Multiple timescales in a model for DNA denaturation dynamics

2009 J. Phys. A: Math. Theor. 42 082003

(http://iopscience.iop.org/1751-8121/42/8/082003)

IOP FTC
IOP PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND THEORETICAL

J. Phys. A: Math. Theor. **42** (2009) 082003 (8pp) doi:10.1088/1751-8113/42/8/082003

FAST TRACK COMMUNICATION

Multiple timescales in a model for DNA denaturation dynamics

Marco Baiesi[1,2,3] **and Roberto Livi**[2,3]

[1] Instituut voor Theoretische Fysica, K. U. Leuven, B-3001, Belgium
[2] CSDC-Dipartimento di Fisica, Università di Firenze, I-50019 Sesto Fiorentino, Italy
[3] Sezione INFN Firenze, I-50019 Sesto Fiorentino, Italy

Received 25 November 2008, in final form 17 December 2008
Published 27 January 2009
Online at stacks.iop.org/JPhysA/42/082003

Abstract
The denaturation dynamics of a long double-stranded DNA is studied by means of a model of the Poland–Scheraga type. We note that the linking of the two strands is a locally conserved quantity; hence, we introduce local updates that respect this symmetry. Linking dissipation via untwist is allowed only at the two ends of the double strand. The result is slow denaturation characterized by two timescales that depend on the chain length L. In a regime up to a first characteristic time $\tau_1 \sim L^{2.15}$, the chain embodies an increasing number of small bubbles. Then, in a second regime, bubbles coalesce and form entropic barriers that effectively trap residual double-stranded segments within the chain, slowing down the relaxation to fully molten configurations, which takes place at $\tau_2 \sim L^3$. This scenario is different from the picture in which the helical constraints are neglected.

PACS numbers: 87.14.gk, 87.15.H−, 61.25.hp, 36.20.−r

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

The Watson–Crick helix is a typical form of a DNA in the cell [1]. In the laboratory, upon heating DNA molecules in a solution, one obtains a helix–coil transition called denaturation. For decades, DNA denaturation has attracted the attention of scientists because it can help to understand important biological processes: for instance, the genetic code can be accessed during transcription and replication by an opening of bubbles [1]. It is experimentally known that the fraction of molten DNA increases for increasing temperature [2]. The first theoretical description of denaturation that can account for this phenomenon was a simple model by Poland and Scheraga (PS) [3]. A model of this kind is now behind software like Meltsim [4], predicting sequence-dependent melting curves, which can then be compared with experiments. Another simple model for the DNA is due to Peyrard and Bishop (PB) [5, 6]. Also, some more detailed yet mesoscopic models have been recently proposed [7, 8], allowing for the numerical study of features that cannot be simulated in all-atoms molecular dynamics.

PS models rarely take into account the topological state of a macromolecule. All polymers that form closed rings have some conserved topological features, as long as the chains cannot break and cross each other. For instance, in circular DNA, such as the genomes of some bacteria, the number of times that the two strands twist around each other (linking number) cannot change. If this constraint is included in a PS model, the thermodynamic denaturation transition is essentially suppressed [9], unless supercoiling effects are considered [10]. The topology is not fundamental for the equilibrium properties of linear polymers, but the fact that chains cannot cross each other is clearly relevant for their dynamics. For instance, in electrophoresis the dynamics of a DNA under an electric field depends on its continuous entangling with the polymers of a gel [11]. This feature is taken into account in models of electrophoresis [11], such as the Rubinstein–Duke 'repton' model [12–14].

While the equilibrium thermodynamics of a DNA has been largely investigated, the dynamics of this macromolecule is of interest as well. The dynamics of thermal denaturation has been studied by means of PS and similar models [15–18], by using the PB model [6], and in more detailed models [7, 8]. For short DNA segments, the rates of opening found in experiments [19] can be estimated by PS models with stochastic dynamics [17]. However, being a coarse grained description, the PS model is particularly useful to study long chains.

The aim of this paper is the study of the denaturation dynamics of a long DNA in a PS model with a stochastic evolution that takes into account the helical structure of the double strand. This was not explicitly included in the standard formulation of the PS model [15, 16]. We show that a dynamics with local preservation of the linking between the chains yields a new scenario, with two timescales. The first regime is dominated by denatured bubbles diffusing into the chain from its ends (where the double chain can freely untwist). This regime ends after a time lapse that scales approximately as the square of the chain length, where the number of bubbles reaches a maximum. Then bubbles start to coalesce, with those near the chain ends that form entropic barriers trapping the helical domains still in excess. Trapping into these metastable states further slows down the denaturation process in the model, leading to thermally equilibrated denaturation only after a second timescale, which grows as the cube of the chain length.

In the PS model, each strand is represented by a chain of length L, where a site i ($1 \leqslant i \leqslant L$) stands for a local portion of the DNA. We associate each site with a segment of ten base pairs, which thus represents a complete helical turn when paired[4]. The state of the chain is stored in a Boolean array σ_i, where $\sigma_i = 1$ if the two segments at index i are paired and $\sigma_i = 0$ otherwise (see figure 1(a)). DNA conformation is thus represented by an alternation of segments of paired bases (sequences of 1's) and of open bubbles (sequences of 0's). For every $\sigma_i = 1$, there is a binding energy $\epsilon = -1$. At a temperature T, this corresponds to a Boltzmann factor $q = e^{-\epsilon/T}$. Thus, a sequence of m paired bases brings a contribution q^m to the global weight W of the configuration. A bubble formed by two complementary strands of length ℓ instead has an entropic contribution accounting for all the possible conformations of a walk of length 2ℓ: if s is the entropy per step of a walk, the constraint to form a loop yields a weight $Bs^{2\ell}\ell^{-c}$, where B is a constant factor. The exponent c can be deduced from self-avoiding walks statistics: with the excluded volume between the chains being fully taken into account [20], it is $c \approx 2.1$. Hence, the weight of a whole configuration is

$$W = (q^{m_1})\left(Bs^{2\ell_1}\ell_1^{-c}\right)\ldots(q^{m_\nu})\left(Bs^{2\ell_\nu}\ell_\nu^{-c}\right)(q^{m_{\nu+1}}), \tag{1}$$

where ν is the number of bubbles. We also set $\sigma_0 = \sigma_{L+1} = 1$, namely each end of a single strand is joined to the corresponding end of the other strand (no Y-fork is formed). Thus,

[4] Note that this choice of coarse graining is not fundamental, as we are just interested in the scaling properties of long chains and not in the microscopic details of the dynamics.

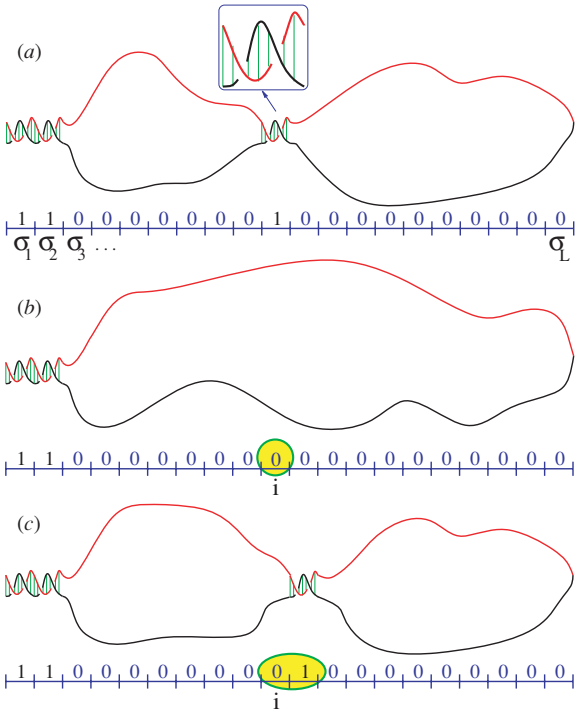

Figure 1. (*a*) Sketch of a DNA configuration and of the relative array σ. (*b*) Configuration obtained by updating (*a*) with (2) at site i. (*c*) Configuration obtained by updating (*a*) with (3) at the same site: in this case, there is local conservation of the linking number.

at high T the equilibrium configuration is a ring of length $2(L+1)$. At low T, on the other hand, typically one finds long double-stranded parts separated by small bubbles. According to this description, the properties of the model at thermodynamic equilibrium can be derived analytically [3, 20].

The simplest dynamical rules that can be assigned to the PS model involve moves where locally one σ_i changes,

$$\sigma_i = 1 \quad \longleftrightarrow \quad \sigma_i = 0. \tag{2}$$

A Metropolis criterion can then be used to choose whether to accept the move. This kind of update resembles the dynamics of adsorption of a polymer onto a wall; see figures 1(*a*), (*b*). However, for $1 \ll i \ll L$, we note that the (dis)appearance of '1' would imply either temporary breaking of one of the chains to (un)twist the two strands there (as it happens, e.g., with topoisomerase enzymes [21]) or a global rotation of 2π of the whole part $<i$ of the chain with respect to the whole part $>i$. The latter case is not in agreement with the idea of a small time step that is intrinsic in (2). Since the update (2) neglects the helix (and the consequent link) of the DNA strands, one needs another dynamics that respects the local topology of dsDNA.

In order to preserve locally the linking number, we adopt a different basic move: one picks a boundary $(i|i+1)$ at random and swaps the relative variables,

$$\sigma_i = x, \quad \sigma_{i+1} = y \quad \longrightarrow \quad \sigma_i = y, \quad \sigma_{i+1} = x, \tag{3}$$

where x and y can be 0 or 1 (if they are equal, the move is trivially the identity). This exchanges the amount of linking of the chains at position i with that at position $i+1$; see the

sketch in figure 1(*c*). In practice, the linking here plays the role of an order parameter that is locally conserved. In order to have evolution of the global linking number, however, one has to also consider some dissipation via untwisting of the double helix. In fact, untwisting with an update like (2) should be valid close to the ends of the dsDNA macromolecule, if they are free to rotate. Two special updates are then introduced by adding to the list of possible boundaries the $(0|1)$ and the $(L|L+1)$ ones:

$$\sigma_1 \to 1-\sigma_1 \qquad \text{if} \quad (0|1) \quad \text{is chosen} \tag{4}$$

$$\sigma_L \to 1-\sigma_L \qquad \text{if} \quad (L|L+1) \quad \text{is chosen.} \tag{5}$$

These two moves can thus change the energy $E = \sum_i \sigma_i$ of the chain, allowing equilibration at every temperature. It is important to note that the move (3) can lead to the nucleation of new bubbles along the whole chain, from the boundaries of already present ones (e.g., $\sigma = \ldots 00001111\ldots \to \sigma = \ldots 00010111\ldots$). The dynamics obeys detailed balance; hence, the reverse process of bubble coalescence can also clearly take place. Therefore, move (3) seems the best approximation for the purpose of describing the local conservation of the linking between two complementary DNA chains. By definition, this model cannot deal with the formation of twisted bubbles, for example by means of breaking of base–pair bonds without untwisting the chain. These configurations, however, are entropically unfavored compared to those in which the twist is concentrated on paired segments and open bubbles are expanded. We thus expect that the approximation of neglecting the formation of twisted bubbles is appropriate in a simple model.

A time step consists in a sequence of L realizations of a basic move, each one with i picked at random, uniformly along the chain. If the dynamics (3)–(5) is used, $i \in [0, L]$, while $i \in [1, L]$ for update (2). Then, according to the Metropolis criterion each move is accepted with probability $p = \min\{1, W_{\text{new}}/W_{\text{old}}\}$, where W_{new} is the weight of the proposed configuration and W_{old} is the weight of the present configuration.

The protocol on which we focus is a quench of a system equilibrated at low T to a regime at very high T. This is indeed the regime that allows us to appreciate more the effects of the new physical ingredients in this model. Equilibrated configurations to start the protocol are generated by setting $q/s^2 = 100$ and by applying multiple (2) updates[5] (this is because they equilibrate faster than (3)). Then, each protocol starts by switching instantaneously to $q/s^2 = 0.01$ at time $t = 0$. We set the parameter $B = 1$, postponing the systematic study of different cases to future works. For the bubble exponent, we use the value $c = 2.14$ [22].

The transient to the new equilibrium is monitored by studying the scaling properties of two quantities: the number of denatured pairs $\delta \equiv \sum_i (1 - \sigma_i)$ and the number of bubbles ν. The former is the quantity normally inferred from UV-absorption experiments [2] while ν is useful for characterizing the state of the system. For convenience, data are binned in time intervals with a constant size $= \log 1.05$ in a log scale. Moreover, an average over at least 1000 trajectories is performed. Figure 2 shows δ and ν versus time in a log–log scale for $L = 500$, both for a dynamics involving only move (2) and for the dynamics (3) introduced in this paper. In the former case, fast denaturation takes place, at a timescale $\tau_0 \approx 1$ that does not scale with the system size L [16].

The dynamics (3)–(5) generates a richer picture. Two characteristic timescales have been highlighted by vertical lines in figure 2. The process goes as follows: at high T, the rate $\sigma_1 = 1 \to \sigma_1 = 0$ is much higher than the rate of the opposite transition. The same is true for site $i = L$. Thus, 0's enter at the boundaries and diffuse toward the center of the chain.

[5] Similar results can be obtained by starting from the fully ordered state, $\sigma_i = 1$ for all i.

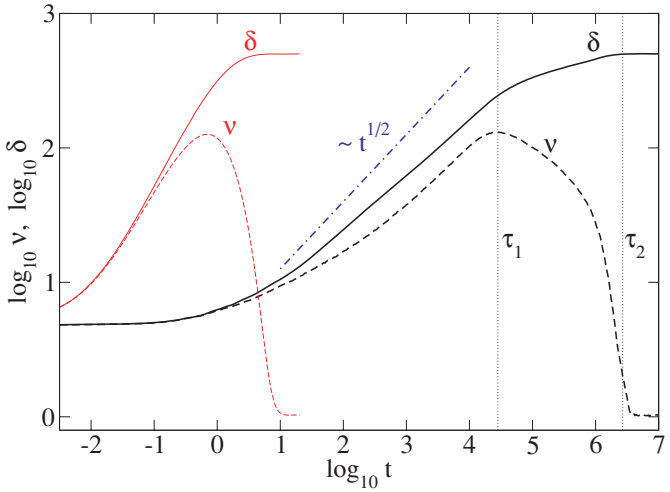

Figure 2. Log–log plot of the number of open sites δ (dense lines) and of the number of bubbles ν (dashed lines) versus time, for $L = 500$. Data shown with thick lines are obtained with our update (3)–(5), while thin lines (red) correspond to data obtained with update (2). The final state for both dynamics is the equilibrium at high T, with $\delta \approx L$ and $\nu \approx 1$. The dot-dashed line represents a scaling $\sim t^{1/2}$.

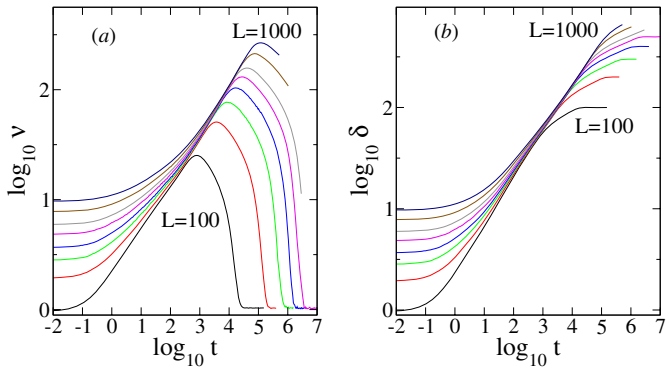

Figure 3. (*a*) Number of bubbles as a function of time, for chain lengths $L = 100, 200, 300, 400, 500, 600, 800$ and 1000 (from bottom to top). (*b*) Number of open sites versus t, the same notation.

This first regime[6] is characterized by an increase of δ and ν that is slower than $\sim t^{1/2}$, see figure 2. A timescale τ_1 is characterized by the maximum of the number of bubbles and it marks the end of the first regime. At times $t > \tau_1$ one observes a second regime in which δ continues to increase while ν decreases, which implies that bubbles coalesce. Finally, equilibrium is reached at a time τ_2, when $\delta \approx L$ and $\nu \approx 1$.

Both τ_1 and τ_2 increase with the system size; see figure 3. We find that both timescales are consistent with an algebraic dependence on L, i.e.

$$\tau_1 \sim L^{z_1}, \qquad \text{with} \quad z_1 \simeq 2.15 \pm 0.10 \tag{6}$$

$$\tau_2 \sim L^{z_2}, \qquad \text{with} \quad z_2 \simeq 3.0 \pm 0.1. \tag{7}$$

[6] In this regime, we do not observe simple dynamical scaling for δ and ν.

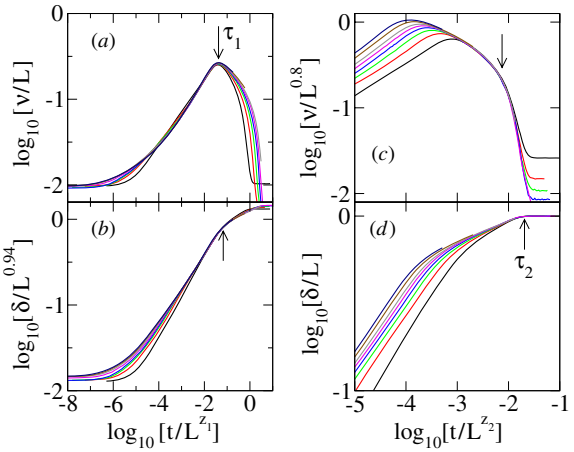

Figure 4. Curves rescaled to collapse at τ_1 with $z_1 = 2.15$ and at τ_2 with $z_2 = 3$; see the notation of figure 3 and equations (6) and (7).

The peak of the ν plots is a particularly clear feature that helps to estimate the value of z_1: peaks are reached at $\tau_1 \simeq 0.04 \times L^{2.15}$. Moreover, around τ_1 we achieve a data collapse of the form ν/L versus t/L^{z_1}. Figure 4(*a*) shows that a critical density of bubbles $\nu(\tau_1)/L$ is reached at τ_1. The values of the critical density of bubbles, as a function of $L^{-1/2}$ and extrapolated for $L \to \infty$, tend to 0.280(1). A similar collapse $\delta/L^{\alpha_1^\delta}$ versus t/L^{z_1} can be attained: the exponent yielding the best rescaling is $\alpha_1^\delta \simeq 0.94$; see figure 4(*b*).

Data collapses can be done also to determine τ_2. At τ_2 by definition we have the full denaturation, i.e. $\delta(\tau_2) \sim L$. To estimate $z_2 \simeq 3$ we have in fact required that $\delta(t)/L \to 1$ for $t \to \tau_2$ (for all L's for which τ_2 could be reached by simulations), as shown in figure 4(*d*). A data collapse of $\nu/L^{\alpha_2^\nu}$ versus t/L^3, with $\alpha_2^\nu = 0.8$ (figure 4(*c*)), confirms that $z_2 \simeq 3$ is the exponent characterizing the timescale τ_2.

The first regime is essentially a diffusion of random walkers, $\sigma_i = 0$ entering from the boundary, and one should expect a temporal domain scaling as the square of the system size. Indeed, we estimate $z_1 = 2$ for $c = 0$, a case in which the open sites are independent of each other and the weight of a configuration just depends on δ. For $c = 2.14$ we instead estimate the small deviation $z_1 \simeq 2.15$ from this classical result, probably due to the bubble weights. The case $c = 0$ is also interesting because it does not display two different timescales but only one. It confirms that the bubble interaction and coalescence is the process leading to that second timescale.

It is possible to predict the value of z_2 with a simple argument: in the regime between τ_1 and τ_2 two bubbles at the chain ends trap the double-stranded parts inside the system, preventing a fast escape of $\sigma_i = 1$ from the boundaries. Let us concentrate on one of the two ends, say $i = 1$, as shown in figure 5, where an exemplified escape of a double-stranded segment is shown. Taking the length ℓ_2 of the forming loop on the right as a reaction coordinate, the free energy $F = -\ln W$ has a profile like that shown in figure 5. It achieves a maximum at state (*c*), where $\ell_2 = \ell_1/2$. The rate of escape from (*a*) to (*e*) is proportional to the barrier jump rate, which is proportional to the ratio of the weights in (*c*) and (*a*), $[(\ell_1/2)^{-c}]^2/\ell_1^{-c} \sim \ell_1^{-c}$. Hence, the time spent for this escape scales as $(\ell_1)^c$. As this has to take place for all ℓ_1 up to the system size, $\tau_2 \sim \sum_{\ell_1 \approx 1}^{L/2} (\ell_1)^c \sim L^{c+1}$, which would imply $z_2 = c + 1 \simeq 3.14$ if the whole dynamics was as simple as that described. Of course, multiple pathways intersect and a description only in terms of a single ℓ_1 might not be exhaustive. Nevertheless, this prediction

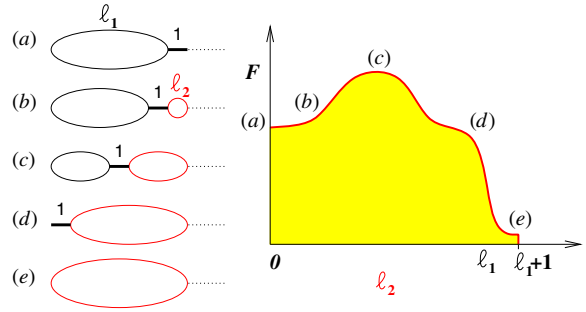

Figure 5. Sketch of the exemplified process of escape of double-stranded segments. Snapshots of some intermediate states and the corresponding free-energy profile (as a function of the length ℓ_2 of the growing loop) are shown.

is remarkably close to the value $z_2 \simeq 3$ obtained by data collapse, suggesting that this is the main mechanism acting in the second regime.

Besides timescales, we also have a time dependence of the number of open bases δ that is different from that predicted by previous models, where one can observe a linear increase of δ with time [6, 15] or $\delta \sim t^{3/4}$ [15], under conditions similar to those discussed in this paper (initial state at low T, denaturation at high T, no external forces). In our model, on the other hand, we observe that δ increases slower than the square root of time.

The fact that polymer chains cannot cross each other is at the basis of our version of the PS model, but of course it is included in many other models, like the model of a polymer diffusing into a gel by de Gennes [23], in which he found that the diffusion constant scales as $1/L^2$. Furthermore, in simulations of polymers in dense melts [24] one observes autocorrelation times scaling as L^3. These long timescales derived from the reptation dynamics of the polymers, which have to diffuse into tubes formed by the melt. We argue that scenarios like this one are similar to the phenomenon predicted by our DNA model, because the two twisted strands constrain the stochastic movements of each other in space during melting.

Acknowledgments

We acknowledge useful discussions with E Orlandini, A Kabakçıoğlu, P De Los Rios, A Flammini, F Piazza, C Maes, E Carlon and G T Barkema. We also thank E Carlon for his useful comments on the manuscript. MB acknowledges support from EC FP6 project 'EMBIO' (EC contract no. 012835) and from KU Leuven grant OT/07/034A.

References

[1] Alberts B *et al* 2002 *Molecular Biology of the Cell* (New York: Gerland Science)
[2] Wartell R M and Benight A S 1985 *Phys. Rep.* **85** 67
[3] Poland D and Scheraga H 1966 *J. Chem. Phys.* **45** 1464
[4] Blake R D *et al* 1999 *Bioinformatics* **15** 370
[5] Peyrard M and Bishop A R 1989 *Phys. Rev. Lett.* **62** 2755
[6] Barbi M, Lepri S, Peyrard M and Theodorakopoulos N 2003 Thermal denaturation of a helicoidal DNA model *Phys. Rev.* E **68** 061909
[7] Drukker K, Wu G and Schatz G C 2001 Model simulations of DNA denaturation dynamics *J. Chem. Phys.* **114** 579–90
[8] Knotts IV T A, Rathore N, Schwartz D C and de Pablo J J 2007 A coarse grain model of DNA *J. Chem. Phys.* **126** 084901

[9] Rudnick J and Bruinsma R 2002 Effects of torsional strain on thermal denaturation of DNA *Phys. Rev.* E **65** 030902(R)
[10] Kabakçıoğlu A, Orlandini E and Mukamel D 2008 Supercoil formation in DNA denaturation arXiv:0811.3229
[11] van Heukelum A and Barkema G T 2002 Lattice models of DNA electrophoresis *Electrophoresis* **23** 2562–8
[12] Rubinstein M 1987 *Phys. Rev. Lett.* **59** 1946–9
[13] Duke T A J 1989 *Phys. Rev. Lett.* **62** 2877–80
[14] Carlon E, Drzewinski A and van Leeuwen J M J 2001 Crossover behavior for long reptating polymers *Phys. Rev.* E **64** 010801
[15] Marenduzzo D, Bhattacharjee S M, Maritan A, Orlandini E and Seno F 2002 Dynamical scaling of the DNA unzipping transition *Phys. Rev. Lett.* **88** 028102
[16] Kunz H, Livi R and Süto A 2007 The structure factor and dynamics of the helix-coil transition *J. Stat. Mech.* **2007** P06004
[17] Hanke A and Metzler R 2003 Bubble dynamics in DNA *J. Phys. A: Math. Gen.* **36** L473
[18] Novotny T, Pedersen J N, Ambjornsson T, Hansen M S and Metzler R 2007 Bubble coalescence in breathing DNA: two vicious walkers in opposite potentials *Europhys. Lett.* **77** 48001
[19] Altan-Bonnet G, Libchaber A and Krichevsky O 2003 *Phys. Rev. Lett.* **90** 138101
[20] Kafri Y, Mukamel D and Peliti L 2000 *Phys. Rev. Lett.* **85** 4988
[21] Dean F B, Stasiak A, Koller T and Cozzarelli N R 1985 *J. Biol. Chem.* **260** 4975
[22] Baiesi M, Carlon E, Kafri Y, Mukamel D, Orlandini E and Stella A L 2003 Inter-strand distance distribution of DNA near melting *Phys. Rev.* E **67** 021911
[23] de Gennes P G 1989 *J. Chem. Phys.* **91** 3252–7
[24] Paul W, Binder K, Heermann D W and Kremer K 1991 *J. Chem. Phys.* **95** 7726–40