J. Phys. A: Math. Gen. **33** (2000) L423–L428. Printed in the UK PII: S0305-4470(00)15829-9

LETTER TO THE EDITOR

Unzipping DNAs: towards the first step of replication

Somendra M Bhattacharjee
Institute of Physics, Bhubaneswar 751 005, India
E-mail: somen@iopb.res.in

Received 21 July 2000

Abstract. It is shown that a double-stranded DNA can be opened by a force only if the force exceeds a critical value, and this unzipping is a critical phenomenon. From the results of an equivalent delocalization in a non-hermitian quantum mechanics problem we show the different scaling behaviours of unzipping by force and thermal melting. Based on this we make a postulate on the first step of replication of DNA.

A double-stranded DNA (dsDNA) melts if the pH or temperature of the solution is changed ('thermal melting') but *in vivo*, during replication, enzymes open it up forming the replication fork, a Y-shaped structure for a linear molecule or an eye in general [1]. It is only recently that the unzipping phenomenon is getting attention, in contrast to melting of dsDNA [2]. In several experiments, the force required to unzip a dsDNA has been measured [3]. Various aspects of the unzipping of dsDNA seem to be explained by assuming thermal equilibrium [4]. The results of the force measurements of [3] distinguish the cases of conventional thermal melting and the unzipping which has been called *directional melting*†

Our aim in this paper is to investigate quantitatively in a simple mathematically tractable model the difference between unzipping by force (or directional melting) and thermal or fluctuation-induced melting of dsDNA. Some of the quantitative questions are the following: (1) Is there a critical force to open up a double-stranded chain in thermal equilibrium, acting say at one end only? (2) Is the nature of the transition different from the thermal melting of the bound pair and is it reflected in the opened region? Based on the results of this, towards the end of this Letter, we speculate on the biological significance of these issues and make a postulate on the enzymatic activity.

In order to focus on the effect of the pulling force (see figure 1) on the bound strands, we take the viewpoint of a simple minimal model that transcends microscopic details but on which further details can be added for a realistic situation. Our approach differs from the previous studies [5–7] in the emphasis on *the opening of the fork*. For simplicity, we consider homo-nucleotides and treat the DNA as consisting of two flexible interacting elastic strings tied together at one end. The strands are pulled at the other end by a force $\boldsymbol{g} = g\hat{\boldsymbol{e}}_g$ in the direction of the unit vector $\hat{\boldsymbol{e}}_g$. The system is in thermal equilibrium [4]. Other features like self-avoidance, winding and heterogeneity can be included, but are ignored in this study. Such a simple model has been found to be useful for many properties of DNA [2, 8] and resembles the model used in the analysis of the experimental data in [4]. The energy contribution from

† Though directional melting is a better nomenclature than unzipping, we use both the terms interchangeably.

0305-4470/00/450423+06$30.00

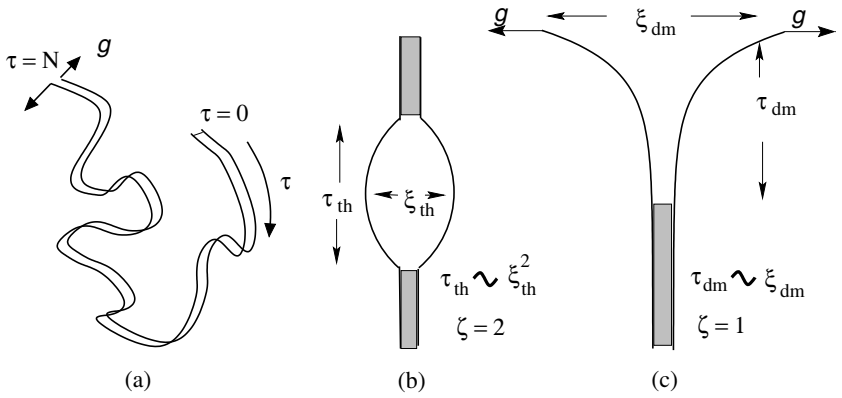

Figure 1. (*a*) Pulling the two strands at one end. (*b*) A 'denatured' bubble due to thermal fluctuation ('melting'), (*c*) an unzipped region slightly below the critical threshold ('directional melting'). In (*b*) and (*c*), shaded parts denote bound regions.

the force can be expressed in terms of the separation of the two strands each of length N. It is given by

$$-\boldsymbol{g}\cdot\boldsymbol{r}(N) = -\int_0^N \mathrm{d}\tau\, \boldsymbol{g}\cdot\frac{\partial \boldsymbol{r}(\tau)}{\partial\tau} \tag{1}$$

where $\boldsymbol{r}(\tau)$ denotes the separation of monomers on the two strands of both monomers at contour length τ along the strand and $\boldsymbol{r}(0) = \boldsymbol{0}$. For a DNA with identical base pairs the Hamiltonian can be written in the relative coordinate as [9]

$$\frac{H}{k_\mathrm{B}T} = \int_0^N \mathrm{d}\tau \left[\frac{\varepsilon}{2}\left(\frac{\partial \boldsymbol{r}(\tau)}{\partial\tau}\right)^2 - \boldsymbol{g}\cdot\frac{\partial \boldsymbol{r}(\tau)}{\partial\tau} + V(\boldsymbol{r}(\tau))\right] \tag{2}$$

where k_B is the Boltzmann constant and ε is the appropriate elastic constant. For notational simplicity, without loss of generality, we choose $\varepsilon = 1$. We take $\boldsymbol{r}(\tau)$ to be d-dimensional, $d = 3$ being of primary interest though other values of d are also considered†.

Base pairings at the same position of the two strands translate to the potential energy being given by $\int \mathrm{d}\tau\ V(\boldsymbol{r}(\tau))$ with $V(\boldsymbol{r})$ a short-range potential whose detailed form is not crucial. We choose $V(\boldsymbol{r})$ to be of square-well type. Our interest is in the free energy per unit length for long chains, $f = -\lim_{N\to\infty} k_\mathrm{B}TN^{-1}\ln Z$ where $Z = \int \mathcal{D}R \exp(-H/k_\mathrm{B}T)$ sums over all the configurations of the chains, and T is the temperature. The use of the continuum version helps us in obtaining a few exact results. A discrete version with a realistic potential can be treated numerically.

The Hamiltonian written in the above form can be thought of as a directed polymer in $d+1$ dimensions about which many results are known [10, 11] for $\boldsymbol{g} = 0$. If we treat τ as a time-like coordinate, then the same Hamiltonian represents, in the path integral formulation, a quantum particle in *imaginary* time. The quantum Hamiltonian is

$$H_q(\boldsymbol{g}) = \frac{1}{2}(\boldsymbol{p} + \mathrm{i}\boldsymbol{g})^2 + V(\boldsymbol{r}) \tag{3}$$

in units of $\hbar(\equiv k_\mathrm{B}T) = 1$ and mass $\varepsilon = 1$, with $\boldsymbol{p}$ as momentum. The ground-state energy of equation (3) determines the free energy per unit length of the DNA. This quantum particle with

† Our interest is in $d = 3$, but $d = 1$ and $d = 2$ provide interesting insights and results as examples of *low-dimensional biology*.

$\boldsymbol{g} \neq 0$ then corresponds to the imaginary vector potential problem much discussed in recent times [12]. We make use of both these pictures in this Letter.

For $d = 1$, the quantum problem with $V(x) = -v_0\delta(x)$, is exactly solvable and is done in [12] as a single-impurity problem. It was shown that there is a critical g_c below which the imaginary vector potential does not affect the bound-state energy, i.e. the quantum particle remains localized near the potential well, while for $g > g_c$ the particle delocalizes. In the polymer picture, this means that in low dimensions, a force beyond a critical strength separates the two strands. It should be pointed out here that the force is applied at one end only, but it is not a boundary or edge effect mainly because of the connectivity of the polymer chain as expressed by equation (1).

Details of the phase transition behaviour of the Hamiltonian of equation (2) for $g = 0$ are known from exact renormalization group (RG) calculations [10, 11]. With $g = 0$, it is known [10, 11] that there is a critical strength v_c ($= 0$ for $d < 2$) of the parameter $v = \int \mathrm{d}\boldsymbol{r}\ V(\boldsymbol{r})/k_BT$ such that there is a bound state of the quantum particle only for $v < v_c$. This transition corresponds to the thermal melting or unbinding of dsDNA because v can be changed by changing pH or T. Renormalization group arguments [11] show that other details of $V(\boldsymbol{r})$ are irrelevant, and such arguments also justify [9] the use of the elastic energy term in equation (2) for $N \to \infty$ even if a polymer is better represented by stiff or worm-like chains.

The important length scales for this critical point, from the bound-state side, come from the typical size of the denatured bubbles of length τ_m along the chain and ξ_m in the spatial extent (figure 1(*b*)). Close to the critical point these length scales diverge as

$$\xi_m \sim |\ \Delta u\ |^{-\nu_m} \qquad \text{and} \qquad \tau_m \sim |\ \Delta u\ |^{-\nu_{\tau,m}} \sim \xi_m^{\zeta_m} \tag{4}$$

where Δu is the deviation from the critical point, and ν_m, $\nu_{\tau,m}$ and ζ_m are the important exponents [10]

$$\nu_m = \nu_{\tau,m}/\zeta_m = 1/\ |\ d-2\ | \qquad \text{and} \qquad \zeta_m = 2. \tag{5}$$

It is this ζ (the dynamic exponent of the quantum problem) that will distinguish the new phenomenon we are trying to understand.

For $g \neq 0$, we choose a d-dimensional square-well potential $V(\boldsymbol{r}) = -V_0$ for $r < r_0$, and 0 otherwise. The well is chosen to be shallow enough to have only one bound state for the $\boldsymbol{g} = 0$ case with energy $E_0 < 0$. The non-hermitian Hamiltonian equation (3) can be connected to the hermitian Hamiltonian at $g = 0$ by

$$U^{-1}H_q(\boldsymbol{g})U = H_q(\boldsymbol{g} = 0) \qquad \text{where} \quad U = \exp(\boldsymbol{g}\cdot\boldsymbol{r}). \tag{6}$$

The wavefunctions are also related by this U-transformation so that if the transformed bound (i.e. localized) state wavefunction remains normalizable, the bound-state energy will not change. The continuum part of the spectrum will have the minimum energy $E = -g^2/2$ (the state with wavevector $\boldsymbol{k} = 0$). For the localized state at $g = 0$, the wavefunction for $r > r_0$ is $\psi_0(\boldsymbol{r}) \sim \exp(-\kappa r)$ where $\kappa = 2\sqrt{|\ E_0\ |}$. The right eigenvector for $H_q(g)$ is then $\psi_R(\boldsymbol{r}) \sim \exp(\boldsymbol{g}\cdot\boldsymbol{r} - \kappa r)$, obtained via the U-transformation. The wavefunction remains normalizable if $g < \kappa$, so that the binding energy stays the same as the $g = 0$ value until $g = g_c \equiv \kappa$. The generic form of the spectrum is shown in figure 2(*a*), which indicates a delocalization transition by tuning g. This is the unzipping transition of DNA at

$$g_c = 2\sqrt{E_0} \sim |v - v_c|^{1/|2-d|} \tag{7}$$

where we used the v-dependences [11] of E_0 close to $v_c(g = 0)$, for general d. This agrees with the exact solution of the one-impurity problem [12] (attractive δ-function potential) in $d = 1$. If one can measure g_c for varieties of double-stranded polymers, then equation (7) can be explicitly verified, with $v - v_c$ as the temperature deviation from the melting transition. The

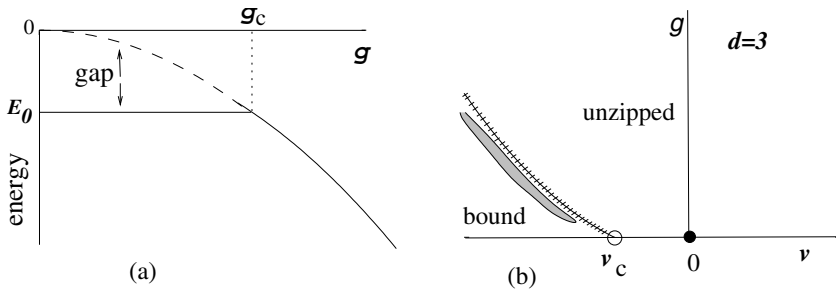

Figure 2. (*a*)The energy spectrum for $g \neq 0$ for the Hamiltonian in equation (3). (*b*) The phase diagram in the v–g plane. Thermal melting takes place along the $g = 0$ line at $v = v_c$. The hatched line indicates the unzipping transition or directional melting. The opening of DNA or initiation is hypothesized to occur in a slight sub-critical region indicated by the grey region.

phase diagram is shown schematically in figure 2(*b*). In the quantum picture, there is a gap in the spectrum (figure 2(*a*)) for $g < g_c$ and the gap vanishes continuously as $\mid g^2 - \kappa^2 \mid \sim \mid g - g_c \mid$ as $g \to g_c-$. Since time in the quantum version corresponds to the contour-length variable, the characteristic length for the unzipping transition is

$$\tau_{dm}(g) \sim \mid g_c - g \mid^{-\nu_\tau} \qquad \text{with} \quad \nu_\tau = 1. \tag{8}$$

The spatial length scale of the localized state is determined by the width of the wavefunction and, for $g = 0$, it is set by κ^{-1}. For $g \neq 0$, the right wavefunction,ψ_R, has a different length scale and this length scale diverges as the wavefunction becomes non-normalizable. The width of ψ_R gives this scale as

$$\xi_{dm}(g) \sim \mid g_c - g \mid^{-\nu_{dm}} \qquad \text{with} \quad \nu_{dm} = 1 \tag{9}$$

for $g \to g_c-$. (This needs to be distinguished from the stretching of a polymer by an external force where, for small force, the end-to-end distance of the polymer scales linearly with the force. There is neither a critical strength nor any diverging scale.) We see that, at the unzipping transition, $\tau_{dm} \sim \xi_{dm}$, and therefore

$$\zeta_{dm} \equiv \frac{\nu_{\tau,dm}}{\nu_{dm}} = 1. \tag{10}$$

The significance of τ_{dm} can be understood if we study the separation of the two chains, i.e. $\langle \boldsymbol{r} \rangle_\tau$, at a distance τ along the chain below the pulled end. This can be evaluated by using the standard rules of quantum mechanics [12]. For infinitely long chains in the sub-critical region, only the bound and the first excited states are sufficient for the computation. One finds, along the pulled direction,

$$\langle r \rangle_\tau \sim \exp[-\tau/\tau_{dm}(g)] \tag{11}$$

where $\tau_{dm}(g)$ is given by equation (8). In other words, $\tau_{dm}(g)$ and $\xi_{dm}(g)$ describe the unzipped part of the two chains near the pulled end (figure 1(*c*)). These length scales diverge with $\zeta_{dm} = 1$ as the critical force is reached from below. (One can also define similar length scales for $g \to g_c+$ where the lengths would describe the bound regions.) The exponential fall-off, equation (11), of the separation from the pulled end immediately gives the picture of a Y-fork as shown in figure 1(*c*). If we make the assumption that the action of the enzyme in opening of a dsDNA is equivalent to an exertion of a force, one might connect the Y-fork of figure 1 to the replication fork [13].

Let us study the behaviour of the free energy. Since the partition function for the Hamiltonian of equation (2) obeys a diffusion-like equation [9], the free energy $\mathcal{F} = F/k_{\mathrm{B}}T - g^2\tau/2$ satisfies the equation

$$\frac{\partial \mathcal{F}}{\partial \tau} = \frac{1}{2}\nabla^2\mathcal{F} - \frac{1}{2}(\nabla\mathcal{F})^2 - \boldsymbol{g}\cdot\nabla\mathcal{F} - v_0\delta_\Lambda(\boldsymbol{r}). \tag{12}$$

The left-hand side represents the free energy per unit length of the chains. Under a scale transformation $x \to bx$, and $\tau \to b^\zeta\tau$, the total free energy remains invariant so that the above equation takes the form

$$\frac{\partial \mathcal{F}}{\partial \tau} = \frac{1}{2}b^{\zeta-2}\nabla^2\mathcal{F} - \frac{1}{2}b^{\zeta-2}(\nabla\mathcal{F})^2 - b^{\zeta-1}\boldsymbol{g}\cdot\nabla\mathcal{F} - b^{\zeta-d}v_0\delta_\Lambda(\boldsymbol{r}). \tag{13}$$

For $g = 0$, equation (13) tells us that $\zeta = 2$ and $d = 2$ are special for the melting transition as we see in equation (5). For the choice $\zeta = 1$, the g-dependent term dominates and all other terms become irrelevant for large length scale b. It is this feature that shows up in the Y-fork of the unzipped chain. The robustness of equations (8)–(11) also follows from this. With the dependence on the potential strength entering only through g_{c}, these are valid along the hatched line of figure 2(*b*), and could be oblivious to the details of the nature of the melting transition.

From the phase diagram of figure 2, we see that for $g \neq 0$ a transition can be induced either by a force or for a given force by tuning v (i.e. temperature or pH). Our results show that the scaling remains the same as given by equations (8)–(10), except that the scaling variable is $v - v_{\mathrm{c}}(g)$ with $v_{\mathrm{c}}(g)$ determined by equation (7). The thermal melting (figure 1) occurs only in the $g = 0$ case characterized by anisotropic scaling, equation (5), of the denatured bubbles. In our simple model, the melting point appears as a multi-critical point in the phase diagram of figure 2(*b*).

The scaling behaviour near the critical force for unzipping of a homo-DNA (or other simpler double-stranded polymers) needs to be studied experimentally, either by varying the force or at a fixed force by changing the temperature or pH. It would also be important to study experimentally the nature and dynamics of activities of DNA polymerase and RNA polymerase after a mechanical unzipping of dsDNA, as a function of the force, especially in the critical region of the unzipping transition.

Although we considered only the equilibrium situation, *a criticality ensures a divergent time scale in the dynamics also*†. Therefore, in the critical region there will be long-range correlations not only in space but also in time. Heterogeneity of the base pairs can be incorporated by taking the interaction to be random [15] with a specific distribution. Such a random case [15] shows a different type of melting behaviour. Self-avoidance and winding can be added to this model by adding a random imaginary scalar potential [16], and a real vector potential [9], respectively, though such general non-hermitian Hamiltonians are little understood at present.

Let us now consider the real biological situation. A fundamental hypothesis of biology is that all activities are mediated by enzymes. The replication (and transcription) starts with the opening of the DNA at certain 'origins' (by one or more enzymes), and after that the polymerization starts at the resulting Y-fork [1]. The actual process is more involved, requiring several proteins or enzymes at various stages, and all of these work at *different points in space and time and in right order*. A question naturally arises: what produces the correlation in space and time? We make the hypothesis that the physical effect of the enzyme at the origin is to pull the strands by a slightly subcritical force (g close to g_{c}) so that a Y-fork (see equation (11)) opens-up. See figure 2. (An eye can be thought of as two connected Y's.) A coupling of

† In a recent preprint, Sebastian [14] has shown a diverging time scale for this problem.

the subsequent processes (untwisting, polymerization, etc) to the correlated critical dynamics of unzipping could be a source of the correlations we see during replication and transcription. The details of such a coupled dynamics will be discussed elsewhere [17].

To conclude, we have shown that a double-stranded DNA can be opened up by a force only if the force exceeds a critical value. This unzipping or directional melting is a genuine critical phenomenon with critical exponents different from thermal melting at zero force in the same model. Exact exponents are obtained. We suggest that experiments be done on synthetic DNA of identical base pairs (or other double-stranded homopolymers) at various forces (in a fixed force ensemble) to study the unzipping phase transition (directional melting) and get the phase diagram of figure 2. It would then be interesting to think of a mechanical unzipping with a sub-critical force and study the replication of DNA (or transcription) by the enzymatic processes in the opened fork; this would identify the correlation between unzipping and replication or transcription.

I thank ICTP, where a major part of this work was done, for warm hospitality (September–October 1999). Mere acknowledgments will probably belittle the influence of a discussion with Mathula Thangarajh that shaped the final form of this work.

References

[1] Kornberg A and Baker T 1992 *DNA Replication* 3rd edn (New York: Freeman)
[2] Wartell R M and Benight A S 1985 *Phys. Rep.* **126** 67
[3] Essevaz-Roulet B, Bockelmann U and Heslot F 1997 *Proc. Natl. Acad. Sci. USA* **94** 11935
[4] Bockelmann U, Essevaz-Roulet B and Heslot F 1998 *Phys. Rev.* E **58** 2386
[5] Viovy J L *et al* 1994 *C. R. Acad. Sci.* **317** 795
Thompson R E and Siggia E D 1995 *Europhys. Lett.* **31** 335
Bustamante C *et al* 1994 *Science* **265** 1599
Vologodskii A 1994 *Macromolecule* **27** 5623
Marko J F and Siggia E D 1995 *Macromolecule* **28** 8759
[6] Haijun Z, Yang Z and Zhong-can O-Y *Phys. Rev. Lett.* **82** 4560
[7] Cocco S and Monasson R 1999 *Preprint* cond-mat/9904277
[8] Peyard M and Bishop A R 1989 *Phys. Rev. Lett.* **62** 2755
Cule D and Hwa T 1997 *Phys. Rev. Lett.* **79** 2375
Campa A and Giansanti A 1998 *Phys. Rev.* E **58** 3585
Bonnet G, Krichevsky O and Libchaber A 1998*Proc. Natl. Acad. Sci. USA* **95** 8602
[9] Doi M and Edwards S F 1986 *The Theory of Polymer Dynamics* (Oxford: Clarendon Press)
[10] Rajasekaran J J and Bhattacharjee S M 1991 *J. Phys. A: Math. Gen.* **24** L1217
Bhattacharjee S M 1992 *Physica* A **186** 183
[11] Kolomeisky E B and Straley J P 1992 *Phys. Rev.* B **46** 12664
[12] Hatano N and Nelson D R 1996 *Phys. Rev. Lett.* **77** 570
Hatano N and Nelson D R 1997 *Phys. Rev.* B **56** 8651
Hatano N and Nelson D R 1998 *Phys. Rev.* B **58** 1998
[13] Bhattacharjee S M 1999 *Preprint* cond-mat/9912279
[14] Sebastian K L 2000 *Preprint* cond-mat/0004384
[15] Mukherji S and Bhattacharjee S M 1993 *Phys. Rev. Lett.* **70** 49
Mukherji S and Bhattacharjee S M 1993 **70** 3359 (erratum)
Mukherji S and Bhattacharjee S M 1993 *Phys. Rev.* E **48** 3483
[16] See, e.g., Izyumov A V and Simons B D 1999 *Europhys. Lett.* **45** 290
[17] Bhattacharjee S M 2000 unpublished