

# Single DNA denaturation and bubble dynamics

Ralf Metzler<sup>1</sup>, Tobias Ambjörnsson<sup>2</sup>, Andreas Hanke<sup>3</sup> and Hans C Fogedby<sup>4,5</sup>

<sup>1</sup> Physics Department, Technical University of Munich, James Franck Strasse, 85747 Garching, Germany

<sup>2</sup> Chemistry Department, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

<sup>3</sup> Department of Physics and Astronomy, University of Texas, 80 Fort Brown, Brownsville, USA

<sup>4</sup> Department of Physics and Astronomy, University of Århus, Ny Munkegade, 8000 Århus C, Denmark

<sup>5</sup> Niels Bohr Institute, Blegdamsvej 17, 2100 København Ø, Denmark

E-mail: [metz@ph.tum.de](mailto:metz@ph.tum.de)

Received 6 July 2008, in final form 20 August 2008

Published 17 December 2008

Online at [stacks.iop.org/JPhysCM/21/034111](http://stacks.iop.org/JPhysCM/21/034111)

## Abstract

While the Watson–Crick double-strand is the thermodynamically stable state of DNA in a wide range of temperature and salt conditions, even at physiological conditions local denaturation bubbles may open up spontaneously due to thermal activation. By raising the ambient temperature, titration, or by external forces in single molecule setups bubbles proliferate until full denaturation of the DNA occurs. Based on the Poland–Scheraga model we investigate both the equilibrium transition of DNA denaturation and the dynamics of the denaturation bubbles with respect to recent single DNA chain experiments for situations below, at, and above the denaturation transition. We also propose a new single molecule setup based on DNA constructs with two bubble zones to measure the bubble coalescence and extract the physical parameters relevant to DNA breathing. Finally we consider the interplay between denaturation bubbles and selectively single-stranded DNA binding proteins.

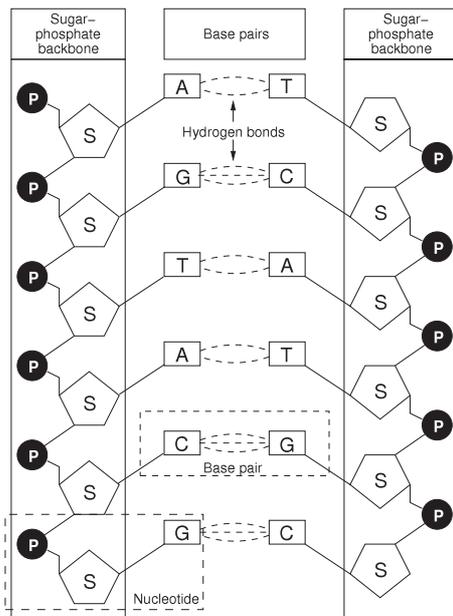
(Some figures in this article are in colour only in the electronic version)

## 1. Introduction

Deoxyribonucleic acid (DNA) is the molecule of life, encoding the complete genetic information of an entire organism. This information is kept in terms of the four letter alphabet comprised by adenine, guanine, cytosine, and thymine. The genetic code is stabilized by base pairing through hydrogen bonding that creates two complementary strands subject to the key-lock principle. This way it is made sure that exclusively AT and GC nucleotides pair. Within this ladder structure (figure 1) the bases and thus the genetic code are protected against unwanted action of chemicals and proteins. The three-dimensional structure of DNA is the famed Watson–Crick double-helix, the equilibrium structure of DNA within a broad range of salt and temperature conditions. Sufficiently close to physiological conditions the typical conformation of double-

stranded DNA is the B form with a pitch of 3.4 Å between successive base pairs and approximately 10.5 base pairs needed to form one complete turn of the helix. This thermodynamic stability, apart from hydrogen bonding between paired bases, is mainly effected by base stacking between nearest neighbour pairs of base pairs [1–6].

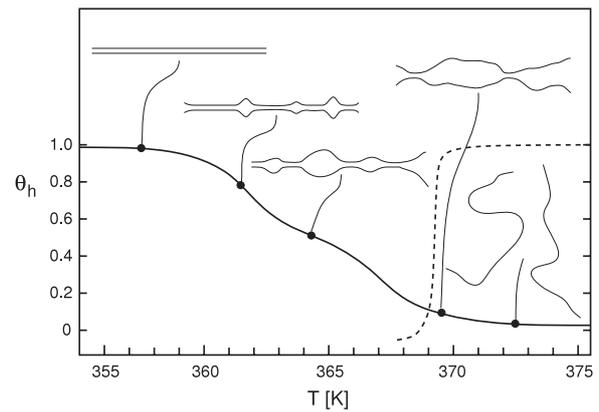
By temperature increase or variation of the pH (titration with acid or alkali) the double-stranded DNA progressively denatures. The comparatively stiff DNA double-strand (persistence length about 50 nm) is thereby interrupted by emerging zones of the flexible single-strand (persistence length about 1 to a few nm). These so-called DNA bubbles then grow and merge until the double-strand is fully molten (figure 2). This is the helix–coil transition. The melting temperature  $T_m$  is experimentally defined as the temperature at which half of the DNA molecule has undergone denaturation [3, 7, 8].



**Figure 1.** Ladder structure of DNA showing the Watson–Crick bonding of the bases A, T, G, and C which are suspended by a sugar-phosphate backbone. Each phosphate carries a negative charge. The longitudinal distance between adjacent base pairs is 3.43 Å while approximately 10.5 base pairs are needed to form a complete helical turn. Under normal salt conditions the persistence length of double-stranded DNA is approximately 50 nm, the hard core diameter is approximately 2 nm. Locally (i.e., for lengths shorter than the persistence length), DNA appears thin and stiff while on longer scales it can be perceived as a flexible polymer. The length of a single DNA molecule varies from several μm of viral DNA over several mm in bacteria up to many cm in eukaryotic cells.

Typically, the denaturation starts in regions rich in the weaker AT base pairs, and subsequently moves to zones of increasing GC content. The occurrence of zones of different stability within the genome was shown to be relevant when separating coding from non-coding regions [9, 10] and is believed to be related to DNA function, for instance the occurrence of weak regions (e.g. the TATA motif) at transcription initiation points.

Albeit rare, already at room temperature thermal fluctuations cause opening events of small intermittent denaturation bubbles [11]. The size of these bubbles fluctuates by step-wise zipping and unzipping of the base pairs at the zipper forks where the bubble connects to the intact DNA double-strand (bubble breathing). Initiation of a bubble in a stretch of the double-strand requires the crossing of a free energy barrier  $F_s$  of approximately 8 kcal mol<sup>-1</sup> (some 10  $k_B T$  at physiological temperature) corresponding to a Boltzmann factor  $\sigma_0 = \exp(-F_s/k_B T) \sim 10^{-5 \dots -3}$ .  $\sigma_0$  is often referred to as the cooperativity factor. Once formed below the melting temperature  $T_m$  a bubble will eventually zip close. Above  $T_m$ , a bubble will preferentially stay open and, if unconstrained, grow in size until it merges with other denaturation bubbles, eventually leading to full denaturation. Constraints against such full unzipping could, for instance, be the build-up of twist in smaller DNA-rings [12], the highly positively supercoiled state (linking excess) in the DNA of extremophile bacteria existing at high temperatures in deep-



**Figure 2.** Thermal denaturation of double-stranded DNA: fraction  $\theta_h$  of double-helical domains within the DNA as a function of temperature. Schematic representation of  $\theta_h(T)$ , showing the increased formation of bubbles and unzipping from the ends, until full denaturation has been reached. Note that bacterial DNA is predominantly circular so that no end effects occur. Also viral DNA circularizes once injected into a host cell.

sea vents [13], or the chemical connection of the two strands by short bulge-loops, compare [14]. In heteropolymer DNA mechanical stretching experiments show that even at the end of the overstretching transition and beyond, the two strands do not separate completely [15–17] but are still held together by isolated GC-rich regions along the chain with average distance of a few hundreds of base pairs [16]. These GC-rich regions break only at a much larger force than the melting force  $F_m$  of the overstretching plateau.

Biologically the physical conformations of DNA molecules are recognized to be of inalienable relevance for its function, see, for instance, the review [18] and references therein. In particular, the existence of intermittent though infrequent bubble domains is important as the opening up of the Watson–Crick base pairs by breaking of the hydrogen bonds between complementary bases disrupts the helical stack. The associated flipping out of the ordered stack of the unpaired bases allows the binding of specific chemicals or proteins, that otherwise would not be able to access the reactive sites of the bases [3, 6, 7, 11]. Indeed there exists a competition of timescales between the survival of DNA bubbles and the binding kinetics of selectively single-stranded DNA binding proteins [19–21]. An important aspect to the biological function of DNA is believed to be that DNA breathing assists transcription initiation [22–25], see below. Altogether it appears fair to say that the quantitative knowledge of the energetics of the denaturation as well as the dynamics of bubbles is imperative to a better understanding of genomic biochemical processes. Additionally DNA denaturation is a fine example of a well defined and chemically stable system whose physical properties can be probed in detail on the level of single molecules. DNA is therefore studied from both viewpoints: biological physics with respect to DNA’s role in biochemical processes and statistical physics for which DNA provides an ideal system to study quantitatively polymer models.

In what follows we will base our analysis on the Poland–Scheraga free energy model treating the DNA molecule as

an Ising-type system of a sequence of ‘spins’ with open or closed states, plus a non-local term that takes care of polymeric effects within denaturation bubbles made up of highly flexible DNA single-strand. A prominent alternative description of DNA denaturation and breathing is the Peyrard Bishop Dauxois (PBD) model [26, 27] based on the set of Langevin equations [28]

$$m \frac{d^2 y_n}{dt^2} = - \frac{dV(y_n)}{dy_n} - \frac{dW(y_{n+1}, y_n)}{dy_n} - \frac{dW(y_n, y_{n-1})}{dy_n} - m\gamma \frac{dy_n}{dt} + \xi_n(t). \quad (1)$$

Here,  $V(y_n) = D_n[\exp(-a_n y_n) - 1]^2$  is a Morse potential for the hydrogen bonding,  $D_n$  and  $a_n$  assuming two different values for AT and GC bps;  $W(y, y') = \frac{k}{2}[1 + \rho \exp\{-\beta(y + y')\}](y - y')^2$  is a nonlinear potential to include bp–bp stacking interactions between adjacent bps  $y$  and  $y'$ . The parameters  $k$ ,  $\rho$ ,  $\beta$ ,  $\gamma$ , and  $m$  are invariants of the sequence. The equation is driven by the thermal noise  $\xi_n(t)$ . Usually, the stochastic equations (1) are integrated numerically [28]. There is also a helicoidal version of the Peyrard–Bishop model to study the torque-induced denaturation of DNA [29, 30]. Due to its formulation in terms of a set of Langevin equations, the DPB model is very appealing, and it is a useful model to study some generic features of DNA denaturation. Its slight disadvantage is that somewhat arbitrary values for the model parameters need to be chosen while (apart from the characteristic timescale) all parameters in the Poland–Scheraga model are available from a large body of experiments.

We first address the denaturation transition at equilibrium both in the absence and presence of an external stretching force. Subsequently we will present two model approaches to the breathing dynamics of a single denaturation bubble. In section 4 we discuss the coalescence dynamics of two DNA bubbles. Finally, in section 5 we address the coupling of the breathing dynamics of a DNA bubble with the binding/unbinding of proteins that specifically bind to single-stranded DNA.

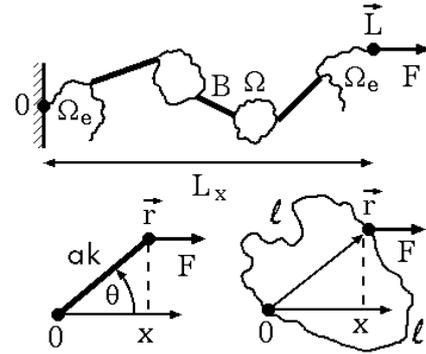
## 2. DNA denaturation in presence of a modest stretching force

A convenient method to treat the denaturation transition is to consider the chain in the grand canonical ensemble in which the total number  $N$  of bps and the end-to-end vector  $\mathbf{L}$  fluctuate. The partition function in  $d = 3$  of the DNA chain under external forcing with force  $F$  in  $x$  direction becomes [31]

$$\mathcal{Z}(z, F) = \sum_{N=1}^{\infty} \int d^3 L \mathcal{Z}_{\text{can}}(N, \mathbf{L}) z^N \exp(\beta F L_x) \quad (2)$$

with  $\beta = 1/(k_B T)$ .  $\mathcal{Z}_{\text{can}}(N, \mathbf{L})$  is the canonical partition function of a chain of  $N$  bps with fixed end-to-end vector  $\mathbf{L}$ ,  $z$  is the fugacity, and  $L_x$  the  $x$ -component of  $\mathbf{L}$  (figure 3). Assuming that bound segments and bubbles are independent,  $\mathcal{Z}$  factories:

$$\mathcal{Z}(z, F) = \Omega_e + \Omega_e \left\{ \sum_{n=0}^{\infty} [B\Omega]^n \right\} B\Omega_e = \Omega_e + \frac{\Omega_e^2 B}{1 - \Omega B}. \quad (3)$$



**Figure 3.** Stretched DNA in the PS model with bound segments  $B$  and denatured loops  $\Omega$ . The DNA is attached between  $\mathcal{O}$  and  $\mathbf{L}$  and subject to the stretching force  $F$  in the  $x$ -direction. Perfect matching in heterogeneous DNA requires both arches of a loop to have equal length  $\ell$ .

The alternating sequence of bound segments and bubbles with weights  $B$  and  $\Omega$  in equation (3) is complemented by the weight  $\Omega_e$  of an open end unit at both ends of the chain. We assume that only one strand of the end unit is bound to the, say, magnetic bead, while the other strand is moving freely. Thus the first term on the right hand side of equation (3) denotes the two unbound single strands of completely denatured DNA; here we assume that one of the two strands is still attached between origin  $\mathcal{O}$  and end point  $\mathbf{L}$ , being subject to the stretching force  $F$ .

A bound segment with  $k = 1, 2, \dots$  bps is modelled as a rigid rod of length  $ak$  where  $a = 0.34$  nm is the length of a bound bp in B-DNA [32]. Here we assume a homopolymer with binding energy  $E_0 < 0$  per base pair. However, we assume perfect matching throughout the transition such that in a denaturation bubble both single-stranded arches carry equal length. This assumption is in line with the above remark that due to stable GC-rich islands in the structure during a force-induced denaturation the sequence of separated denaturation bubbles and intact double-strand persists to much larger forces than the melting force  $F_m$  of the overstretching plateau [16]. The statistical weight of a segment with fixed number  $k$  and fixed orientation is then  $\omega^k$  with  $\omega = \exp(\beta \epsilon)$  and  $\epsilon = -E_0 > 0$ . Assuming that  $k$  fluctuates with fixed fugacity  $z$ , and rotates around one end while subject to the force  $F$  (figure 3), the statistical weight of the segment for fixed  $z$  and  $F$  becomes [31]

$$B(z, \omega, F) = \frac{1}{2y} \ln \left( \frac{1 - \omega z e^{-y}}{1 - \omega z e^y} \right), \quad y \equiv \beta F a. \quad (4)$$

At  $F = 0$ ,  $B(z, \omega, 0) = \omega z / (1 - \omega z)$  as found previously for free DNA [33].

A denatured loop is considered as a closed random walk with  $2\ell$  monomers, corresponding to  $\ell$  broken bps. The loop starts at  $\mathcal{O}$  and visits the point  $\mathbf{r}$  after  $\ell$  monomers (figure 3). The number of configurations of such a loop becomes

$$\Omega(\ell, \mathbf{r}) = C_0(2\ell) p_\ell(\mathbf{r}) \quad (5)$$

where  $C_0(2\ell)$  counts the configurations of a loop of length  $2\ell$  starting at  $\mathcal{O}$  and  $p_\ell(\mathbf{r})$  is the probability that the loop

visits  $\mathbf{r}$  after  $\ell$  monomers. For an ideal random walk in  $d = 3$ ,  $C_0(2\ell) \sim \mu^{2\ell} \ell^{-3/2}$  ( $\mu$  is the connectivity constant<sup>6</sup>) and  $p_\ell(\mathbf{r}) \sim \mathcal{R}^{-3} \exp[-\lambda(r/\mathcal{R})^2]$  where  $\lambda > 0$ ,  $r = |\mathbf{r}|$ ,  $\mathcal{R} = b\ell^{1/2}$  is the scaling length of the walk and  $b$  its persistence length. The coefficient  $b$  is proportional to the persistence length. Thus,  $\Omega(\ell, \mathbf{r}) \sim s^\ell \ell^{-3} \exp[-\lambda(r/\mathcal{R})^2]$  where  $s = \mu^2$ . We assume that  $\mathbf{r}$  moves freely and is subject to the force  $F$  in the positive  $x$ -direction. The weight of an ideal random loop for fixed  $\ell$  and  $F$  is given by the Gaussian integral

$$\Omega(\ell, F) = \int d^3r \Omega(\ell, \mathbf{r}) e^{\beta Fx} = A s^\ell \ell^{-c} \exp(\alpha y^2 \ell) \quad (6)$$

where  $A$  is an amplitude proportional to the cooperativity factor  $\sigma_0$ ,  $c = 3/2$ , and  $\alpha = b^2/(4\lambda a^2)$ .

Generally, the loop free energy in both presence and absence of the external forcing is of the power-law form

$$\Omega \simeq \ell^{-c}. \quad (7)$$

For free DNA it was found that the nature of the denaturation transition is determined by the value of the critical exponent  $c$ : for  $c \leq 1$  there is no phase transition in the thermodynamic sense; for  $1 < c \leq 2$  the transition is second order, and for  $c > 2$  it is first order [33–35]. One finds  $c = 3/2 < 2$  if the loops are ideal random walks. Self-avoiding interactions within a loop modify this value to  $c = 3\nu = 1.76$  with  $\nu = 0.588$  in  $d = 3$ . In both cases the transition is second order. Including self-avoiding interactions between denatured loops and the rest of the chain was found to produce  $c = 2.12 > 2$ , driving the transition to first order [9, 33, 36]. These results suggest that the inclusion of self-avoiding interactions generally shifts the loop exponent  $c$  to larger values, possibly effecting a change of the transition from second to first order.

Using scaling arguments in the presence of self-avoiding interactions within a loop we find a modified expression for the statistical weight [31]

$$\Omega(\ell, F) = A s^\ell \ell^{-c} y^{1/(2\nu)-1} \exp(\alpha y^{1/\nu} \ell) \quad (8)$$

for  $\kappa = \beta b F \ell^\nu \rightarrow \infty$  and with the new loop exponent in  $d = 3$ ,

$$c = 4\nu - 1/2 = 1.85. \quad (9)$$

Thus, in the presence of self-avoiding interactions within a denatured loop and  $F > 0$  the transition remains second order, but moves closer to first order compared to free DNA (with  $c = 3\nu = 1.76$  obtained within the same approach). In the Gaussian limit the same result obtains as in the absence of the force corresponding to the ideal Hookean chain behaviour of a phantom chain.

Within this formalism it is also possible to obtain the force–extension behaviour of the chain as well as the temperature–force phase diagram, see figure 4 [31, 33]<sup>7</sup>. The

<sup>6</sup> For an ideal chain embedded in  $d$ -dimensional space  $\mu = 2d$  while for a self-avoiding walk the connectivity constant becomes reduced compared to that value. On average  $\mu \approx 4.68$  in  $d = 3$  [37].

<sup>7</sup> To obtain the phase diagram shown in figure 4 we ignore the singularity of  $\Omega_e$  in equation (3) and only consider the singularity that occurs if the denominator  $1 - \Omega B$  approaches zero. The reason is that for heteropolymer DNA even at the end of the overstretching transition and beyond the two strands do not separate completely but are still held together by isolated, GC-rich islands along the chain. Thus the size of the end units  $\Omega_e$  is bounded by the first GC-rich region at either end of the DNA and the statistical weight  $\Omega_e$  cannot diverge.

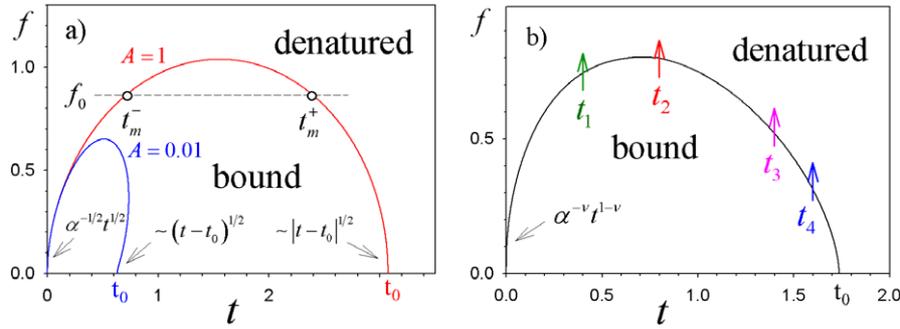
shape of the transition line  $f_m(t)$  depends on  $A$ ,  $\alpha$ , and  $s$ . Figure 4(a) shows  $f_m(t)$  for  $A = 1$ ,  $\alpha = 1$ , and  $s = 5$  for the case that denatured loops are ideal random walks ( $\theta = 0$ ,  $\nu = 1/2$ ). The transition line for a more realistic value  $A \ll 1$  is also shown (here  $A = 0.01$ ).<sup>8</sup> The line  $f_m(t)$  separates a finite region of bound states from an infinite region of denatured states. The point  $(t_0, f = 0)$  with  $t_0 = t_m(f = 0)$  corresponds to the traditional melting transition for free DNA ( $F = 0$ ). The line  $f_m(t)$  for  $A = 1$  contains a region in which  $f_m(t)$  decreases with  $t$ , such that increased stretching forces  $f$  lower the melting temperature  $t_m(f)$ , corresponding to force-induced destabilization of DNA [32]. Interestingly, for  $A = 0.01$  the line  $f_m(t)$  is not single-valued. Moreover,  $f_m(t)$  vanishes for both  $t \rightarrow t_0$  (as  $|t - t_0|^{1/2}$ ) and  $t \rightarrow 0$  (as  $\alpha^{-1/2} t^{1/2}$ ). This ‘reentrant behaviour’ [38] means that for given  $0 < f_0 < f_{\max}$ , where  $f_{\max}$  is the maximum of  $f_m(t)$ , the chain does not only denature at a large  $t_m^+(f_0)$  but also at a small  $t_m^-(f_0)$ . This behaviour can be traced back to a balance of the terms  $(\beta F a)^2$  and  $\beta F a$  in  $z_m(F) = \exp(-\alpha y^2)/s$  and equation (4), respectively. For  $(\beta F a)^2 \ll \beta F a$ , i.e.,  $k_B T \gg F a$ , the melting transition at  $t_m^+(f_0)$  is mainly driven by the entropy gain on creation of fluctuating loops, similar as for free DNA. For  $k_B T \ll F a$  the transition at  $t_m^-(f_0)$  is due to the fact that  $B[z_m(F), \omega, F]$  decreases with  $y = \beta F a = f/t$  in the denatured state, due to the rapid decay of  $z_m(F)$  (cf equation (4))<sup>9</sup>. Figure 4(b) shows the line  $f_m(t)$  for self-avoiding loops with  $c = 1.85$  demonstrating analogous behaviour.

At very high forces corrections to this treatment are expected. However the fact that already at moderate (in fact, any positive) external force  $F$  the value of the critical exponent  $c$  changes indicates that the force-induced denaturation employed in single molecule experiments is physically different from thermal denaturation. This is intuitively clear as the pulling alters not only the free energy of intact base pairs but also the number of accessible degrees of freedom of the polymer loops forming the denaturation bubbles. Recently, the Poland–Scheraga model for stretched DNA was extended to model both the double-stranded segments and the single-stranded segments forming a bubble as freely jointed chains, including distinct bending rigidities for both types of segments [39]. Phase diagrams and force–extension curves were obtained by generalizing the matrix technique for single persistent chains to describe the branching bubbles. The authors found reentrance behaviour similar to that observed in figure 4 in that the DNA may be bound at intermediate values of the force but melts at both weak and strong force [39].

We here treat the DNA denaturation in presence of the external stretching force in analogy to thermal denaturation. The transition, that is, goes from the double-stranded state to the fully denatured single-strand. While this view is

<sup>8</sup> The amplitude  $A$  is proportional to the cooperativity factor  $\sigma_0 \ll 1$ .

<sup>9</sup> This relies on the assumption that  $p_\ell(\mathbf{r})$  is Gaussian for ideal random loops and as described above equation (8) for self-avoiding loops. For very large  $\beta F$  denatured loops are stretched out and aligned along  $F$  so that the partition function is dominated by parameter values for which  $p_\ell(\mathbf{r})$  deviates from this form. A suitable  $p_\ell(\mathbf{r})$  should be used to obtain the phase diagram in this regime.



**Figure 4.** Transition lines  $f_m = F_m a / \epsilon$  as function of  $t = k_B T / \epsilon$  for  $\alpha = 1, s = 5$  for denatured loops modelled as (a) ideal random walks and (b) self-avoiding walks. Note the reentrant behaviour at lower temperatures where the required melting force decreases.

in accord with a large body of experiments [19, 40–42] and theoretical approaches [32, 43, 44] as well as recent simulations [45] one cannot exclude the possibility that an intermediate state of DNA exists, so-called S-DNA. A number of recent contributions address this question [39, 44, 46–49] but for now this point remains unresolved.

### 3. Single DNA bubble dynamics

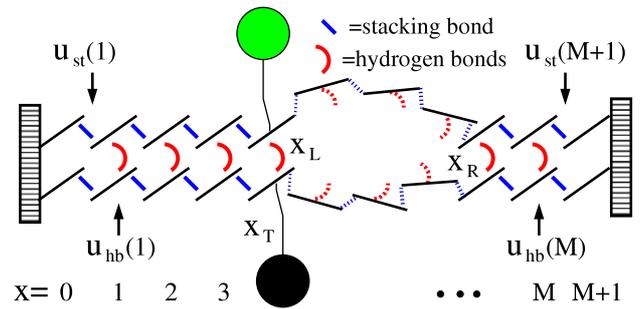
Below the melting temperature  $T_m$ , DNA bubbles are intermittent, i.e., they form spontaneously due to thermal fluctuations and after some time close again. DNA breathing can be thought of as a biased random walk in the phase space spanned by the bubble size  $m$  and its position denoted, e.g., by the left zipper fork position  $x_L$  [24, 25]. The bubble creation can be viewed as a nucleation process, whereas the bubble lifetime corresponds to the survival time of the first passage problem of relaxing to the  $m = 0$  state after a random walk in the  $m > 0$  half-space [24, 25, 50–52]. Apart from NMR techniques [6, 11] bubble breathing could be measured on the single DNA bubble level by fluorescence correlation spectroscopy [14]. This technique employs a designed stretch of DNA, in which weaker AT bps form the bubble domain, that is clamped by stronger GC bonds. In the bubble domain, a fluorophore–quencher pair is attached, see figure 5. Once the bubble is created, fluorophore and quencher are separated, and fluorescence occurs.

#### 3.1. Continuum approach for homopolymer DNA

Originally bubble breathing was considered in a random energy model with scaling arguments and numerical solution [53] and for a homopolymer by mapping on a Fokker–Planck equation for a random walker in the bubble free energy landscape with approximate analytical and numerical solution [50]. An analytical approach to bubble breathing in a homopolymer DNA with explicit solution for the distribution of bubble lifetimes is indeed possible by mapping onto the quantum Coulomb problem [54, 55] as we discuss here. In the following subsection we consider explicitly given DNA sequences in a discrete approach.

The Poland–Scheraga free energy for a single bubble has the continuum form [50, 54]

$$\mathcal{F} = \gamma_0 + \gamma x + c k_B T \ln x \quad (10)$$



**Figure 5.** Clamped DNA domain with internal bps  $x = 1-M$ , statistical weights  $u_{hb}(x), u_{st}(x)$ , and tag position  $x_T$ . The DNA sequence enters through the statistical weights  $u_{st}(x)$  and  $u_{hb}(x)$  for disrupting stacking and hydrogen bonds respectively. The bubble breathing process consists of the initiation of a bubble and the subsequent motion of the forks at positions  $x_L$  and  $x_R$ . See [25] for details.

in terms of the bubble size  $x \geq 0$ . Expression (10) corresponds to a logarithmic sink in  $\mathcal{F}$  at  $x = 0$  proportional to the loop exponent  $c$ . To be consistent with the notation in [50, 54] we choose the bubble initiation free energy  $\gamma_0 = -k_B T \log \sigma_0$  (i.e., equal to the boundary energy  $F_s$  mentioned earlier) instead of the cooperativity factor  $\sigma_0$  and the free energy  $\gamma = -k_B T \log(u_{hb} u_{st})$  to break a base pair in an already existing loop made up of the Boltzmann factor for the stacking free energy  $u_{st}$  and the corresponding factor for the hydrogen bonding  $u_{hb}$  used in the discrete model below. We recognize from equation (10) that a characteristic bubble size is set by  $x_1 = c k_B T / |\gamma|$ . We rewrite the free stacking energy in terms of  $\gamma \equiv \gamma_1 (T_m - T) / T_m$  through the melting temperature  $T_m$ , and similarly, we introduce  $\epsilon = \gamma_1 / [2k_B] (T_m^{-1} - T^{-1})$ .

For large bubble size  $x > x_1$  the linear term dominates and the free energy grows as  $\mathcal{F} \sim \gamma_0 + \gamma x$ . For small bubbles  $x < x_1$  (or close to  $T_m$ , where  $\gamma(T) \approx 0$ ) the free energy is characterized by the logarithmic sink but has, strictly speaking, a minimum at  $\mathcal{F} = \gamma_0$  for zero bubble size. We distinguish two temperature ranges: (i) For  $\gamma < 0$ , i.e.,  $T > T_m$ ,  $\mathcal{F}$  has a maximum  $\mathcal{F}_{\max} = \gamma_0 + c k_B T (\log x_1 - 1)$  at  $x = x_1$ . The free energy profile thus defines a Kramers escape problem in the sense that an initial bubble can grow in size corresponding to the complete denaturation of the double-stranded DNA. The

escape probability  $P_{\text{esc}} \propto \exp(-\Delta\mathcal{F}/k_B T)$ , where the free energy barrier is  $\Delta\mathcal{F} = ck_B T(\log x_1 - 1)$ . Thus

$$P_{\text{esc}} \propto \left( \frac{ck_B T}{|\gamma|} \right)^{-c} \quad (11)$$

has a power-law dependence on temperature typical for entropic barriers. In contrast a Kramers escape across a high energetic barrier leads to an Arrhenius behaviour. An example for the latter would be the initiation process of a bubble during which the barrier  $\sigma_0 = \exp(-\beta F_s)$  needs to be crossed.

(ii) For  $\gamma > 0$ , i.e.,  $T < T_m$ , the free energy increases monotonically from  $\mathcal{F} = \gamma_0$  at  $x = 0$  and the finite size bubbles are stable. The change of sign of  $\gamma$  at  $T = T_m$  thus defines the bubble melting.

The gradient of the free energy profile then enters as a force term in a Langevin equation for the bubble size  $x$ . Such a treatment is possible since  $x$  is the slow variable of the system compared to the polymeric degrees of freedom of a bubble and even the entire chain unless the chain size becomes too large. The Langevin equation can then be mapped onto the Fokker-Planck equation for the probability density  $P(x, t)$  to find a bubble of size  $x$  at time  $t$ :

$$\frac{\partial P(x, t)}{\partial t} = Dk_B T \frac{\partial}{\partial x} \left( \left[ \frac{c}{x} - \epsilon \right] + \frac{\partial^2}{\partial x^2} \right) P(x, t). \quad (12)$$

Here  $D$  is the noise strength of the thermal environment measured in units of  $k_B T$  and time. It is now the task to derive from this dynamical description physically relevant and measurable quantities. These are the bubble lifetime and its distribution as well as autocorrelation functions of the bubble dynamics. We here concentrate on the former while addressing the autocorrelation function in the subsequent section dealing with the discrete formalism. More details on the autocorrelation function in the continuum limit can be found in [54–56].

The single bubble dynamics can be analysed in different ways; namely in terms of the underlying Langevin equation including the interpretation of the single bubble dynamics below the melting temperature as a noisy finite time singularity. Alternatively a weak noise analysis allows one to interpret the dynamics through orbitals in phase space portraits. Finally, one may turn to the Fokker-Planck equation (12). For more details we refer to [50, 54, 55].

To determine the lifetime distribution of a bubble once opened we face a technical problem posed by the  $c/x$  term in the drift term of equation (12). One way to circumvent this is to map this Fokker-Planck equation onto the corresponding imaginary time Schrödinger equation of the quantum Coulomb problem [54, 55]. From this formulation one is able to deduce the behaviour of the bubble lifetime. We distinguish three cases.

(i) *Below the melting temperature.* At  $T < T_m$  one can determine the density of the bubble lifetime distribution analytically in the long time limit obtaining

$$\wp(t) \simeq x_0^{1+c} e^{\epsilon|x_0|} e^{-\epsilon^2 t/2} t^{-3/2-c/2}. \quad (13)$$

Thus, we observe a power-law behaviour  $t^{-3/2-c/2}$  with an exponential cutoff at  $\tau = 2/\epsilon^2$  such that the bubble lifetime is always finite. This form for  $\wp(t)$  generalizes the expression of the first passage time density of a bubble without entropy loss correction (i.e.,  $c = 0$ ) with constant drift  $|\epsilon|$  towards bubble closure [50]. For the mean bubble lifetime we find the approximate expression

$$T = \int_0^\infty t \wp(t) dt \simeq \frac{x_0 K_{(c-1)/2}(x_0|\epsilon|)}{|\epsilon| K_{(c+1)/2}(x_0|\epsilon|)}. \quad (14)$$

For large sufficiently large values of  $x_0|\epsilon|$  the ratio of the two Bessel functions tends to 1, in particular, for the Gaussian chain limit  $c = 3/2$  we find  $K_{1/2}(x_0|\epsilon|)/K_{3/2}(x_0|\epsilon|) = 1/(1 + |\epsilon|/x_0)$ . This result for  $T$  includes the characteristic bubble lifetime  $x_0/|\epsilon|$  when the loop entropy correction is neglected ( $c = 0$ ) [50].

(ii) *At the melting temperature.* Right at  $T = T_m$  the drift exerted by the free stacking energy  $\epsilon$  vanishes, and the dynamics is almost free diffusion. The result for the density of bubble lifetimes reads

$$\wp(t) = \frac{2x_0^{1+c}}{\Gamma(1/2 + c/2)} e^{-x_0^2/2t} (2t)^{-3/2-c/2} \quad (15)$$

and is normalized and exact for all times. In this case the power-law  $t^{-3/2-c/2}$  determines the long time behaviour. While for free diffusion ( $c = 0$ ) the corresponding mean bubble lifetime  $\int_0^\infty t \wp(t) dt$  diverges [50], for all  $c > 1$  we encounter the mean bubble lifetime

$$T = \frac{x_0^2}{c-1} \quad (16)$$

which interestingly grows as the square of the initial bubble size in contrast to the linear scaling in the case of diffusion with linear drift in case (i). In addition to the finite mean bubble lifetime a value  $c > 2$  would also cause a power-law decay  $C(t) \sim t^{1-c/2}$  of the associated correlation function at long times in contrast to the plateau  $C(t) \sim 1$  reached for  $1 < c < 2$  [56].

(iii) *Above the melting temperature.* At  $T > T_m$  the situation is opposite to case (i); namely the drift is now directed towards the complete denaturation of the chain. In a long chain the one bubble picture would no longer hold and bubble coalescence needs to be taken into account. However in shorter DNA constructs preferring one single bubble the density of bubble lifetimes would decay exponentially [54, 55].

### 3.2. Discrete approach and sequence dependence

The natural coordinate for the unzipping and zipping of base pairs in DNA breathing dynamics is the location  $x$  of a respective base pair along the chemical backbone of the DNA molecule. By its very nature this is a discrete variable. While in the continuum approach one may include certain given distributions of more and less stable regions (predominantly GC-rich versus predominantly AT-rich) the use of a truly discrete  $x$  allows one to consider any given sequence. This

is of particular importance when analysing actual biologically relevant sequences or those designed sequences that are used in a given experiment. Such a discrete approach in terms of the master equation will be described here. We note that a disadvantage of this method is the limited system size one can *de facto* analyse due to computational constraints.

With a discrete coordinate we are also able to explicitly distinguish hydrogen bonding and stacking energies and use the parameters for the free energies from Krueger *et al* [6]. For the setup sketched in figure 5 we then find the partition function. The positions  $x_L$  and  $x_R$  of the zipper forks correspond to the right and leftmost closed bp of the bubble.  $x_L$  and  $x_R$  are stochastic variables, whose time evolution in the energy landscape defined by the partition factor ( $m \geq 1$ )

$$\mathcal{Z}(x_L, m) = \frac{\xi'}{(1+m)^c} \prod_{x=x_L+1}^{x_L+m} u_{\text{hb}}(x) \prod_{x=x_L+1}^{x_L+m+1} u_{\text{st}}(x) \quad (17)$$

characterizes the bubble dynamics.  $\mathcal{Z}$  is written in terms of  $x_L$  and bubble size  $m = x_R - x_L - 1$ , with  $\mathcal{Z}(m=0) = 1$ . Here,  $\xi' = 2^c \xi$ , where  $\xi \approx 10^{-3}$  is the ring factor for bubble initiation from [6] that is related to the cooperativity parameter  $\sigma_0 \approx 10^{-5}$  [7, 57] by  $\sigma_0 = \xi \exp(\epsilon_{\text{st}})$  [6]. For the entropy loss on forming a closed polymer loop we assign the factor  $(1+m)^{-c}$  [57, 58] and take  $c = 1.76$  for the critical exponent [35]. This corresponds to the Flory form  $3\nu$  for the entropy loss factor for a polymer ring with excluded volume. The best known value for  $\nu$  is 0.588 [59–61]. Note that there exist alternative models taking into account the self-avoiding interactions of the bubble with the rest of the chain, leading to an increased value for  $c$  ( $c \approx 2.1$ ) such that the denaturation transition becomes first order [33, 35]. Note also that a bubble with  $m$  open bps requires breaking of  $m$  hydrogen bonds and  $m+1$  stacking interactions.

The zipper forks move step-wise  $x_{L/R} \rightarrow x_{L/R} \pm 1$  with rates  $t_{L/R}^{\pm}(x_{L/R}, m)$ . We define for bubble size decrease

$$t_L^+(x_L, m) = t_R^-(x_L, m) = k/2 \quad (m \geq 2) \quad (18)$$

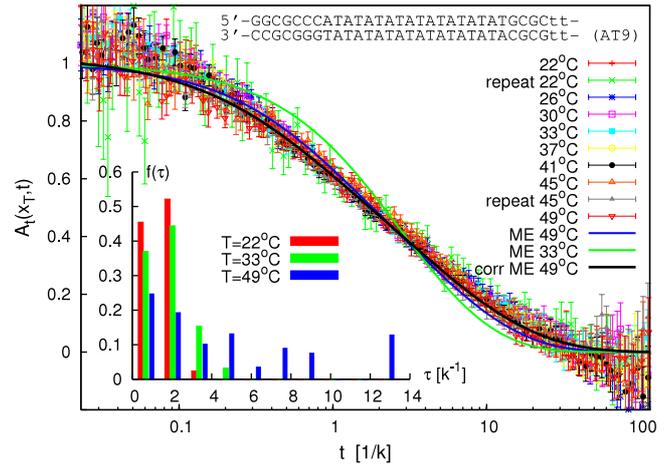
for the two forks<sup>10</sup>. The rate  $k$  characterizes a single bp zipping. Its independence of  $x$  corresponds to the view that bp closure requires the diffusional encounter of the two bases and bond formation; as sterically AT and GC bps are very similar,  $k$  should not significantly vary with bp stacking.  $k$  is the only adjustable parameter of our model, and has to be determined from experiment or future MD simulations. The factor 1/2 is introduced for consistency [51, 52]. Bubble size increase is controlled by

$$\begin{aligned} t_L^-(x_L, m) &= k u_{\text{st}}(x_L) u_{\text{hb}}(x_L) s(m)/2, \\ t_R^+(x_L, m) &= k u_{\text{st}}(x_R + 1) u_{\text{hb}}(x_R) s(m)/2, \end{aligned} \quad (19)$$

for  $m \geq 1$ , where  $s(m) = \{(1+m)/(2+m)\}^c$ . Finally, bubble initiation and annihilation from and to the zero bubble ground state,  $m=0 \leftrightarrow 1$  occur with rates

$$\begin{aligned} t_G^+(x_L) &= k \xi' s(0) u_{\text{st}}(x_L + 1) u_{\text{hb}}(x_L + 1) u_{\text{st}}(x_L + 2) \\ t_G^-(x_L) &= k. \end{aligned} \quad (20)$$

<sup>10</sup> Due to intrachain coupling (e.g., Rouse), larger bubbles may involve an additional ‘hook factor’  $m^{-\mu}$  [51, 52].



**Figure 6.** Scaling plot of  $A_t(x_T, t)$  at various  $T$  for the sequence AT9 from [14] as indicated in the figure. This experimental construct is designed with a weak AT-rich bubble domain in the core, a GC clamp at both ends and additional bulge-loop of DNA single-strand consisting of four T bases. The symbols represent experimental data at various temperatures, see [24, 25] for more details. We also include results from our master equation model. Inset: relaxation time spectrum. See text for more details.

The rates  $t$  fulfil detailed balance conditions. The annihilation rate  $t_G^-(x_L)$  is twice the zipping rate of a single fork, since the last open bp can close either from the left or right. Due to the clamping,  $x_L \geq 0$  and  $x_R \leq M+1$ , ensured by reflecting conditions  $t_L^-(0, m) = t_R^+(x_L, M - x_L) = 0$ . The rates  $t$  together with the boundary conditions fully determine the bubble dynamics.

In the FCS experiment fluorescence occurs if the bps in a  $\Delta$ -neighbourhood of the fluorophore position  $x_T$  are open [14]. Measured fluorescence time series thus correspond to the stochastic variable  $I(t)$ , that takes the value 1 if at least all bps in  $[x_T - \Delta, x_T + \Delta]$  are open, else it is 0. The time averaged  $\langle \bar{I} \rangle$  fluorescence autocorrelation

$$A_t(x_T, t) = \overline{I(t)I(0)} - \overline{I(t)}^2 \quad (21)$$

for the sequence AT9 for various temperatures from [14] are rescaled in figure 6.

DNA breathing is described by the probability distribution  $P(x_L, m, t)$  to find a bubble of size  $m$  located at  $x_L$  whose time evolution follows the master equation

$$\frac{\partial P(x_L, m, t)}{\partial t} = \mathbb{W}P(x_L, m, t). \quad (22)$$

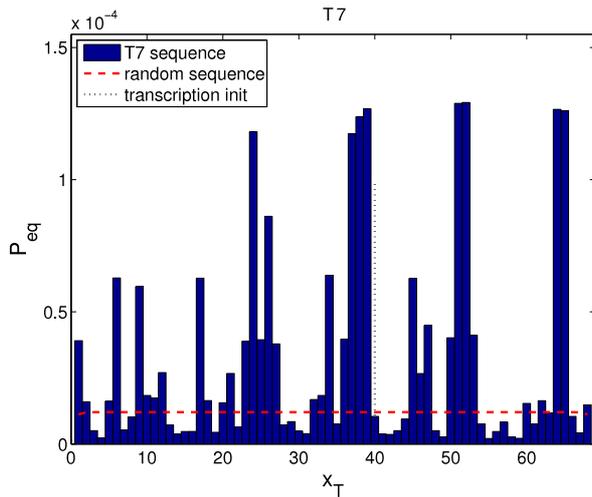
The transfer matrix  $\mathbb{W}$  incorporates the rates  $t$ . Detailed balance guarantees equilibration toward

$$P_{\text{eq}} = \lim_{t \rightarrow \infty} P(x_L, m, t) = \frac{\mathcal{Z}(x_L, m)}{\mathcal{Z}}, \quad (23)$$

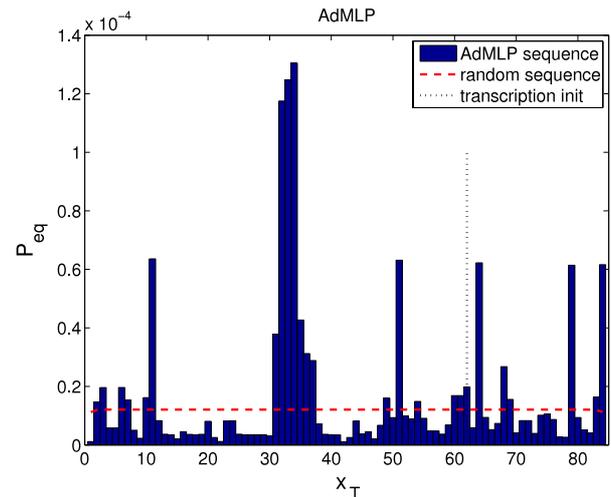
with  $\mathcal{Z} = \sum_{x_L, m} \mathcal{Z}(x_L, m)$  [51, 52, 62]. The master equation and the explicit construction of  $\mathbb{W}$  are discussed at length in [25, 51, 52, 63]. Eigenmode analysis and matrix diagonalization produce all quantities of interest such as the ensemble averaged autocorrelation function

$$A(x_T, t) = \langle I(t)I(0) \rangle - \langle I \rangle^2. \quad (24)$$





**Figure 8.** Equilibrium opening probability of base pairs in the sequence of the bacteriophage T7 core promoter. The dotted line marks the transcription initiation site.



**Figure 9.** Equilibrium opening probability of base pairs in the sequence of the adenovirus major late promoter.

it is located 25 base pairs upstream to the transcription site. The TATA box is normally bound by the TATA binding protein (TBP) during transcription. The TBP unwinds the DNA and strongly bends it. At a later stage the TATA box is bound by RNA polymerase and transcription commences<sup>12</sup>. The high proneness towards bubble formation at the TATA box is therefore believed to actively contribute to transcription initiation<sup>13</sup>.

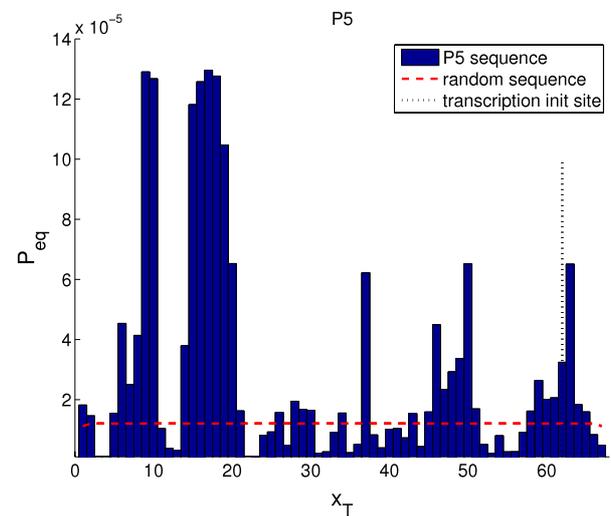
**3.3.1. Bacteriophage T7 core promoter.** Its sequence is displayed in equation (26). It contains the TATA box at base pair labels 36–39. Figure 8 shows the equilibrium probabilities for the base pairs to be open. In this example the TATA box is located right next to the transcription start site. From the graph one can see that indeed the simultaneous opening probability of four base pairs is significantly increased at the position of the TATA box. Note the level of the opening probability of a random sequence also drawn in the figure. Accordingly several domains of significantly increased bubble probability exist along this sequence.

**3.3.2. Adenovirus major late promoter.** Its 86 base pair sequence



contains a transcription start site at the position labelled TSS, compare figure 9. In this example the (extended) TATA box is located upstream at the base pair label—31. In this example the TATA box is extremely more likely to open simultaneously than any other domain along the sequence.

<sup>12</sup> Other proteins are also involved in this quite complex process.  
<sup>13</sup> In particular with the stability parameters from Krueger et al [6] the stacking free energy of a TA/AT pair of base pairs essentially vanishes.



**Figure 10.** Equilibrium opening probability of base pairs in the sequence of the Adeno Associated Viral P5 promoter.

**3.3.3. Adeno associated viral P5 promoter.** This sequence consists of the 69 base pairs



and supports binding of TBP at the (extended) TATA box as well as the binding of the Yin Yang 1 (YY1) transcription factor. YY1 is known to interact with the TBP [67]. YY1 binds to a specific sequence element of the form CCATNTT in the sequence. As can be seen from figure 10 these two binding motifs have a significantly higher cooperative opening probability than any other sequence element of this promoter. The analysis also shows a broader but lower peak around the transcription start site.

In summary this analysis shows that indeed local instability of the DNA sequence appears to occur at specific

binding sequences for proteins involved in transcription initiation. Whether it is just the lower free energy needed to break these sequences or indeed rare bubble openings at these site that help the protein binding remains an open question.

#### 4. DNA bubble coalescence

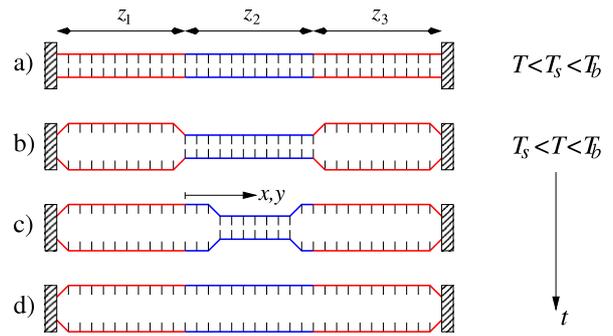
It has been shown in a quantitative analysis that the experimentally accessible autocorrelation function is sensitive to the stacking parameters of DNA [24, 25]. However, it has not been fully appreciated to what extent the fluorophore and quencher molecules, that are attached to the DNA construct in the experiments reported in [14, 64, 68], influence the stability of DNA. Moreover, the zipping rates measured in the single molecule fluorescence setup differ from those determined in NMR experiments [11, 14]. We here propose and study an alternative setup for the single molecule fluorescence investigation of DNA breathing, as shown in figure 11, that may improve and complement the single molecule data obtained from a DNA construct with a single bubble domain<sup>14</sup>. In this setup, a short stretch of DNA, clamped at both ends, is designed such that two soft zones consisting of weaker AT bps are separated by a more stable barrier region rich in GC bps. For simplicity, we assume that both soft zones and barrier are homopolymers with a bp-dissociation free energy  $\Delta G_s$  and  $\Delta G_b$ , respectively, and, in accordance with the experimental findings of [14], we neglect secondary structure formation in the barrier zone. At temperatures higher than the melting temperature  $T_s$  of the soft zones but still lower than the melting temperature  $T_b$  of the barrier region, thermal fluctuations will gradually dissociate the barrier, until the two bubbles coalesce. Once coalesced, the free energy corresponding to one cooperativity factor  $\sigma_0 \approx 10^{-5 \dots -3}$  is released, stabilizing the coalesced bubble against reclosure of the barrier. Moreover there exists a significant dynamic barrier stemming from the necessity of diffusional encounter of two bases in order to reanneal the barrier. Both points lead to a long lifetime of the coalesced state. This fact should allow for a meaningful measurement of the coalescence time in experiment, and therefore provide a new and sensitive method to measure DNA stability data and base pair zipping rates. We also study the case when the system is prepared as above and then  $T$  suddenly increased such that  $T > T_b > T_s$  so that the system is driven towards coalescence. In both cases the two boundaries between bubbles and barrier perform a (biased) random walk in opposite free energy potentials.

The statistical weight of the construct before coalescence,

$$\mathcal{L}_{X,Y} = (\xi e^{N_L \beta \varepsilon'}) e^{(X-Y+N)\beta \varepsilon} (\xi e^{N_R \beta \varepsilon'}), \quad (29)$$

at  $T_b > T > T_s$  involves the cooperativity factor  $\xi \approx 10^{-5}$  for each bubble, and a Boltzmann factor for each broken bp with free energies  $\varepsilon' > 0$  and  $\varepsilon < 0$ , compare reference [24].

<sup>14</sup> Due to the stabilization of the coalesced bubble as discussed below this setup would allow for measurements of a first passage for the merging of the two initial bubbles and thus distinguish this setup from the open-close dynamics in the previous experiments.



**Figure 11.** Schematic of the DNA construct for bubble coalescence. Note that the position of both ends of the barrier region are measured from the same point (the position of the leftmost barrier base pair).

Upon coalescence, the boundary free energy corresponding to one factor  $\xi$  is released,

$$\mathcal{L}_{\text{coal}} = \xi e^{(N_L + N_R)\beta \varepsilon' + N\beta \varepsilon}, \quad (30)$$

stabilizing the system against immediate transition back to a two bubble state. It is this distinctive feature that should render this setup an interesting model system for single molecule analyses of DNA denaturation dynamics, as the coalesced state can be determined by measuring first passage time statistics (corresponding to the introduction of an absorbing boundary condition at the point of coalescence).

In our analysis we use a continuum approach to the stochastic motion of the two zipping forks at either end of the barrier zone with locations  $x$  and  $y$ . The probability density  $P(x, y, t)$  then follows the bivariate Fokker-Planck equation [69]

$$\frac{\partial}{\partial t} P(x, y, t) = \left( \left[ \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right] - 2f \frac{\partial}{\partial x} + 2f \frac{\partial}{\partial y} \right) \times P(x, y, t), \quad (31)$$

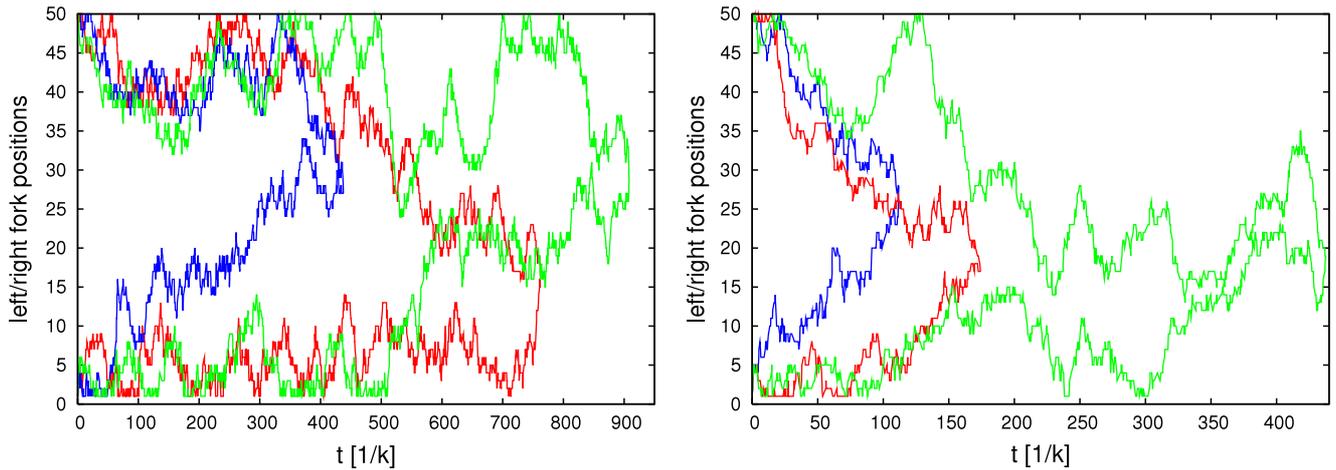
with the dimensionless force  $f = N(u - 1)/(1 + u)$  and time rescaled by  $k(1 + u)/2N^2$ . Equation (31) is completed by the initial condition  $P(x, y, 0) = \delta(x - x_0)\delta(y - y_0)$  and the reflecting boundary conditions (the bubbles in the soft zones are assumed to be open at all times)

$$\left( \frac{\partial}{\partial x} - 2f \right) P(x, y, t) \Big|_{x=0} = \left( \frac{\partial}{\partial y} + 2f \right) P(x, y, t) \Big|_{y=1} = 0. \quad (32)$$

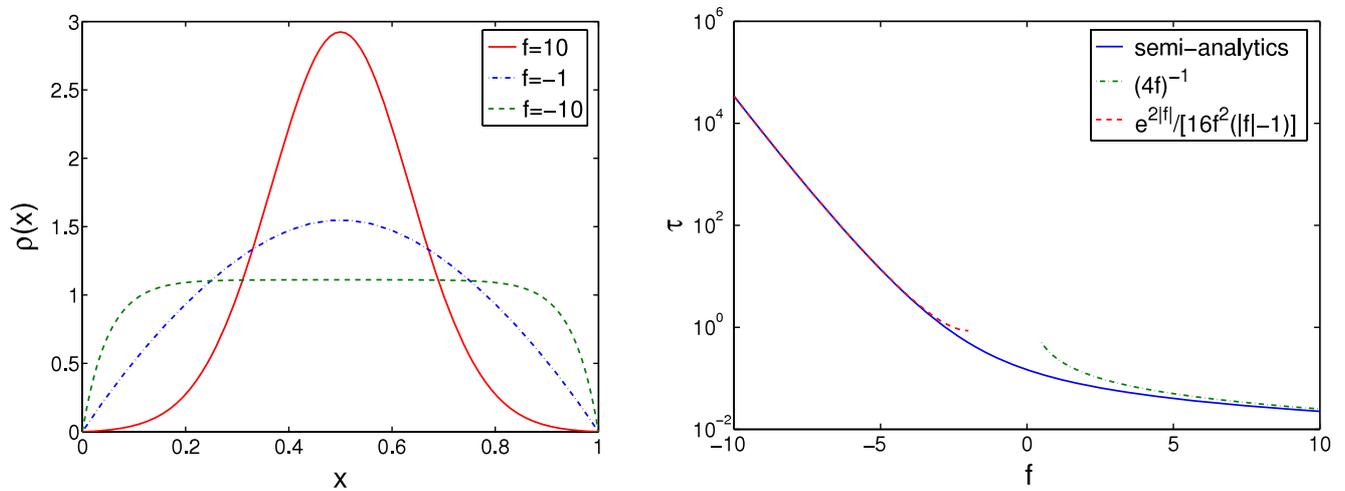
Moreover, we impose the absorbing boundary condition  $P(x, x, t) = 0$ . This defines the vicious walker property [70], terminating the process when the two walkers meet. The fact that the two walker move in opposite potentials actually make this problem a previously unsolved case of vicious walkers models [69].

Typical examples of individual trajectories resulting from a Gillespie algorithm are displayed in figure 12, where traces of the two interfaces (forks) cornering the barrier region are shown. Bubble coalescence terminates each pair of trajectories.

The analysis in [69] reveals the distribution of coalescence positions (i.e., where the two zipper forks eventually meet) and



**Figure 12.** Trajectories of the random motion of the two bubble forks.



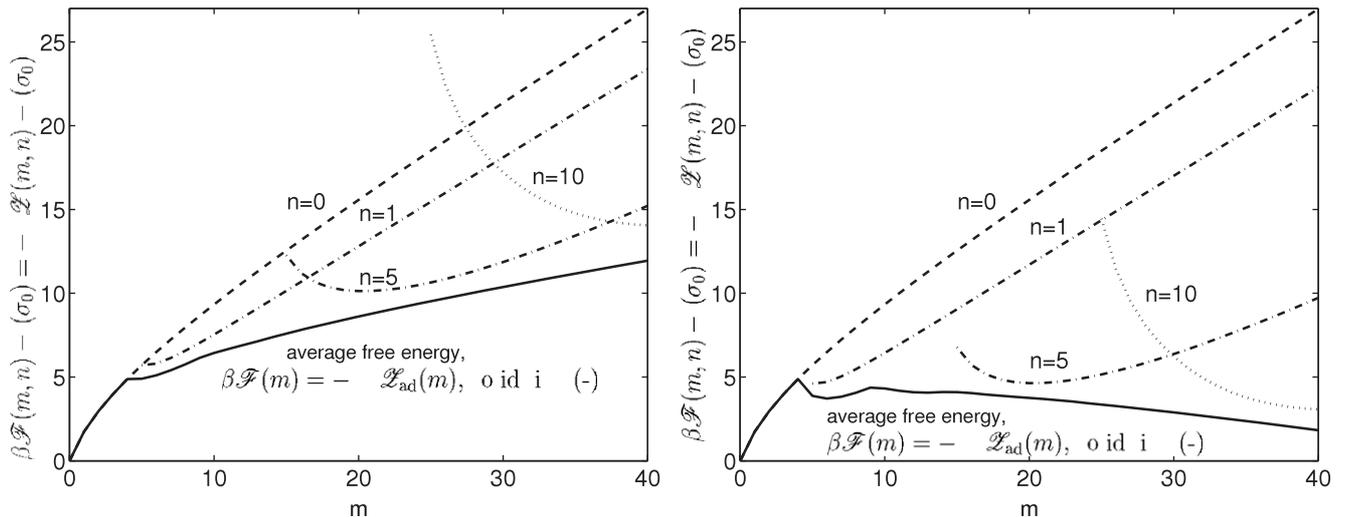
**Figure 13.** Left: distribution of coalescence positions within the rescaled barrier zone  $[0, 1]$ . Right: distribution of coalescence times.

the coalescence times, as shown in figure 13. The curves for the PDF  $\rho(x)$  of the coalescence position exhibit a pronounced crossover from a relatively sharply peaked form to an almost flat behaviour. The former occurs for large positive force  $f$ , corresponding to a strong drift toward a potential well, with negligible influence of the boundary conditions. In contrast, for large negative  $f$ , corresponding to a high barrier for coalescence, the insensitivity of  $\rho(x)$  to the position  $x$  can be explained in terms of a simple Arrhenius argument: the probability of the walker to be at a position  $x$  is proportional to the Boltzmann weight,  $\exp(-\beta\phi(x))$ , where  $\phi(x) = -\int^x F(x') dx'$  is the free energy corresponding to the force  $F(x)$ . Then, the joint probability to have both walkers meet at the same position is given by the product  $\exp(-\beta[\phi_L(x) + \phi_R(x)]) \approx \text{const}$  as the two walkers are in opposite linear potentials and the position dependence of the exponent cancels out. This simple picture necessarily breaks down close to the boundaries. (ii) The  $f$ -dependence of the mean first passage time  $\tau$  crosses over from the  $\tau \simeq 1/f$  behaviour typical for diffusion in a strong positive force pushing the two walkers together, to the exponential form  $\tau \simeq \exp(2|f|)$  of

the associated Kramers problem. The former problem was studied in [50] by neglecting the boundaries and switching to the relative coordinate description which enables one to find the analytic result  $\tau = 1/(4f)$ . For the Kramers problem ( $f \ll -1$ ) the analytic solution for both  $\rho(x) = [1 - e^{-2|f|x} - e^{-2|f|(1-x)}]|f|/(|f|-1)$  and  $\tau = e^{2|f|}/[16f^2(|f|-1)]$  can be found rather easily [71] by the expansion into the lowest two eigenmodes of  $p(x, t|x_0)$ .

### 5. Coupled dynamics of DNA bubbles and selectively single-strand DNA binding proteins

A traditional puzzle had been the question why the presence of selectively single-strand DNA binding proteins (SSBs) does not lead to full denaturation of the DNA [1]. While ideas about a kinetic block were brought forth relatively early [72–74], experimentally this puzzle could only be solved by single molecule methods in which the denaturation was not induced by temperature but force. In a series of experiments the binding and unbinding kinetics of SSBs and their mutants to DNA denaturation bubbles and the



**Figure 14.** Effective free energy of the SSB–DNA bubble interaction in the limit  $\gamma \gg 1$  (—), and free energy landscape for various fixed  $n$  ( $u = 0.6, M = 40, c = 1.76, \lambda = 5$ ). Left:  $\kappa = 0.5$ ; right: stronger binding,  $\kappa = 1.5$ . In the latter case the binding strength of the SSB suffices to cause a decreasing effective free energy and therefore induce full denaturation of the DNA. Due to the finite size effects the nucleation barrier for initiation of SSB exchange has to be crossed.

resulting effect on the denaturation force were studied in great detail [19, 75, 76]. Here we discuss a simple model for the SSB–DNA interaction in a homopolymer approach by a master equation approach [51, 52].

The quantity of interest is the joint probability  $P(m, n, t)$  to have a bubble consisting of  $m$  broken bps, and  $n$  SSBs bound to the two arches of the bubble. In addition to the rates for bubble increase and decrease, the rates for SSB-binding and unbinding are necessary to define the breathing dynamics in the presence of SSBs. On the statistical level, the effect of the SSBs becomes coupled to the motion of the zipper forks. Thus, the rate for bubble size decrease is proportional to the probability that no SSB is located right next to the corresponding zipper fork; and the rate for SSB-binding is proportional to the probability that there is sufficient unoccupied space on the bubble. Binding is allowed to be asymmetric with respect to the two arches of the bubble, and is related to a parking lot problem in the following sense. The number  $\lambda$  of bases occupied by a bound SSB is usually (considerably) larger than one. In order to be able to bind in between two already bound SSBs, the distance between these two SSBs must be larger than  $\lambda$ . The larger  $\lambda$  the less efficient the SSB-binding becomes, similar to parking large cars on a parking lot designed for small vehicles. Apart from the binding size  $\lambda$  of the SSBs, two additional physical parameters come into play: the unbinding rate  $q$  of the SSBs, and their binding strength  $\kappa = c_0 K^{\text{eq}}$  consisting of the volume concentration  $c_0$  of SSBs and the equilibrium binding constant  $K^{\text{eq}} = v_0 \exp(\beta|E_{\text{SSB}}|)$ , with the typical SSB volume  $v_0$  and binding energy  $E_{\text{SSB}}$ .

The coupled dynamics of SSB-binding and bubble breathing is discussed in [51, 52]; similar effects in end-denaturing DNA were studied in [68] in detail. Here, we report the behaviour of the effective free energy landscape in the limit of fast SSB-binding in the sense that the dimensionless parameter  $\gamma \equiv q/k$  of SSB-unbinding and bubble zipping

rates is large,  $\gamma \gg 1$ . This limit allows one to average out the SSB-dynamics and to calculate an effective free energy, in which the bubble dynamics with the slow variable  $m$  runs off. The result for two different binding strengths  $\kappa$  is shown in figure 14, along with the free energies corresponding to keeping  $n$  fixed. It is distinct that while for lower  $\kappa$  the presence of SSBs diminishes the slope of the effective free energy, for larger  $\kappa$  the slope actually becomes negative. In the first case, that is, the bubble opening is more likely, but still globally unfavourable. In the latter case, the presence of SSBs indeed leads to full denaturation. One observes distinct finite size effects due to  $\lambda > 1$ : only when the bubble reaches a minimal size  $m \geq \lambda$ , SSB-binding may occur, a second SSB is allowed to bind to the same arch only once  $m \geq 2\lambda$ , etc. This effect also produces the nucleation barrier for full denaturation in the right plot of figure 14. Similar finite size effects were investigated for biopolymer translocation in [77, 78]. We note that the transition to denaturation could also be achieved by reaching a smaller positive slope of the effective free energy in the presence of SSBs, and additional titration or change of the effective temperature through actual temperature change or mechanical stretching as performed in the experiments reported in [19, 75, 76].

## 6. Concluding remarks

DNA possesses a number of properties that render it a very attractive model system. Thus the study of the DNA denaturation transition has occupied statistical physics for around five decades. DNA is comparatively thin and stiff locally while its overall length is fully macroscopic. Thus it is probably the closest available example for testing the predictions from polymer physics. In particular single DNA can be probed and manipulated and its interactions with binding proteins and chemicals investigated. This includes the monitoring of single DNA bubbles and their interaction

with specifically single-stranded DNA binding proteins, both described here, as well as the interaction of DNA with intercalators [40]. By now single molecule assays can also be used to study the search mechanisms of DNA binding proteins scanning it for specific binding sites relevant in gene regulation and DNA repair [20, 79–81]. This attractiveness of DNA combines with its ultimate role as the molecule of life and therefore is one of the finest examples where the interests of biological physics meet those of biochemists and molecular biologists.

The label *century of biology* is frequently bestowed upon the 21st (e.g., [82]). In the wash of the success of biology, molecular and systems biology in particular, one experiences a mushrooming number of works in the biological physics sector. Indeed many of these problems pose very attractive and new questions to physicists and along with the availability of single molecule techniques prompt new advances, for instance in statistical physics.

On a general level the contribution of physics to biology lies in the number of various concepts that physicists can provide, ranging from novel single molecule probing methods to the theoretical framework that physicists are trained in. This has been true for the development of the statistical and dynamic models to describe DNA denaturation and breathing, or for a large part of the development of facilitated diffusion models for dilute solutions peaking in the Berg–von Hippel model [83], and possibly will pertain even more to the current undertaking to understand transport and regulation processes in living cells and organisms. Thus attempts to understand the topology of cellular networks such as protein networks or regulatory networks more generally caused the development of a new area in physics and mathematics whose implications are now feeding back to biology [87]. Or, instead of reducing biochemical pathways in the cell to sets of rate equations the spatial aspect of diffusion at low concentrations, for instance, of regulatory proteins, are being promoted as important ingredients [79, 84–86].

The DNA mechanisms outlined in this paper are expected to contribute to the way chemicals and proteins bind to DNA and initiate subsequent biochemical reaction cascades. We showed here that we are on the way to a quantitative understanding of DNA breathing and that it can be probed on the level of single DNA molecules. Techniques such as fluorescence correlation spectroscopy or optical tweezers will most likely become crucial when trying to understand DNA breathing under cellular conditions or conditions of reconstituted crowding when, among other things, the entropic contribution on base pair denaturation should become less relevant but DNA bubbles may be promoted by local DNA structure due to the packaging of DNA (e.g., twist-induction of bubbles). An essential question is how under more realistic conditions the stability parameters of DNA become modified, in particular, whether the asymmetry between different nearest neighbour combinations of base pairs becomes more pronounced. Equally important will be the question how the local stability is affected by the fact that *in vivo* the DNA is decorated by a variety of other binding proteins all of which also contribute to change the local structure.

Thus the simultaneous presence of several transcription factors necessary to trigger a certain biochemical pathway may significantly alter the opening probabilities of the transcription start sites etc, and therefore also their occurrence frequency. A first step towards a better understanding of such complex interactions may be to introduce labelled DNA constructs and to follow their behaviour when crowding is increased and other species of binding proteins introduced. Such knowledge will also be relevant to better control DNA-targeting therapies.

Another example for the challenges ahead is the current lack of understanding of non-local biochemical processes in living cells under conditions of molecular crowding [88–91]. It is being realized that knowledge obtained under dilute conditions *in vitro* does not necessarily translate to the situation *in vivo* and this point will need considerable more quantitative investigation. As it stands the input from biological physics will be crucial, for example regarding the diffusive processes. It appears that subdiffusion of biopolymers occurs in conditions of molecular crowding [92–94], this being the likely source for strong scatter of time averages and apparent diffusivities of single trajectories [92, 95, 96], requiring great care in the quantitative analysis [97, 98]. In the end a more local picture of *in vivo* gene regulation will emerge.

## Acknowledgments

We would like to thank Ralf Blossey for the invitation to contribute this article to this special issue. We are also happy to acknowledge discussions with Suman K Banik, Maxim Frank-Kamenetskii, Olek Krichevsky, Michael Lomholt, Tomáš Novotný, Igor M Sokolov, and Mark C Williams. RM acknowledges partial funding of this research by the Canada Research Chair programme and the National Sciences and Engineering Research Council of Canada. The work of AH was supported by the NIH through grant 1 SC3 GM083779-01 and by the AFOSR through grant FA9550-06-1-0408.

## References

- [1] Kornberg A and Baker T A 1992 *DNA Replication* (New York: Freeman)
- [2] Watson J D and Crick F H C 1953 *Nature* **171** 737
- [3] Frank-Kamenetskii M D 1997 *Phys. Rep.* **288** 13
- [4] Delcourt S G and Blake R D 1991 *J. Biol. Chem.* **266** 15160
- [5] SantaLucia J Jr 1998 *Proc. Natl Acad. Sci.* **95** 1460
- [6] Krueger A, Protozanova E and Frank-Kamenetskii M D 2006 *Biophys. J.* **90** 3091
- [7] Poland D and Scheraga H A 1970 *Theory of Helix–Coil Transitions in Biopolymers* (New York: Academic)
- [8] Wartell R M and Benight A S 1985 *Phys. Rep.* **126** 67
- [9] Carlon E, Malki M L and Blossey R 2005 *Phys. Rev. Lett.* **94** 178101
- [10] Yeramian E 2000 *Gene* **255** 139  
Yeramian E 2000 *Gene* **255** 151
- [11] Guéron M, Kochoyan M and Leroy J L 1987 *Nature* **328** 89
- [12] Thumm W, Seidl A and Hinz H-J 1988 *Nucleic Acids Res.* **16** 11737
- [13] López-García P and Forterre P 1997 *Mol. Microbiol.* **23** 1267
- [14] Altan-Bonnet G, Libchaber A and Krichevsky O 2003 *Phys. Rev. Lett.* **90** 138101
- [15] Smith S B, Cui Y J and Bustamante C 1996 *Science* **271** 795

- [16] Wenner J R, Williams M C, Rouzina I and Bloomfield V A 2002 *Biophys. J.* **82** 3160
- [17] Clausen-Schumann H, Rief M, Tolksdorf C and Gaub H E 2000 *Biophys. J.* **78** 1997
- [18] Metzler R, Ambjörnsson T, Hanke A and Levene S 2007 *J. Comput. Theor. Nanosci.* **4** 1
- [19] Pant K, Karpel R L and Williams M C 2003 *J. Mol. Biol.* **327** 571
- [20] Sokolov I M, Metzler R, Pant K and Williams M C 2005 *Biophys. J.* **89** 895
- [21] Ambjörnsson T and Metzler R 2005 *J. Phys.: Condens. Matter* **17** S1841  
Ambjörnsson T and Metzler R 2005 *Phys. Rev. E* **72** 030901(R)
- [22] Choi C H, Kalosakas G, Rasmussen K Ø, Hiromura M, Bishop A R and Usheva A 2004 *Nucleic Acids. Res.* **32** 1584
- [23] Kalosakas G, Rasmussen K Ø, Bishop A R, Choi C H and Usheva A 2004 *Europhys. Lett.* **68** 127
- [24] Ambjörnsson T, Banik S K, Krichevsky O and Metzler R 2006 *Phys. Rev. Lett.* **97** 128105
- [25] Ambjörnsson T, Banik S K, Krichevsky O and Metzler R 2007 *Biophys. J.* **92** 2674
- [26] Peyrard M and Bishop A R 1989 *Phys. Rev. Lett.* **62** 2755
- [27] Dauxois T, Peyrard M and Bishop A R 1993 *Phys. Rev. E* **44** R44
- [28] Alexandrov B S, Wille L T, Rasmussen K Ø, Bishop A R and Blagoev K B 2006 *Phys. Rev. E* **74** 050901(R)
- [29] Cocco S and Monasson R 1999 *Phys. Rev. Lett.* **83** 5178
- [30] Barbi M, Lepri S, Peyrard M and Theodorakopoulos N 2003 *Phys. Rev. E* **68** 061909
- [31] Hanke A, Ochoa M G and Metzler R 2008 *Phys. Rev. Lett.* **100** 018106
- [32] Rouzina I and Bloomfield V A *Biophys. J.* **80** 882  
Rouzina I and Bloomfield V A *Biophys. J.* **80** 894
- [33] Kafri Y, Mukamel D and Peliti L 2000 *Phys. Rev. Lett.* **85** 4988  
Kafri Y, Mukamel D and Peliti L 2002 *Eur. Phys. J. B* **27** 132
- [34] Fisher M E 1966 *J. Chem. Phys.* **44** 616
- [35] Richard C and Guttman A J 2004 *J. Stat. Phys.* **115** 925
- [36] Carlon E, Orlandini E and Stella A L 2002 *Phys. Rev. Lett.* **88** 198101
- [37] de Gennes P G 1979 *Scaling Concepts in Polymer Physics* (Ithaca, NY: Cornell University Press)
- [38] Orlandini E, Bhattacharjee S M, Marenduzzo D, Maritan A and Seno F 2001 *J. Phys. A: Math. Gen.* **34** L751
- [39] Rahi S J, Hertzberg M P and Kardar M 2008 arXiv:0806.2837
- [40] Vladescu I D, McCauley M J, Rouzina I and Williams M C 2005 *Phys. Rev. Lett.* **95** 158102
- [41] Shokri L, McCauley M J, Rouzina I and Williams M C 2008 *Biophys. J.* **95** 1248
- [42] Wenner J R, Williams M C, Rouzina I and Bloomfield V A 2002 *Biophys. J.* **82** 3160
- [43] Rudnick J and Kuriabova T 2008 *Phys. Rev. E* **77** 051903
- [44] Kreuzer H J, Einert T and Netz R R, unpublished
- [45] Santosh M and Maiti P K 2009 *J. Phys.: Condens. Matter* **21** 034113
- [46] Storm C and Nelson P C 2003 *Phys. Rev. E* **67** 051905
- [47] Cocco S, Yan J, Leger J-F, Chatenay D and Marko J F 2004 *Phys. Rev. E* **70** 011910
- [48] Whitelam S, Pronk S and Geissler P L 2008 arXiv:0806.0505
- [49] Piana S 2005 *Nucleic Acids Res.* **33** 7029
- [50] Hanke A and Metzler R 2003 *J. Phys. A: Math. Gen.* **36** L473
- [51] Ambjörnsson T and Metzler R 2005 *J. Phys.: Condens. Matter* **17** S1841
- [52] Ambjörnsson T and Metzler R 2005 *Phys. Rev. E* **72** 030901(R)
- [53] Hwa T, Marinari E, Sneppen K and Tang L-H 2003 *Proc. Natl Acad. Sci. USA* **100** 4411
- [54] Fogedby H C and Metzler R 2007 *Phys. Rev. Lett.* **98** 070601
- [55] Fogedby H C and Metzler R 2007 *Phys. Rev. E* **76** 061915
- [56] Bar A, Kafri Y and Mukamel D 2007 *Phys. Rev. Lett.* **98** 038103
- [57] Blake R D, Bizzaro J W, Blake J D, Day G R, Delcourt S G, Knowles J, Marx K A and SantaLucia J Jr 1999 *Bioinformatics* **15** 370
- [58] Fixman M and Freiere J J 1977 *Biopolymers* **16** 2693
- [59] Le Guillou J C and Zinn-Justin J 1977 *Phys. Rev. Lett.* **39** 95
- [60] Caracciolo S, Causo M S and Pelissetto A 1998 *Phys. Rev. E* **57** R1215
- [61] Hsu H-P, Nadler W and Grassberger P 2004 *Macromolecules* **37** 4658
- [62] van Kampen N G 1992 *Stochastic Processes in Physics and Chemistry* (Amsterdam: North-Holland)
- [63] Ambjörnsson T, Lomholt M A, Banik S K and Metzler R 2007 *Phys. Rev. E* **75** 021908
- [64] Krichevsky O and Bonnet G 2002 *Rep. Prog. Phys.* **65** 251
- [65] Banik S K, Ambjörnsson T and Metzler R 2005 *Europhys. Lett.* **71** 852
- [66] Alberts B, Johnson A, Lewis J, Raff M, Roberts K and Walter P 2002 *Molecular Biology of the Cell* (New York: Garland)
- [67] Fry C G and Farnham P J 1999 *J. Biol. Chem.* **274** 29583
- [68] Ambjörnsson T and Metzler R 2005 *J. Phys.: Condens. Matter* **17** S4305
- [69] Novotny T, Pedersen J N, Hansen M S, Ambjörnsson T and Metzler R 2007 *Europhys. Lett.* **77** 48001
- [70] Fisher M E 1984 *J. Stat. Phys.* **34** 667
- [71] Pedersen J N, Novotny T, Hansen M S, Ambjörnsson T and Metzler R, unpublished
- [72] Karpel R L 1990 *The Biology of Non-Specific DNA-Protein Interactions* ed A Revzin (Boca Raton, FL: CRC Press)
- [73] Jensen D E, Kelly R C and von Hippel P H 1976 *J. Biol. Chem.* **251** 7215
- [74] Karpel R L 2002 *IUBMB Life* **53** 161
- [75] Pant K, Karpel R L, Rouzina I and Williams M C 2004 *J. Mol. Biol.* **336** 851
- [76] Pant K, Karpel R L, Rouzina I and Williams M C 2005 *J. Mol. Biol.* **349** 317
- [77] Ambjörnsson T and Metzler R 2004 *Phys. Biol.* **1** 77
- [78] Ambjörnsson T, Lomholt M and Metzler R 2005 *J. Phys.: Condens. Matter* **17** S3945
- [79] van den Broek B, Lomholt M A, Kalisch S-M J, Metzler R and Wuite G J L 2008 *Proc. Natl Acad. Sci. USA* **105** 15738–42
- [80] Wang Y M, Austin R H and Cox E C 2006 *Phys. Rev. Lett.* **97** 048302
- [81] Elf J, Li G-W and Xie X S 2007 *Science* **316** 1191
- [82] Venter C and Cohen D 2004 *New Persp. Quart.* **21** 73
- [83] von Hippel P H and Berg O G 1989 *J. Biol. Chem.* **264** 675
- [84] Kolesov G, Wunderlich Z, Laikova O N, Gelfand M S and Mirny L A 2007 *Proc. Natl Acad. Sci. USA* **104** 13948
- [85] Warren P B and ten Wolde P R 2004 *J. Mol. Biol.* **342** 1379
- [86] Lomholt M A, van den Broek B, Kalisch S-M J, Wuite G J L and Metzler R, unpublished
- [87] Barabási A-L and Oltvai Z N 2004 *Nat. Rev. Gen.* **5** 101
- [88] Ellis R J and Minton A P 2003 *Nature* **425** 27
- [89] Rivas G, Ferrone F and Herzfeld J 2004 *EMBO Rep.* **5** 23
- [90] Kornberg A 2000 *For the Love of Enzymes* (Cambridge, MA: Harvard University Press)
- [91] Kornberg A 2000 *J. Bacteriol.* **182** 3613
- [92] Golding I and Cox E C 2006 *Phys. Rev. Lett.* **96** 098102
- [93] Weiss M, Elsner M, Kartberg F and Nilsson T 2004 *Biophys. J.* **87** 3518
- [94] Banks D S and Fradin C 2005 *Biophys. J.* **89** 2960
- [95] Tolić-Nørrelykke I M, Munteanu E L, Thon G, Oddershede L and Berg-Sorensen K 2004 *Phys. Rev. Lett.* **93** 078102
- [96] Platani M, Goldberg I, Lamond A I and Swedlow J R 2002 *Nat. Cell Biol.* **4** 502
- [97] He Y, Burov S, Metzler R and Barkai E 2008 *Phys. Rev. Lett.* **101** 058101
- [98] Lubelsky A, Sokolov I M and Klafter J 2008 *Phys. Rev. Lett.* **100** 250602