) for a finite value of N can be determined with arbitrary precision by a transfer matrix approach. By taking finite differences of the free energy for different values of x we can determine the force and the extensibility at a given temperature. The inverse extensibility shows the finite size scaling form

$$\chi^{-1} = N^{-d_\chi} \mathcal{F}((g - g_c(T))N^{\phi}), \tag{15}$$

and a data collapse is obtained for $d_\chi = 2$ and $\phi = 1$ as determined by the Bhattacharjee–Seno method [23]. The data collapse is shown in figure 5(a). Keeping in mind that χ is extensive, these exponents show that the extensibility has a divergence as

$$\chi \sim |g - g_c(T)|^{-1}, \tag{16}$$

as has been predicted in continuum models [1].

A similar scaling form with x as the variable allows us to locate the end point of the coexistence region in the fixed distance ensemble. We used the same procedure at different temperatures and trace out the coexistence region for the DNA problem in the fixed distance ensemble. The results (circles) obtained via this procedure agree very well with the exact results (solid curve) in figure 2. That the finite size exponent $\phi = 1$ in equation (16) follows from a careful analysis of the singularities of the generating function used for the exact solution [5, 6].

Another quantity of interest is the length m, of the unzipped segment in the Y-fork as a function of the separation x between the end monomers of two strands of DNA at a fixed temperature. Phase coexistence means the DNA is segregated into an unzipped region of length m and a zipped region of length $N - m$. If the equation of state of the unzipped chain of length N is written as $x = N\mathcal{A}(g/T)$, then for a given separation x, one expects the length of the unzipped chain to be

$$m(x) = x/\mathcal{A}(g_c(T)/T) = x/\tanh(g_c(T)/T), \tag{17}$$

where we have used the equation of state as given by equation (7).

From the transfer matrix approach, the length of the unzipped chain of the Y-fork can be determined. It exhibits a finite size scaling of the form

$$m(x) = N^q \mathcal{M}(x/N^p) \tag{18}$$

with $q = p = 1$. Figure 5(b) shows the data collapse at $T = 1.5$ for different chain lengths. We find that $m(x)$ behaves linearly for small end separation x. The straight line is the best fit to this linear region. In figure 5(c) we have plotted the slope of this linear region as a function of temperature. The points are the slope of the fitted line and the solid curve is the known analytical result from equation (17). This result corroborates the basic tenet of *coexistence* of the two phases on the one-dimensional DNA chain.

4.1. Specific heat

The divergence of χ and the temperature dependence of m at coexistence can be used to determine the specific heat behaviour. There are two specific heats to be considered, the constant x specific heat c_x and the constant force specific heat c_g.

The constant x specific heat (per base pair) follows from the bulk free energy

$$Nf(T, x) = mf_{\mathrm{uz}}(T) + (N - m) f_{\mathrm{zp}}(T, x). \tag{19}$$

Taking derivatives and noting that the two free energies are the same at coexistence, the constant x specific heat $c_x(T)$ close to the transition is given by

$$c_x(T) \approx \frac{m}{N} c_{x,\mathrm{uz}}(T) + \frac{N - m}{N} c_{\mathrm{zp}}(T) + \frac{\mathrm{d}m}{\mathrm{d}T}\mathcal{L}, \tag{20}$$

where $\mathcal{L}$ is the latent heat. At constant x, as the temperature is changed, the heat supplied goes partly in changing the temperature of the individual components and also in converting a part of the zipped chain to an unzipped form, i.e. changing m. For the last process, a latent heat is required. This is the content of the above equation (20). As the phase boundary is approached, m/N approaches 1 smoothly; see equation (17). Beyond the phase boundary, there is only unzipped chain. Therefore, the constant x specific heat as a function of T shows a *jump discontinuity* at the phase transition point though it is a first-order transition. The peculiarity comes from the fact that for the most of the temperature range in the constant distance ensemble, we see phase coexistence, i.e. a Y-fork.

If we consider the specific heat at a constant force, then we cross the phase boundary at the transition point only, while at all other temperatures we are strictly in one phase. Therefore, below the transition point ($T < T_{\mathrm{c}}(g)$) we see the specific heat of the zipped chain, while above ($T > T_{\mathrm{c}}(g)$) we see that of the unzipped chain. At the transition point for the given force ($T = T_{\mathrm{c}}(g)$), the latent heat supplied goes into changing the phase but not the temperature. Consequently, a delta function peak (divergence) appears in c_g at $T = T_{\mathrm{c}}(g)$. This divergence of c_g also follows from the thermodynamic relation of the two specific heats,

$$c_g - c_x = T\chi_T \left(\frac{\partial g_{\mathrm{c}}(T)}{\partial T}\bigg|_x \right)^2. \tag{21}$$

Since c_x is finite at all T, we see that the divergence of c_g at $T_{\mathrm{c}}(g)$ is linked to the divergence of χ at that point. A similar analysis can be done for the re-entrant part of the phase boundary also.

The analytical results for the thermodynamic limit and the numerically calculated specific heat for a finite length of DNA with a force that admits re-entrance are shown in figure 6. The two peaks for the two transitions are prominent in the finite size results.

5. Closed loops in the phase diagram

The s-dependent phase diagrams can be used to interpret the effect of helicases and other motor proteins on DNA which keep the separation between the two strands constant. They may be

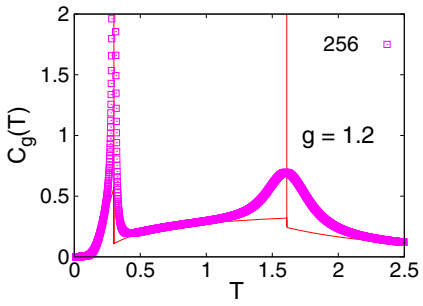

Figure 6. Specific heat c_g as a function of temperature for a value of g that allows re-entrance (path FG in figure 2). The solid line is from analytical results in the $N \to \infty$ limit and points are from numerical results for a finite chain of length $N = 256$.

classified as fixed distance objects (FDOs). Take for example dnaA, which is known to initiate the replication process. It generally starts at a point near an end which means $s < 1$. As it starts moving towards the end, s starts increasing and the s dependence of the phase boundary of figure 3 (equation (9)) shows that at some value of s for the given T, the end points get released. This is accompanied by a sudden change in the force (equations (11), (12)) on the FDO and this sudden change can be responsible for its offloading. Similarly, helicases have components for maintaining a fixed distance between the two strands, to behave as FDOs. The motor action and any further destabilizing effects (like force in the case of pcrA) are equivalent to a thermodynamic force on the junction of the two coexisting phases at the Y-fork ('domain wall'). The opening of the DNA by a helicase can then be interpreted as a motion of the wall or a propagating front through the double-stranded DNA under a local time-dependent instability. The velocity of the front (i.e. opening of the DNA) can be related to the elastic behaviour of the wall [19].

Let us now discuss the features that one should see if one moves in a closed loop on the phase diagram.

Suppose we start with a DNA in its zipped phase (point A in figure 3(a)) in the fixed distance ensemble and use a tip (say of AFM) to isothermally increase the separation between the monomers of two strands at a fraction $s > 1/2$ from the anchored end. As we move on line AB in the phase diagram, a bubble starts emerging for small x. The size (i.e. length) of the bubble grows with separation until we cross the solid curve at B$'$ where the free end of the DNA gets unzipped into two single strands. This curve (for B$'$) represents the jump from one coexistence region to another with a *sudden, detectable change in force* (equations (11) and (12)). Point B$'$ corresponds to the separation $X(s, T)$ of equation (10). Once in the unzipped region at point B, we raise the temperature to point C. There is no change in the phase though the elastic response of the unzipped chains will show a T dependence. On releasing the tip, i.e. along the path CD, we first see the formation of a Y-fork at C$'$ via zipping of the base pairs *from the anchored end* and a sudden change in force will be felt. This force remains constant until we reach the point D$'$ when the free end of the DNA zips leading to the formation of a bubble. The force gets doubled at this point. By a reduction of T we may then close the loop to get back the original state.

A closed loop EFGH of figure 3(b) will show only the zipped and the unzipped phases, with the Y-fork appearing at special values of the force and/or temperature.

A similar closed loop cycle may be performed for $s < 1/2$. Again we start from the zipped phase at a low temperature (point A in figure 4(a)) and increase the separation between

the strands at fraction $s < 1/2$ keeping the temperature constant. For any x, we get the 'eye' with a zipped tail ('tadpole'). This state is characterized by all open bonds at the anchored end and the free end of DNA is still in its zipped state. If the temperature is increased, the eye phase of DNA gets denatured at a well defined temperature *which is below the bulk melting temperature* T_m. At C one gets two single strands. If we decrease x at this new temperature (path CD) the two strands of DNA start feeling each other's presence at C′ where we cross the solid curve at a (temperature-dependent) critical separation $X(s, T)$ at which both the ends of the DNA get re-zipped and a bubble is formed on the DNA. By cooling one gets back into the eye phase. Further decrease of separation brings the DNA back to its initial zipped phase.

In the fixed force ensemble, we start from the zipped region at low temperature (point E). The closed loop EFGH takes us from the zipped part to the eye (at F), then to the unzipped part (at G) and then back to the zipped phase (at H, E). We see a bubble formation only at specific points on the cycle.

6. Conclusion

We have shown how the response function, or extensibility, of a double-stranded DNA to a pulling force can be used to obtain the thermodynamic phase diagram. Our numerical procedure on short chains reproduced the exact phase boundary. We suggest that a similar procedure be adopted for results from experiments which are necessarily on short DNA chains. The behaviour of the specific heat is also studied. The signature of the phase diagram can be felt in a thermodynamic cycle. Differences in ensemble (fixed distance versus fixed force) are also shown in various situations, especially in closed loops. These are verifiable in single-molecule experiments with DNA.

References

[1] Bhattacharjee S M 2000 *J. Phys. A: Math. Gen.* **33** L423 (*Preprint* cond-mat/9912297)
[2] Sebastian K L 2000 *Phys. Rev.* E **62** 1128
[3] Marenduzzo D, Trovato A and Maritan A 2001 *Phys. Rev.* E **64** 031901
[4] Orlandini E *et al* 2001 *J. Phys. A: Math. Gen.* **34** L751
[5] Marenduzzo D *et al* 2002 *Phys. Rev. Lett.* **88** 028102
[6] Kapri R, Bhattacharjee S M and Seno F 2004 *Phys. Rev. Lett.* **93** 248102
[7] Zhou H 2000 *Preprint* cond-mat/0007015
[8] Chen J Z Y 2002 *Phys. Rev.* E **66** 031912
[9] Lubensky D K and Nelson D R 2000 *Phys. Rev. Lett.* **85** 01572
Lubensky D K and Nelson D R 2002 *Phys. Rev.* E **65** 031917
[10] Kafri Y, Mukamel D and Peliti L 2002 *Eur. Phys. J.* B **27** 135
[11] Lam P-M *et al* 2004 *Biopolymers* **73** 293
[12] Singh N and Singh Y 2005 *Eur. J. Phys.* E **17** 7
[13] Kumar S, Giri D and Singh Y 2005 *Europhys. Lett.* **70** 15
[14] Kumar S, Giri D and Bhattacharjee S M 2005 *Phys. Rev.* E **71** 051804
[15] Kapri R and Bhattacharjee S M 2005 *Phys. Rev.* E **72** 051803
(Kapri R and Bhattacharjee S M 2005 *Preprint* cond-mat/0505315)
[16] Essevaz-Roulet B, Bockelmann U and Heslot F 1997 *Proc. Natl Acad. Sci. USA* **94** 11935
[17] Danilowicz C *et al* 2004 *Phys. Rev. Lett.* **93** 078101
[18] Bhattacharjee S M and Seno F 2003 *J. Phys. A: Math. Gen.* **36** L181
[19] Bhattacharjee S M 2004 *Europhys. Lett.* **65** 574
[20] Bianco P *et al* 2001 *Nature* **409** 374
[21] Ha T *et al* 2002 *Nature* **419** 638
[22] Alberts B 2003 *Nature* **421** 431
[23] Bhattacharjee S M and Seno F 2001 *J. Phys. A: Math. Gen.* **34** 6375