INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **17** (2005) S1841–S1869 doi:10.1088/0953-8984/17/20/013

Binding dynamics of single-stranded DNA binding proteins to fluctuating bubbles in breathing DNA

Tobias Ambjörnsson and Ralf Metzler

NORDITA–Nordic Institute for Theoretical Physics, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark

E-mail: ambjorn@nordita.dk and metz@nordita.dk

Received 23 November 2004, in final form 24 January 2005
Published 6 May 2005
Online at stacks.iop.org/JPhysCM/17/S1841

Abstract

We investigate the dynamics of a single local denaturation zone in a DNA molecule, a so-called DNA bubble, in the presence of single-stranded DNA binding proteins (SSBs). In particular, we develop a dynamical description of the process in terms of a two-dimensional master equation for the time evolution of the probability distribution of having a bubble of size m with n bound SSBs, for the case when m and n are the slowest variables in the system. We derive explicit expressions for the equilibrium statistical weights for a given m and n, which depend on the statistical weight u associated with breaking a base-pair interaction, the loop closure exponent c, the cooperativity parameter σ_0, the SSB size λ, and binding strength κ. These statistical weights determine, through the detailed balance condition, the transfer coefficient in the master equation. For the case of slow and fast binding dynamics the problem can be reduced to one-dimensional master equations. In the latter case, we perform explicitly the adiabatic elimination of the fast variable n. Furthermore, we find that for the case that the loop closure is neglected and the binding dynamics is vanishing (but with arbitrary σ_0) the eigenvalues and the eigenvectors of the master equation can be obtained analytically, using an orthogonal polynomial approach. We solve the general case numerically (i.e., including SSB binding and the loop closure) as a function of statistical weight u, binding protein size λ, and binding strength κ, and compare to the fast and slow binding limits. In particular, we find that the presence of SSBs in general increases the relaxation time, compared to the case when no binding proteins are present. By tuning the parameters, we can drive the system from regular bubble fluctuation in the absence of SSBs to full denaturation, reflecting experimental and *in vivo* situations.

1. Introduction

The Watson–Crick double-helix, or, more precisely, its B-form, is the thermodynamically stable configuration of a DNA molecule under physiological conditions (normal salt and room/body temperature). This stability is effected first by Watson–Crick H-bonding, that is essential for the specificity of base-pairing, i.e., for the key-lock principle according to which the base (nucleotide) adenine (A) exclusively binds to thymine (T), and guanine (G) only to cytosine (C). Base-pairing therefore guarantees the high level of fidelity during replication and transcription. The second contribution to DNA-helix stability comes from base-stacking between neighbouring base-pairs: through hydrophobic interactions between the planar aromatic bases, that overlap geometrically and electronically, the base-pair stacking stabilizes the helical structure against the repulsive electrostatic force between the negatively charged phosphate groups located at the outside of the DNA double-strand. While hydrogen bonds contribute only little to the helix stability, the major support comes from base-stacking [1, 2].

An important feature of double-stranded DNA (dsDNA) is the ease with which its component chains can come apart and rejoin, without damaging the chemical structure of the two single-strands. This is crucial to many physiological processes such as replication via the proteins DNA helicase and polymerase, and transcription through RNA polymerase. During these processes, the proteins unzip a certain region of the double-strand, to obtain access to the genetic information stored in the bases in the core of the double-helix [1, 3, 4]. This unzipping corresponds to breaking the hydrogen bonds between the base-pairs. Classically, the so-called melting and reannealing behaviour of DNA has been studied in solution *in vitro* by increasing the temperature, or by titration with acid or alkali. During thermal melting, the stability of the DNA duplex is related to the content of triple-hydrogen-bonded G–C base-pairs: the larger the fraction of G–C pairs, the higher the required melting temperature or pH value. Thus, under thermal melting, dsDNA starts to unwind in regions rich in double-hydrogen-bonded A–T base-pairs, and then proceeds to regions of progressively higher G–C content [1, 2]. Conversely, molten, complementary chains of single-stranded DNA (ssDNA) begin to reassociate and eventually reform the original double-helix under incubation at roughly 25° below the melting temperature T_m [1]. The relative amount of molten DNA in a solution can be measured by UV spectroscopy, revealing large changes in absorption in the presence of perturbed base-stacking [5]. Careful melting studies allow one to obtain accurate values for the stacking energies of the various combinations of neighbouring base-pairs, a basis for detailed thermodynamic modelling of DNA-melting and DNA-structure *per se* [6, 7]. In fact, thermal melting data have been used to identify coding sequences of the genome due to the different G–C content [8–10].

While the double-helix is the thermodynamically stable configuration of the DNA molecule below T_m (or at non-denaturing pH), even at physiological conditions there exist local denaturation zones, so-called DNA-bubbles, predominantly in A–T-rich regions of the genome [5, 11]. Driven by ambient thermal fluctuations, a DNA-bubble is a dynamical entity whose size varies by thermally activated zipping and unzipping of successive base-pairs at the two forks where the ssDNA-bubble is bordered by the dsDNA-helix. This incessant zipping and unzipping leads to a random walk in the bubble-size coordinate, and to a finite lifetime of DNA-bubbles under non-melting conditions, as eventually the bubble closes due to the comparatively large loop initiation barrier [5, 11]. DNA-breathing opens up a few tens of base-pairs [12]. It has been demonstrated recently that by fluorescence correlation methods the fluctuations of DNA-bubbles can be explored on the single molecule level, revealing a multistate kinetics that corresponds to the picture of successive zipping and unzipping of single base-pairs[1]. At room temperature, the characteristic closing time of an unbounded

[1] Essentially, the zipper model advocated by Kittel [13].

base-pair was found to be in the range 10–100 μs corresponding to an overall bubble lifetime in the range of tens of milliseconds [14]. The multistate nature of the DNA-breathing was also confirmed by a UV-light absorption study [15].

The presence of fluctuating DNA-bubbles is essential to the understanding of the binding of single-stranded DNA binding proteins (SSBs) that selectively bind to ssDNA, and that play important roles in replication, recombination and repair of DNA [16]. One of the key tasks of SSBs is to prevent the formation of secondary structure in ssDNA [17]. From the thermodynamical point of view one would therefore expect SSBs to be of an effectively helix-destabilizing nature, and thus to lower T_m [18]. However, it was found that neither the gp32 protein from the T4 phage nor *E. coli* SSBs do [18, 19, 17]. An explanation to this apparent paradox was suggested to consist in a kinetic block, i.e., a kinetic regulation such that the rate constant for the binding of SSBs is smaller than the one for bubble closing [19, 20]. This hypothesis could recently be verified in extensive single molecule setups using mechanical overstretching of dsDNA by optical tweezers in the presence of T4 gene 32 protein [21, 22].

In what follows, we develop a quantitative dynamical model for the description of the competition between SSB-binding and DNA-bubble fluctuations. Based on the free energy of the Poland–Scheraga approach to DNA-melting, we establish a master equation that governs the size change of a bubble simultaneously to the number change of bound proteins. In comparison to previous approaches to DNA-bubble fluctuations in the absence of binding proteins [23, 24], our discrete formalism allows for the explicit consideration of the cooperative bubble initiation brought about by the activation factor σ_0 due to the breaking of stacking interactions [5–7, 11]. We consider cases of fast and vanishing binding/unbinding of SSBs, for which the $(2+1)$-dimensional description (coordinates m and n plus time t) can be reduced to $(1+1)$ dimensions, as well as the general case. In particular, we quantify the free energy landscape and the lifetime of a bubble as influenced by the presence of SSBs. After setting up the general framework in the next section, we proceed by defining the transfer coefficients in our master equation. We then discuss three cases of SSB-bubble dynamics by means of the relaxation time and the effective free energy, before drawing our conclusions. In the four appendices, we collect some more formal derivations.

2. General two-dimensional framework

Bearing in mind typical *in vitro* studies on designed DNA with a poly-(AT) bubble zone[2], we consider a dsDNA segment with M internal base-pairs that is clamped at the two ends and immersed in a bath of single-stranded DNA binding proteins (or single-stranded DNA binders, SSBs). Apart from the connection to the experimental setup, we note that at temperatures well below the melting temperature T_m, a one-bubble picture is a good approximation due to the rather high bubble initiation barrier; see appendix D for details. A schematic diagram of the system we have in mind is depicted in figure 1. We assume that there are *two* slow variables in the problem: m, which is the number of broken base-pairs in the bubble; and n, which is the number of SSBs bound to the bubble. The aim is to understand how m and n change in time.

The individual dynamics of bubble breathing, and SSB attachment and detachment, are stochastic processes, and hence must be described in terms of a Langevin equation, or a probability distribution. Moreover, bubble fluctuation and SSB binding dynamics are coupled

[2] In the experimental protocol of [25], different variants of AT-zones were compared, some of which could in principle produce secondary structure. Dynamically, no difference in behaviour could be discerned, such that it seems legitimate to ascribe to the homopolymer, poly-(AT) bubble a somewhat more general meaning; analogously, we could have a poly-(GC) bubble with physical loop-clamps in mind, that would show similar breathing at slightly more elevated T. The investigation of the dynamical properties of heteropolymers is left for future studies.

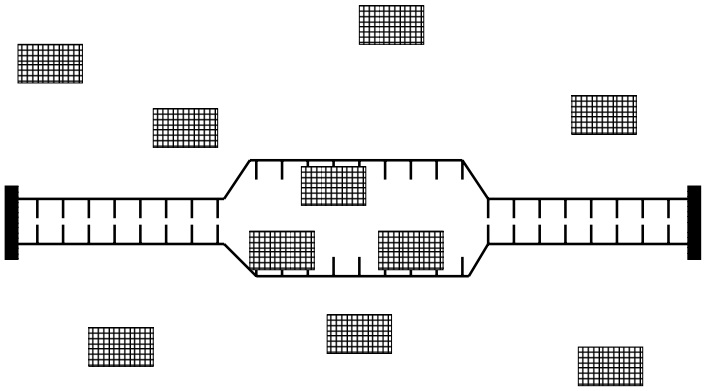

Figure 1. DNA bubble in a region which is clamped at the ends, and immersed in a bath of single-stranded DNA binding proteins (or single-stranded DNA binders, SSBs).

processes. We introduce a joint probability distribution $P(m, n, t)$ for the probability of having a bubble of size m with n bound SSBs at time t, and proceed by assuming that transitions can occur only one step in the forward or backward direction (i.e., m or n are each increased or decreased by one). This assumption is reasonable since the dynamics for nanometre-sized particles moving in a liquid medium is strongly overdamped (low Reynolds number motion) and zipping requires the hinge-like guidance by a zipped vicinal base-pair; this multistep kinetics picture is also supported by experiment [25]. The transition probability for forward and backward motion is described by the transfer coefficients $\mathsf{t}^+(m, n)$, $\mathsf{t}^-(m, n)$ (governing an increase/decrease in bubble size m) and $\mathsf{r}^+(m, n)$, $\mathsf{r}^-(m, n)$ (controlling an increase/decrease in the number n of bound SSBs), respectively. We limit the study to bubbles, that are initiated in the middle part of the potential DNA bubble zone, for temperatures below the melting temperature in order to avoid complications associated with end effects [26]. We then describe the dynamics through a master equation for the probability distribution $P(m, n, t)$,

$$\begin{aligned}\frac{\partial}{\partial t}P(m,n,t) &= \mathsf{t}^+(m-1,n)P(m-1,n,t) + \mathsf{t}^-(m+1,n)P(m+1,n,t)\\ &\quad - (\mathsf{t}^+(m,n) + \mathsf{t}^-(m,n))P(m,n,t)\\ &\quad + \mathsf{r}^+(m,n-1)P(m,n-1,t) + \mathsf{r}^-(m,n+1)P(m,n+1,t)\\ &\quad - (\mathsf{r}^+(m,n) + \mathsf{r}^-(m,n))P(m,n,t). \qquad (1)\end{aligned}$$

Notice that the maximum number of bound SSBs, for a given bubble size m, is

$$n^{\max}(m) = 2[m/\lambda], \qquad (2)$$

where λ is the number of bases occupied by a binding protein if bound onto one of the single-strands, i.e., the size of an SSB measured in units of a single base. For SSBs, typical sizes λ are of the order of 10 bases. Moreover, $[\cdot]$ is the Landau bracket returning the integer value of the argument, and the maximum number of SSBs *per strand* is $n^{\max}/2$. Thus, the bubble size m constrains n such that $0 \leqslant n \leqslant n^{\max}(m)$. For a completely unzipped DNA region, $m = M$, the maximum number of bound SSBs is $N = n^{\max}(M) = 2[M/\lambda]$. The (m, n)-lattice, on which the above problem is defined, is illustrated in figure 2.

At this point a few words concerning the boundary conditions we want to impose on equation (1) are in order. We require reflecting boundary conditions at $m = M$, i.e.,

$$\mathsf{t}^+(m = M, n) = 0. \qquad (3)$$

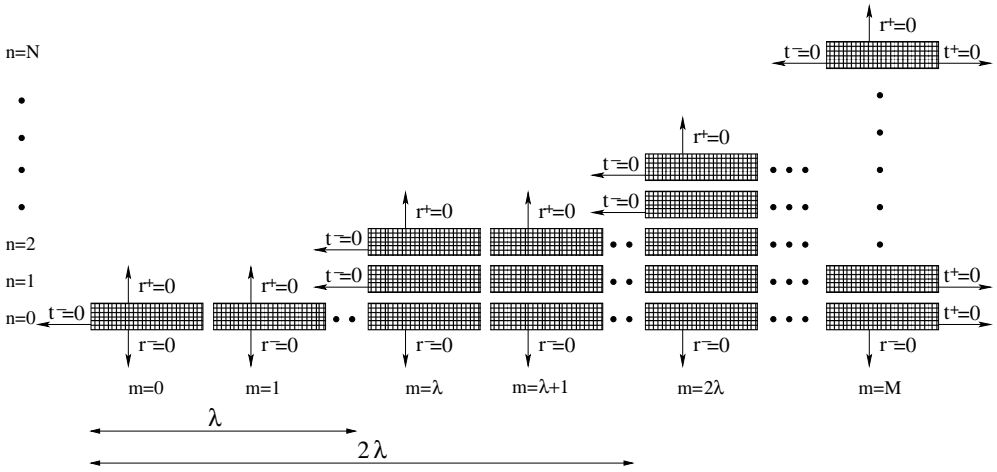

Figure 2. The lattice, on which the general two-dimensional master equation is solved; see equation (1) (compare also to the corresponding eigenvalue equation (12)). The boundary conditions given by equations (3)–(7) are also schematically illustrated. The maximum n-value is $N = n^{\max}(M) = 2[M/\lambda]$.

This condition states that, due to the clamping, the bubble size cannot increase beyond $m = M$. Analogously,

$$\mathsf{t}^{-}(m = 0, n = 0) = 0. \tag{4}$$

Furthermore, it is obvious that the bubble size m and the number n of bound SSBs are coupled in the sense that if m equals an integer multiple of λ and $n = n^{\max}(m)$ or $n = n^{\max}(m) - 1$, then at least one of the two strands is completely filled and it is not possible for the bubble to decrease in size, i.e., we have

$$\begin{aligned} &\mathsf{t}^{-}(m = k\lambda, n = n^{\max}(m)) = 0, \\ &\mathsf{t}^{-}(m = k'\lambda, n = n^{\max}(m) - 1) = 0, \end{aligned} \tag{5}$$

where $k, k' = 0, 1, \ldots, M$, as is illustrated in figure 2. Similarly, for the transfer coefficients for the number of SSBs attached to the bubble, it has to be guaranteed that

$$\mathsf{r}^{-}(m, n = 0) = 0, \tag{6}$$

and

$$\mathsf{r}^{+}(m, n = n^{\max}(m)) = 0. \tag{7}$$

These conditions declare that if there are no SSBs bound to the DNA, no unbinding can take place; and that when the bubble is completely filled with SSBs, no additional binding proteins can bind[3]. We note the fact that the *form* of the coupled DNA bubble and SSB

[3] For equation (1) to be completely specified we need to describe the transfer coefficients just outside the lattice as well; from the detailed balance condition (17) and the boundary conditions (3)–(7), we have

$$\mathsf{t}^{-}(m = M + 1, n) = 0 \tag{8}$$

and

$$\mathsf{t}^{+}(m = k\lambda - 1, n^{\max}(m) < n \leqslant n^{\max}(m + 1)) = 0, \tag{9}$$

as well as

$$\mathsf{r}^{+}(m, n = -1) = \mathsf{r}^{-}(m, n = n^{\max}(m) + 1) = 0. \tag{10}$$

dynamical equations as contained in equation (1), together with the boundary conditions (3)–(7), necessarily have to be imposed whenever m and n are the relevant slow variables of the problem (as long as the dynamics is strongly damped). Explicit expressions for the transfer coefficients for interior points on the lattice (see figure 2) will be constructed in the next section, based on the detailed balance condition.

We are mainly interested in the dynamics caused by the fluctuations. Instead of the master equation in time as given by equation (1), we use the corresponding eigenvalue equation to obtain the mode relaxation. To this end, we decompose the probability density $P(m, n, t)$ into the eigenmodes [27]

$$P(m, n, t) = \sum_p c_p Q_p(m, n) \exp(-\eta_p t) \tag{11}$$

where the index p labels different eigenmodes. The coefficients c_p are determined by the initial condition. Inserting this expansion into equation (1), we arrive at the eigenvalue equation

$$\begin{aligned} &\mathsf{t}^+(m-1,n)Q_p(m-1,n) + \mathsf{t}^-(m+1,n)Q_p(m+1,n) \\ &\quad - (\mathsf{t}^+(m,n) + \mathsf{t}^-(m,n))Q_p(m,n) \\ &\quad + \mathsf{r}^+(m,n-1)Q_p(m,n-1) + \mathsf{r}^-(m,n+1)Q_p(m,n+1) \\ &\quad - (\mathsf{r}^+(m,n) + \mathsf{r}^-(m,n))Q_p(m,n) = -\eta_p Q_p(m,n), \end{aligned} \tag{12}$$

with eigenvalues η_p and eigenvectors $Q_p(m, n)$. We prefer using an eigenvalue approach (spectral representation) to the present problem rather than solving the master equation in real time, since the eigenvalue approach avoids time discretization problems. Also, the eigenvalue approach directly provides the spectral density of relaxation times, which can be experimentally measured (see below). In general we have to solve equation (12) numerically (although limiting behaviour is analytically obtained in a later section), the procedure of which is described in appendix A.

A convenient, experimentally accessible, quantity is the equilibrium bubble size autocorrelation function $A(t) = \langle \Delta m(t) \Delta m(0) \rangle = \langle m(t)m(0) \rangle - (\langle m \rangle_{\mathrm{eq}})^2$, where $\langle \cdots \rangle_{\mathrm{eq}}$ denotes the equilibrium value. This correlation function can be written as[4]

$$A(t) = \int \mathrm{d}\tau \exp(-t/\tau) f(\tau) \tag{13}$$

where we have introduced the relaxation time spectrum

$$f(\tau) \equiv \sum_{p \neq 0} A_p \delta(\tau - \tau_p) \tag{14}$$

with amplitudes

$$A_p \equiv \left(\sum_{m,n} m Q_p(m, n) \right)^2 \tag{15}$$

[4] By definition, we have

$$\langle m(t)m(0) \rangle = \sum_{n,n'} \sum_{m,m'} m m' P(m, n, t; m', n'),$$

where $P(m, n, t; m', n')$ is the probability density for the initial condition $m = m'$ and $n = n'$. By using the eigenvalue expansion, equation (12), the orthogonality relation [27]

$$\sum_{m,n} \frac{Q_{p'}(m, n) Q_p(m, n)}{Q_0(m, n)} = \delta_{pp'},$$

together with the assumption that the system initially is in equilibrium, $P(m, n, t = 0; m', n') = \delta_{mm'} \delta_{nn'} Q_0(m, n)$, one straightforwardly proves equations (13)–(15).

and $\tau_p \equiv 1/\eta_p$. The general theory for processes as described by equation (12) guarantees that there is one zero eigenvalue, $\eta_0 = 0$, and that all other $\eta_p > 0$, $p = 1, \ldots, M$ [27]. We label the eigenvalues such that $\eta_0 = 0 < \eta_1 < \eta_2 < \ldots$. The eigenvector corresponding to the eigenvalue $\eta_0 = 0$ is just the equilibrium distribution, i.e., $Q_0(m, n) = P^{\mathrm{eq}}(m, n)$, where $P^{\mathrm{eq}}(m, n) = \lim_{t\to\infty} P(m, n, t)$. Note that we exclude the eigenvalue $\eta_0 = 0$ in the spectrum (14). Throughout this study, we will show examples of spectral densities $f(\tau)$, in order to illustrate the relaxation behaviour of DNA bubbles under different conditions. A related quantity of interest is the longest relaxation time, $\tau_{\mathrm{relax}} \equiv 1/\eta_1$, which sets the characteristic time for system equilibration. τ_{relax} can, for instance, be studied as a function of the dimensionless number γ, the ratio between the rate constant for SSB unbinding and the rate constant for base-pair closing, the statistical weight u, and the SSB binding strength κ. We are going to introduce these parameters in the next section in connection with the models for the bubble partition function and the statistical weights for SSB (un)binding[5].

3. Definition of the transfer coefficients $\mathsf{t}^{\pm}$ and $\mathsf{r}^{\pm}$

In the previous section we introduced the master equation (1) involving the transfer rates $\mathsf{t}^{\pm}$ and $\mathsf{r}^{\pm}$, and we specified the boundary conditions. We now proceed by constructing these transfer rates based on the assumption that m and n are the slow variables of the system, that we require to eventually settle into equilibrium. This is guaranteed by imposing detailed balance on the dynamics, i.e., demanding that the number of transitions per time from state m into state $m-1$ balances the number of transitions per time from state $m-1$ into state m (and similarly for n). The condition of detailed balance therefore reads

$$\mathsf{t}^{+}(m-1, n)Z(m-1, n) = \mathsf{t}^{-}(m, n)Z(m, n) \tag{17}$$

and

$$\mathsf{r}^{+}(m, n-1)Z(m, n-1) = \mathsf{r}^{-}(m, n)Z(m, n), \tag{18}$$

where $Z(m, n)$ is the (equilibrium) statistical weight for a given m and n as defined below. The dynamical equation (1) together with the rate constants of the form above make sure that the equilibrium (in the Boltzmann sense) distribution is reached for sufficiently long times. We point out that the detailed balance conditions above do not fully determine (only constrain) the transfer coefficients. Physically reasonable, explicit expression are constructed below.

Let us first consider the statistical weight $Z(m, n)$ to find the system in state (m, n), from which in turn we are going to define the transfer rates. Using the one-bubble approximation (see appendix D), we write

$$Z(m, n) = Z^{\mathrm{bubble}}(m)Z^{\mathrm{bind}}(m, n) \tag{19}$$

as a product of the (free) bubble weight and the weight for the degrees of freedom of the SSB binding states. We adopt the Zimm–Poland–Scheraga model for a denaturation bubble according to which the rather stiff dsDNA segments are considered to be in a zero-entropy state and to carry the enthalpic contributions from the bound base-pairs, whereas the flexible ssDNA-bubbles correspond to entropy reservoirs; see [5, 11, 28, 29] for details. The bubble statistical weight then becomes

$$Z^{\mathrm{bubble}}(m) = \sigma_0 u^m (1+m)^{-c}, \qquad \text{for } m \geqslant 1, \tag{20}$$

[5] Furthermore, it is sometimes instructive to investigate the difference between the longest relaxation time and the second longest, τ_2, i.e.,

$$\Delta\tau = 1/\eta_1 - 1/\eta_2. \tag{16}$$

The quantity $\Delta\tau$ contains information that allows a rough discrimination between 'multi-exponential' and 'exponential' relaxation dynamics.

and $Z^{\rm bubble}(m=0)=1$. We have above introduced the following physical parameters characterizing the bubble breathing: (i) the statistical weight

$$u=\exp(-\beta E), \tag{21}$$

with $\beta=1/k_{\rm B}T$, where $k_{\rm B}$ is the Boltzmann constant and T the temperature of the surrounding solution. Moreover, $E=E(T)$ denotes the free energy associated with breaking a hydrogen bond (note that this energy differs between A–T and G–C bonds) [6]; (ii) the non-universal prefactor σ_0 that measures the loop initiation energy associated with breaking the stacking interactions necessary to open the bubble. Determined from DNA melting data, σ_0 typically varies between 10^{-3} and 10^{-5}, corresponding to an activation barrier of some 7–12 $k_{\rm B}T$ [5–7, 11]; (iii) and the loop closure exponent c that stems from the entropy loss upon loop formation of a polymer of length m, m^{-c} [30]. To take care of persistence length effects in the ssDNA bubble, we use the modified form $(1+m)^{-c}$ [31]. The exponent c, for a polymer ring assumes the value 3/2 in the Gaussian (phantom) chain limit, and $c\approx 1.76$ for self-avoiding chains [30][6]. For an infinite chain in the absence of SSBs, the melting temperature $T_{\rm m}$ is defined by $E(T_{\rm m})=0$, or, $u=1$. We point out that, following [41], including an external torque τ has the effect of changing $u\to u\exp(\beta\theta_0\tau)$, where $\theta_0=2\pi/10.35$ is the twist angle per base of the double helix. Thus, experimentally, u can be changed by an applied external torque on the DNA rather than by changing the temperature. This can be achieved in the optical tweezers setup applied in [21], in which the DNA is pulled until reaching the overstretching transition; due to the possibility that the tweezers beads can rotate freely, the dsDNA is progressively unwound.

The statistical weight $Z^{\rm bind}(m,n)$ for SSB (un)binding in equation (19) is given by

$$Z^{\rm bind}(m,n)=\Omega^{\rm bind}(m,n)\kappa^n, \tag{22}$$

where $\Omega^{\rm bind}(m,n)$ counts the possible configurations that can be explored by n SSBs bound to a bubble consisting of m broken base-pairs, and κ is the binding strength. The number of degrees of freedom,

$$\Omega^{\rm bind}(m,n)=\sum_{n_1=0}^{n}\omega^{\rm bind}(m,n_1)\omega^{\rm bind}(m,n-n_1)\Big|_{n_1\leqslant n^{\rm max}/2}^{n-n_1\leqslant n^{\rm max}/2}, \tag{23}$$

can be expressed in terms of the measure $\omega^{\rm bind}(m,n_1)$, which counts the number of ways of arranging n_1 SSBs onto m binding sites, i.e., onto the m available bases on one of the two single-strands in the bubble. The conjugated $\omega^{\rm bind}(m,n-n_1)$ counts the degrees of freedom on the second single-strand. The explicit expression for $\omega^{\rm bind}$ becomes (compare [42–44])

$$\omega^{\rm bind}(m,n)=\binom{m-(\lambda-1)n}{n}=\frac{(m-(\lambda-1)n)!}{n!(m-\lambda n)!}. \tag{24}$$

Throughout this study we assume that the SSB binding is effectively insensitive to DNA base structure differences of the two strands. The quality of the binding is characterized by the dimensionless binding strength [44]

$$\kappa=c_0K^{\rm eq}=\exp\{\beta|E_{\rm bind}|+\ln(v_0c_0)\}, \tag{25}$$

where

$$K^{\rm eq}=v_0\exp(\beta|E_{\rm bind}|) \tag{26}$$

[6] The value of c in fact determines the order of the phase transition of DNA melting within the Poland–Scheraga approach [32]: if $c>2$, the transition is first order, while for $c<2$, it is of second order. Whereas there exist models that predict a $c>2$ [33, 34], this issue has received some debate [35–40]. The bubble *dynamics* appears fairly insensitive to whether c is slightly above or below 2 [23, 24]. In contemporary bioinformatics, the choice $c\approx 1.76$ appears standard, compare, for instance, [6].

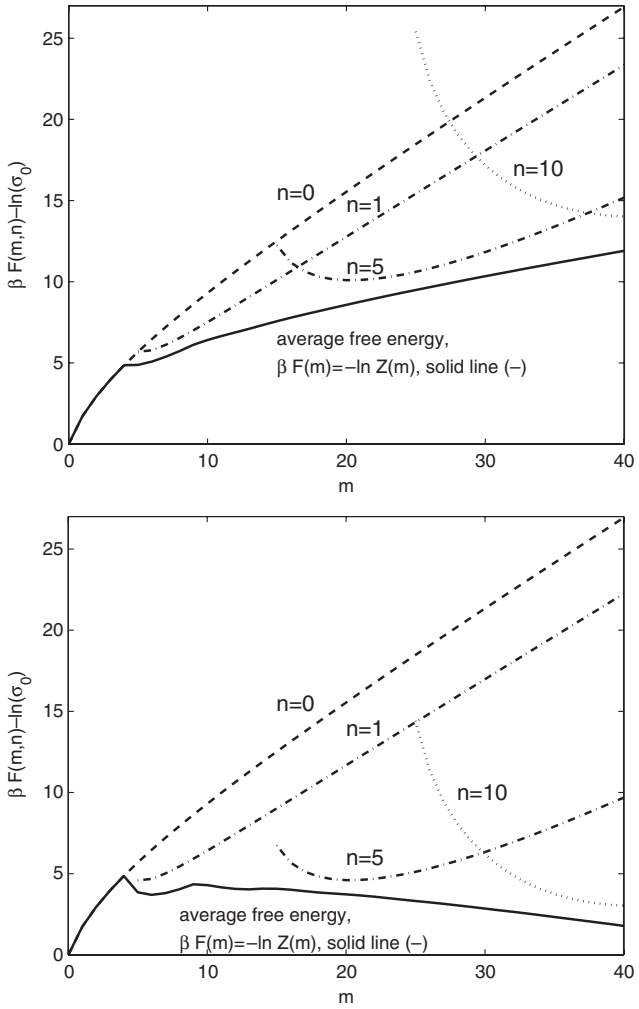

Figure 3. The 'free energy' $F(m, n)$ for fixed number n of bound SSBs as function of bubble size m for various n. We consider (top) relatively weak binding (or high SSB concentration), $\kappa = 0.5$; and (bottom) stronger binding, $\kappa = 1.5$. Notice that, for small n, $F(m, n)$ increases rapidly with increasing m, i.e., if only a few SSBs are bound there is a tendency for the bubble to close. When n is larger, $F(m, n)$ has a smaller or negative slope. The solid curves represent the averaged free energy $\beta F(m) = -\ln Z(m)$, where $Z(m)$ is given in equation (56). $F(m)$ corresponds to the free energy profile relevant for fast SSB dynamics, as derived in section 4.2. Notice that this averaged free energy has a positive (negative) slope for weak (strong) SSB binding for large m, for the parameters used here: $u = 0.6$, $M = 40$ $c = 1.76$, and $\lambda = 5$. In the latter case, the SSB dynamics would therefore lead to full denaturation in an unclamped DNA molecule.

is the equilibrium binding constant, v_0 denotes the typical volume occupied by an SSB, E_{bind} is the binding energy of an SSB to the DNA, and c_0 is the number-per-volume concentration of SSBs in the solution. In figure 3, the 'free energy' $F(m, n) = -k_{\text{B}}T \ln Z(m, n)$ is plotted using the results above. Notice that if n is small (large), then $\beta F(m, n)$ increases rapidly (slowly) with increasing m. The solid curves represent the averaged free energy $\beta F(m) = -\ln Z(m)$, where $Z(m)$ corresponds to the adiabatic limit as given in equation (56); $F(m)$ corresponds to the free energy profile relevant for fast SSB dynamics derived in section 4.2.

Having defined explicit expressions for the statistical weights $Z(m,n)$, we are ready to construct physically realistic transfer coefficients. For the bubble opening rate, that we assume to be independent of the number n of bound SSBs[7], we choose

$$\mathrm{t}^{+}(m,n) = \mathrm{t}^{+}(m) = k(1+m)^{-\mu}\frac{Z^{\mathrm{bubble}}(m+1)}{Z^{\mathrm{bubble}}(m)}, \tag{27}$$

so that, using equation (20), we arrive at the bubble initiation rate for leaving the ground state $m = 0$,

$$\mathrm{t}^{+}(0) = k\sigma_0 us(0), \tag{28}$$

that includes the cooperativity factor σ_0 to overcome the base-stacking interactions on introducing a bubble ($m = 0 \to m = 1$) into the previously intact double-strand. For general $m \geqslant 1$, we obtain the forward rates

$$\mathrm{t}^{+}(m) = ku(1+m)^{-\mu}s(m), \qquad \text{for } m \geqslant 1. \tag{29}$$

These expressions state that a forward step (i.e., an increase in bubble size from m to $m+1$) is determined by the statistical weight u associated with breaking a base-pair, multiplied by the loop closure factor

$$s(m) = \left(\frac{1+m}{2+m}\right)^{c}. \tag{30}$$

Notice that $s(m) \approx 1$ for large m. The constant factor k represents the typical rate for (un)zipping, and is determined by the quantum chemistry of H-bond formation or breaking. In our phenomenological approach, it is a free parameter.

Finally, in equation (28), we also introduced an additional factor $(1+m)^{-\mu}$ for the following reason. A local zipping or unzipping event not only involves the pulling or pushing of the immediate base pairs at the zipper fork, but, due to the connectivity of the flexible ssDNA strands, by necessity involves the motion of several Kuhn segments adjacent to the zipper fork. The effective number of segments engaged in an (un)zipping event depends on the bubble size m, as we are going to show now. In a phenomenological argument, we include this collective effect in the spirit of the scaling approach brought forth in [45]. We note that the validity of this argument should be verified by simulations; however, for completeness we include it in our dynamical formulation. To this end, consider the plane that is perpendicular to the plane spanned by the open base-pair, and that contains the local tangent to the double-strand segment connected to the zipper fork[8]. Following one of the strands from the zipper fork along its contour until the first point where it returns to the plane, then the length of this hook is proportional to m^{ν}. This corresponds to the random walk character of a random polymer: in a Gaussian chain of length M, it takes on average $\simeq M^{1/2}$ steps to return to the plane; in a self-avoiding chain, the analogous number is expected to scale like $\simeq M^{\nu}$. This quantity scales like the gyration radius $R_{\mathrm{g}} \simeq M^{\nu}$, which makes sense, as the typical length of a hook must lie between the monomer size and the overall chain length, and R_{g} is the only such length scale available. Pulling at the end of such a hook results in the motion of this hook, but does not propagate to the remainder of the strand because physically tension cannot be transferred beyond it (imagine pulling at the end of a rope immersed, in a coiled up configuration, into water). The presence of this hook effect reduces the closing rate, and, by detailed balance, the

[7] It may be argued that, in contrast to SSBs bound in the 'middle' of a bubble, SSBs bound next to one of the zipper forks of the bubble may ease the opening of further base-pairs. Moreover, some types of SSBs are known to bind cooperatively. Such effects can, in principle, be included in the model. For the sake of simplicity, we neglect such effects in this study.

[8] This plane appears a natural choice given the hinge-like structure of the zipper fork.

forward rate involves pushing away all segments in a hook. We take $\mathsf{t}^+(m,n) \propto k(1+m)^{-\nu}$, which should be appropriate for Rouse-like dynamics for which the total friction is obtained by summing up the friction from different monomers (hydrodynamic interactions are hence neglected). Note that we included the shift by 1 in the factor $(1+m)^{-\nu}$, to keep consistent with the notation for the loop closure according to [31]. We also note that the inclusion of hydrodynamical interactions may change the exponent μ. The bubble size being of the order of a few tens of broken base-pairs, at temperatures well below T_{m} this truly polymeric effect may be expected to be of lesser significance. At higher temperatures leading to bubbles of the size of a few hundreds of broken base-pairs, i.e., roughly a hundred Kuhn segments, it may become relevant. We therefore leave the exponent μ in expression (29) as a free parameter, that has to be determined by more detailed studies[9].

For the bubble closing rate, we choose the expression

$$\mathsf{t}^-(m,n) = km^{-\mu}\frac{Z^{\mathrm{bind}}(m-1,n)}{Z^{\mathrm{bind}}(m,n)} = km^{-\mu}\frac{\Omega^{\mathrm{bind}}(m-1,n)}{\Omega^{\mathrm{bind}}(m,n)}. \tag{31}$$

The transfer coefficient $\mathsf{t}^-(m,n)$ is thus determined by the ratio between the number of available configurations (for a given n) between bubbles of size $m-1$ and m. This choice of $\mathsf{t}^-(m,n)$ realistically describes the fact that a region, which is almost fully occupied by SSBs (n large), is less likely to decrease in size due to steric constraints. We note that, in general, $\mathsf{t}^-(m,n) \leqslant k$, and that $\mathsf{t}^-(m,n=0) = km^{-\mu}$. The $m^{-\mu}$-factor again accounts for the fact that the zipping process, due to the connectivity of the chain, involves the motion of several bases[10]. Although the detailed balance condition (17) does not uniquely determine the transfer coefficients, we believe that the above choice is physically reasonable: the rate k has the simple physical interpretation as the characteristic inverse DNA zipping time in the absence of SSBs for a bubble of size $m=1$. Also, with the expressions above for the transfer coefficients, the forward rate has a simple Arrhenius form (being proportional to the statistical weight $u = \exp(-E/k_{\mathrm{B}}T)$, i.e., to the Boltzmann factor associated with breaking a base-pair interaction).

Let us now consider the transfer coefficients associated with the increase or decrease of n by one. We choose

$$\begin{aligned}\mathsf{r}^+(m,n) &= (n+1)q\frac{Z(m,n+1)}{Z(m,n)} = (n+1)\gamma k\frac{Z^{\mathrm{bind}}(m,n+1)}{Z^{\mathrm{bind}}(m,n)} \\ &= (n+1)\gamma k\kappa\frac{\Omega^{\mathrm{bind}}(m,n+1)}{\Omega^{\mathrm{bind}}(m,n)},\end{aligned} \tag{32}$$

and

$$\mathsf{r}^-(m,n) = nq = n\gamma k \tag{33}$$

where q is the rate constant associated with the unbinding of a single protein. We have for convenience defined a dimensionless parameter

$$\gamma \equiv \frac{q}{k} \tag{34}$$

as the ratio between the rate constant for SSB unbinding and the rate constant for bubble closing. When $\gamma \ll 1$ ($\gamma \gg 1$) the SSB binding is much slower (faster) than the typical bubble base-pair closing time. In order to show that the above choice of transfer coefficients

[9] Additional corrections to the dynamical behaviour may result from the dynamics of the two vicinal DNA double-strands, into which the bubble is embedded. This will be of importance to long DNA molecules, whereas for the short experimental DNA construct we have in mind, this effect should again be of higher order.

[10] We note that the boundary conditions as contained in equation (5) are implicit in equation (31); the number of ways of binding $n^{\max}(m) = 2[m/\lambda]$ or $n^{\max}(m) - 1$ SSBs to a bubble with $m-1$ broken base-pairs is zero if m equals an integer number of λ, i.e., we have $\Omega^{\mathrm{bind}}(k\lambda - 1, n^{\max}(m)) = 0$ and $\Omega^{\mathrm{bind}}(k\lambda - 1, n^{\max}(m) - 1) = 0$.

is physically realistic, we consider the case $\lambda = 1$, for which the above expression reduces to standard results. For λ, it is straightforward to show that $\mathsf{r}^+(m,n) = q\kappa(2m-n)$. This means that the rate for the process $n \to n+1$ is proportional to the binding strength $\kappa = c_0 K^{\rm eq}$ (i.e., proportional to the concentration of SSBs in solution), and to the number of unoccupied binding sites $(2m-n)$, as it should (see [27], chapter VI). For the general expression, equation (32), the forward transfer coefficients $\mathsf{r}^+(m,n)$ are proportional to κ, multiplied with a correction factor, that accounts for the fact that for a given m the number of allowed configuration changes with the number n of bounds SSBs. The transfer coefficients $\mathsf{r}^-(m,n)$, as given by equation (33), correspond to the backward rate being proportional to the number of bound SSBs [27][11]. The above choice, equations (32) and (33), for the rate coefficients satisfies the detailed balance condition (18) as it should.

4. Eigenvalue spectrum of the master equation

In this section, we explore two limiting cases of the general dynamics contained in the $(2+1)$-dimensional master equation (1), these being (1) the case of no binding (or absence of SSBs), and (2) the case of fast binding. In the former, the description naturally reduces to a $(1+1)$-dimensional master equation, whereas in the latter we perform a procedure of adiabatically removing the fast SSB variable, revealing a 'dressed' $(1+1)$-dimensional description with an effective bubble free energy. (3) In the last subsection, we then embark on the general case. For the case of vanishing SSB dynamics we can solve the eigenvalue equation corresponding to the master equation analytically, if the loop closure effects are neglected. We compare the approximate results to the general numerical solution of the full $(2+1)$-dimensional problem. The focus is on the spectral density of relaxation times for the combined DNA bubble and SSB dynamics.

4.1. Vanishing SSB dynamics

Let us at first consider the bubble dynamics in the absence of SSBs, or, equivalently, of SSBs with vanishing binding strength, $\kappa = 0$, or extremely slow binding dynamics. In this limiting case, our discrete model features a few advantages over the continuum model of [24], and therefore a few words on this case are in order. Apart from being the more physical approach, given the discrete nature of both (un)zipping base-pair by base-pair and SSB (un)binding, the discrete description allows the explicit incorporation of the cooperativity parameter σ_0 for loop initiation. This latter feature is not possible in the continuum Fokker–Planck approach, as there the drift term involves the force as given by the *gradient* of the free energy.

The dynamical equation in the absence of SSB binding can be obtained from the original, full description, by putting $n = 0$ in equation (1) (we assume that initially no SSBs are bound to the DNA), to find

$$\frac{\partial}{\partial t}\bar{P}(m,t) = \bar{\mathsf{t}}^+(m-1)\bar{P}(m-1,t) + \bar{\mathsf{t}}^-(m+1)\bar{P}(m+1,t) - (\bar{\mathsf{t}}^+(m) + \bar{\mathsf{t}}^-(m))\bar{P}(m,t), \quad (35)$$

where we introduced the shorthand notations $\bar{P}(m,t) = P(m,n=0,t)$ and $\bar{\mathsf{t}}^{\pm}(m) = \mathsf{t}^{\pm}(m,n=0)$. The boundary conditions become $\bar{\mathsf{t}}^+(-1) = \bar{\mathsf{t}}^-(0) = 0$ and $\bar{\mathsf{t}}^-(M+1) = \bar{\mathsf{t}}^+(M) = 0$. While the forward transfer coefficients are $\bar{\mathsf{t}}^+(m) = \mathsf{t}^+(m)$, with $\mathsf{t}^+(m)$ given in equations (28) and (29), the backward transfer coefficients now read $\bar{\mathsf{t}}^-(m) = km^{-\mu}$, i.e.,

[11] We note that the boundary condition given by equation (7) is implicit in equation (32); it is not possible to attach $n^{\max}(m)+1$ SSBs to a bubble of size m, and hence $\Omega^{\rm bind}(m, n^{\max}(m)+1) = 0$. Also, the boundary condition (6) is implicit in equation (33).

they are determined by the constant rate k and the hook exponent μ. The eigenvalue equation corresponding to equation (35) has the comparatively simple structure

$$\bar{t}^{+}(m-1)\bar{Q}_p(m-1)+\bar{t}^{-}(m+1)\bar{Q}_p(m+1)-(\bar{t}^{+}(m)+\bar{t}^{-}(m))\bar{Q}_p(m)=-\eta_p\bar{Q}_p(m), \quad (36)$$

where $p = 0, 1, \ldots, M$. Based on the eigenvectors and eigenvalues of this equation, any quantity of interest may be constructed. Below, we show that we can solve the eigenvalue equation (36) analytically in the absence of the loop closure factor, but with arbitrary cooperativity parameter σ_0. In the general case, the solution is obtained numerically.

Neglecting loop closure (i.e., setting $s(m) = 1$ in equations (28) and (29)) as well as the hook effect (i.e., setting $\mu = 0$), but keeping an arbitrary σ_0, the transfer coefficients take on the form

$$\begin{aligned}&\bar{t}^{+}(0)=k\sigma_0 u,\\ &\bar{t}^{+}(m)=ku, \qquad \text{for } m\geqslant 1,\\ &\bar{t}^{-}(m)=k,\end{aligned} \quad (37)$$

and, apart from the initiation term, correspond to constant drift-diffusion in bubble size space. The eigenvalue equation (36) with these rate coefficients can be solved analytically by using the orthogonal polynomial approach described in appendix B; in general, an eigenvalue problem of the type considered here can be transformed into a recurrence relation for the orthogonal polynomials $\bar{\Psi}_m$, $m = 0, \ldots, M$. Through these polynomials eigenvalues and eigenvectors can be constructed. Accordingly, we solve equations (B.8) and (B.9) for $\bar{\Psi}_m$, using the transfer coefficients (37). We obtain (for $m = 1$ in equation (B.9))

$$\bar{\Psi}_1(\eta_p)=-\eta_p+k(1+\sigma_0), \quad (38)$$

while for $m \geqslant 2$, we have the three-term recurrence relation

$$\bar{\Psi}_m(\eta_p)+(\eta_p-k(u+1))\bar{\Psi}_{m-1}(\eta_p)+k^2u\bar{\Psi}_{m-2}(\eta_p)=0. \quad (39)$$

We follow [46], and introduce the ansatz

$$\eta_p=k[u+1-2u^{1/2}\cos\omega_p] \quad (40)$$

for the eigenvalues, so that equation (39) becomes

$$\bar{\Psi}_m(\omega_p)-2ku^{1/2}\cos\omega_p\bar{\Psi}_{m-1}(\omega_p)+k^2u\bar{\Psi}_{m-2}(\omega_p)=0. \quad (41)$$

For $\omega_p \neq 0$, equations (41), (B.9) and (38) are satisfied by the solution

$$\begin{aligned}\bar{\Psi}_m(\omega_p)&=\frac{(k^2u)^{m/2}}{\sin\omega_p}(\sin[(m+1)\omega_p]-\delta\sin[m\omega_p])\\ &=(k^2u)^{m/2}(U_m(\cos\omega_p)-\delta U_{m-1}(\cos\omega_p)),\end{aligned} \quad (42)$$

as may be verified by direct substitution. We have above introduced the Chebyshev polynomials of the second kind $U_m(x)$. The eigenvalues are determined by equation (B.6), leading to

$$g(\omega_p)\equiv\sin[(M+1)\omega_p]-\delta\sin[M\omega_p]=0, \quad (43)$$

with

$$\delta\equiv(1-\sigma_0)u^{1/2}. \quad (44)$$

This, together with equation (40), determines the eigenvalues. Equation (43) can alternatively be written in terms of the Chebyshev polynomials according to

$$U_M(\cos\omega_p)-\delta U_{M-1}(\cos\omega_p)=0. \quad (45)$$

It is straightforward to show that the M solutions to the eigenvalue equation for $\delta > 0$ range in the intervals[12]

$$\frac{(p-1)\pi}{M+1} < \omega_p < \frac{p\pi}{M+1} \tag{46}$$

for $p = 1, \ldots, M$. This result is useful when solving equation (40) numerically. For the special case $\delta = 0$ (i.e., trivial cooperativity $\sigma_0 = 1$), we obtain the standard (compare [46]), explicit solutions $\omega_p = p\pi/(M+1)$ for $p = 1, \ldots, M$[13]. The eigenvectors are given by

$$Q_p = (1, \chi_1(\eta_p), \ldots, \chi_M(\eta_p)), \tag{48}$$

with

$$\begin{aligned}\chi_m(\eta) = \frac{u^{m/2}}{\sin\omega}&\{\sin[(m+1)\omega_p] + u^{-1/2}(u(\sigma_0-1)-1)\sin[m\omega_p] \\ &+ (1-\sigma_0)\sin[(m-1)\omega_p]\}, \\ = u^{m/2}&\{U_m(\cos\omega_p) + u^{-1/2}(u(\sigma_0-1)-1)U_{m-1}(\cos\omega_p) \\ &+ (1-\sigma_0)U_{m-2}(\cos\omega_p)\}\end{aligned} \tag{49}$$

and the solutions above are completed by the eigenvalue–eigenvector pair

$$\begin{aligned}&\eta_0 = 0, \\ &Q_0 = (1, \sigma_0 u, \sigma_0 u, \ldots, \sigma_0 u^M).\end{aligned} \tag{50}$$

These results above completely determine the eigenvalue and eigenvector spectrum. We have reduced the problem of computing the eigenvalues and eigenvectors to an $(M+1)\times(M+1)$ matrix to that of solving equation (43). As before, introducing the relaxation times $\tau_p = 1/\eta_p$, we see from equation (40) that

$$\tau_{\min}(u) \leqslant \tau_p \leqslant \tau_{\max}(u), \tag{51}$$

where

$$\begin{aligned}&\tau_{\min}(u) = k^{-1}(1+u^{1/2})^{-2}, \\ &\tau_{\max}(u) = k^{-1}(1-u^{1/2})^{-2}.\end{aligned} \tag{52}$$

For $p = 1, \ldots, M$, the inequality (52) enforces a constraint on the allowed eigenvalues, for a given u. Combining this statement with equation (46), we find that the largest (smallest)

[12] Equation (46) is proven in the following way: first, consider the cases $p = 2, \ldots, M$. We are interested in investigating the function $g(\omega)$ in the interval $[(p-1)\pi/(M+1),\ p\pi/(M+1)]$. One straightforwardly shows that the $g(\omega)$ have different sign at the end points of this interval; hence $g(\omega)$ passes zero at least once in the considered region. The behaviour of $g(\omega)$ in the interval $[0, \pi/(M+1)]$ requires special considerations. We here limit the discussion to the physically realistic case $\delta > 0$. We then see that $g(\omega = 0) = 0$, $g(\omega = \pi/(M+1)) < 0$, and that the derivative of $g(\omega)$ is positive at $\omega = 0$. These results guarantee that there is at least one solution to equation (43) in the interval $[0, \pi/(M+1)]$. Collecting the results above, we have thus shown that there is *at least* one solution to equation (43) in each of the M intervals $[(p-1)\pi/(M+1),\ p\pi/(M+1)]$. However, since the general theory of W-matrices of the form considered here guarantees that there are M non-zero eigenvalues we can conclude that there is *exactly* one solution in each of the considered interval, which proves equation (46).

[13] A different way of solving equation (43) numerically is to transform it into a polynomial equation. We define $x = \cos\omega_p$. By using equation (1.331) in [47], we can rewrite expression (43) according to

$$\begin{aligned}g(\omega_p) = 2^M \sin\omega_p \Bigg\{ &x^M - \frac{\delta}{2}x^{M-1} + \sum_{k=1}^{[M/2]} (-1)^k 4^{-k} \binom{M-k}{k} x^{M-2k} \\ &- \frac{\delta}{2} \sum_{k=1}^{[(M-1)/2]} (-1)^k 4^{-k} \binom{M-k-1}{k} x^{M-2k-1} \Bigg\} = 0.\end{aligned} \tag{47}$$

The polynomial equation above can be straightforwardly solved on a computer. The roots to this equation, together with equation (40), lead to the eigenvalues. We note that equation (47) follows also from the series expansion of the Chebyshev polynomials of the second kind.

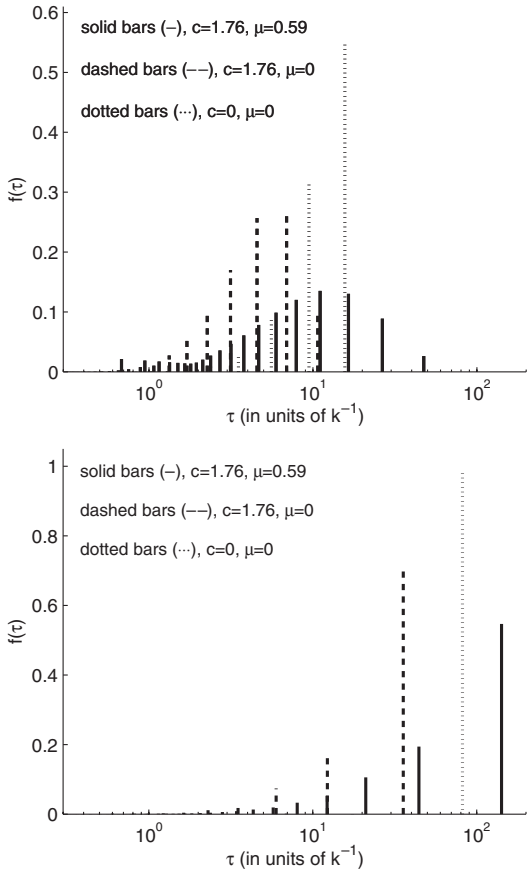

Figure 4. The spectral density $f(\tau)$ of relaxation times τ is plotted for two different u (i.e., for two different temperatures or applied torques). The top (bottom) graph corresponds to $u = 0.6$ ($u = 0.9$), i.e., the top (bottom) graph is for the case when the DNA region is well below (close to) the melting temperature. Notice the logarithmic abscissa. The solid bars correspond to the case with loop closure and the hook effect ($c \neq 0$ and $\mu \neq 0$), and for the dashed bars we included the loop correction, but not the hook effect ($c \neq 0$, $\mu = 0$). The dotted bars are for the case that both the loop closure and hook correction are excluded ($c = 0$ and $\mu = 0$). The corresponding free energy is shown in figure 5. Note that for the case when u is well below the melting temperature (top), the spectrum consists of many closely spaced eigenvalues (multiexponential behaviour), whereas closer to the melting temperature (bottom) there is one (slow) dominant eigenvalue (exponential behaviour). Including the hook correction shifts the spectrum towards longer relaxation times, as expected. The length of the DNA region was taken to be $M = 20$, and the cooperativity parameter is $\sigma_0 = 10^{-3}$.

eigenvalue approaches $\tau_{\max}$ ($\tau_{\min}$) as $M \to \infty$. We also see that the relaxation time $\tau_{\text{relax}} = 1/\eta_1$ is larger for the case of non-zero δ than for the case of $\delta = 0$; thus, the incorporation of a cooperativity parameter smaller than one gives rise, for finite sized DNAs, to an increased relaxation time. We also point out that $\tau_{\max}$ tends to infinity as $u \to 1$, i.e., the relaxation time τ_{relax} is expected to become extremely large and eventually diverges as the melting temperature T_{m} is approached ($u \to 1$), as it should.

In figure 4, the spectral density is shown for two different u-values, while the corresponding free energy profiles are shown in figure 5. Notice that for the case when u is well below the

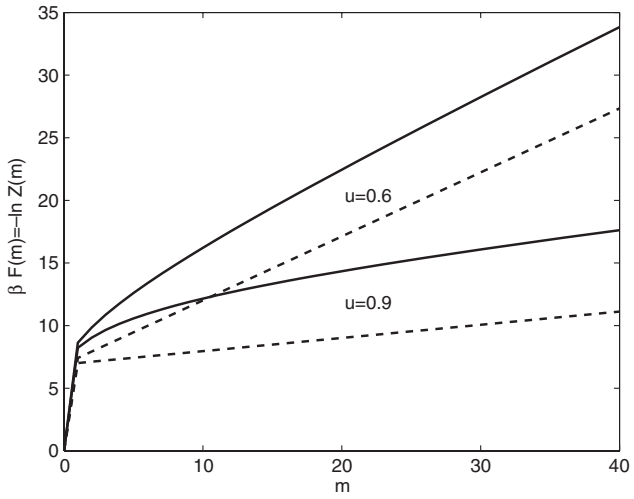

Figure 5. The free energy $F(m)$ as a function of m for the two different u used in figure 4. Solid lines: including the loop closure. Dashed lines: excluding the loop closure. The cooperativity parameter is $\sigma_0 = 10^{-3}$, and the loop closure exponent $c = 1.76$.

melting temperature, the spectrum consists of many 'closely' spaced eigenvalues corresponding to a multiexponential dynamics, whereas closer to the melting temperature there is one (slow) dominant eigenvalue, so that the system essentially behaves like a two-state system, with monoexponential relaxation. Thus, the spectral density is sensitive to the temperature (and in general to external torque, etc), and therefore provides a convenient fingerprint of the system under investigation. Including the hook correction ($\mu \neq 0$), i.e., considering that a closing (opening) base-pair needs to pull (push) that portion of the connected ssDNA until the first significant bend, shifts the spectrum towards longer relaxation times.

In figure 6, the relaxation time τ_{relax} as a function of u is plotted for two different sizes of the bubble region ($M = 10$ and $M = 20$). Note that τ_{relax} increases as u approaches the melting temperature ($u \to 1$), as shown above. Also note that the loop correction becomes more pronounced closer to the melting temperature. We point out that the dependence of τ_{relax} on σ_0 is more pronounced for shorter DNA regions, as discussed above, while the effect of the loop closure is less pronounced for a longer DNA region, as is seen in figure 6.

4.2. Fast binding dynamics, $\gamma \gg 1$

Let us now address the other limiting behaviour, that is, the case of fast binding, when $\gamma = q/k \gg 1$. We define the effective probability distribution $\tilde{P}(m,t) \equiv \sum_n P(m,n,t)$. Using the adiabatic elimination technique described in appendix C, equation (1) reduces to the (1 + 1)-dimensional master equation

$$\frac{\partial}{\partial t}\tilde{P}(m,t) = \tilde{\mathrm{t}}^{+}(m-1)\tilde{P}(m-1,t) + \tilde{\mathrm{t}}^{-}(m+1,n)\tilde{P}(m+1,t) - (\tilde{\mathrm{t}}^{+}(m) + \tilde{\mathrm{t}}^{-}(m))\tilde{P}(m,t), \tag{53}$$

where the effective transfer coefficients are given through

$$\tilde{\mathrm{t}}^{\pm}(m) = k \sum_{n=0}^{n^{\max}(m)} \mathrm{t}^{\pm}(m,n)\frac{Z(m,n)}{Z(m)}. \tag{54}$$

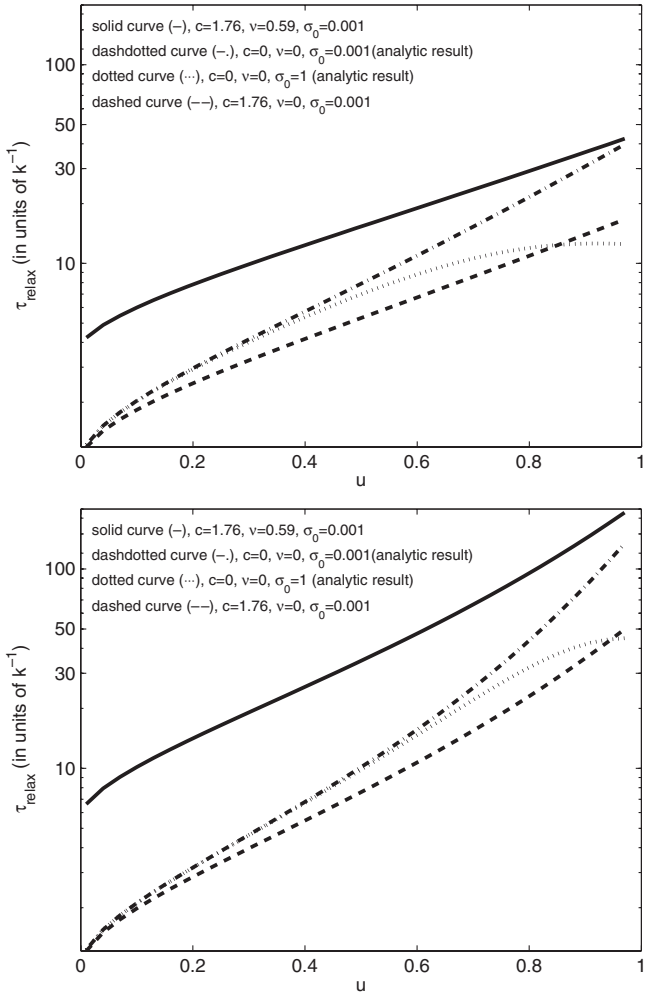

Figure 6. Plot of the relaxation time $\tau_{\rm relax}$ as a function of the statistical weight u (which in turn depends on temperature and applied external torque). Note the logarithmic scale on the vertical axis. Top: short DNA region with $M = 10$. Bottom: longer DNA region with $M = 20$. Also note that the dependences on σ_0 and the loop correction are smaller for the longer DNA region. Including the hook effect ($\mu \neq 0$) significantly increases the relaxation time, as one would expect. As u approaches the melting temperature ($u \to 1$), the relaxation time $\tau_{\rm relax}$ increases dramatically, and finally diverges.

It is straightforward to show that these transfer coefficients satisfy the detailed balance condition (compare appendix C)

$$\tilde{\mathrm{t}}^{+}(m-1)Z(m-1) = \tilde{\mathrm{t}}^{-}(m)Z(m), \tag{55}$$

where

$$Z(m) = Z^{\rm bubble}(m)Z^{\rm bind}(m), \tag{56}$$

and where we have introduced the binding partition function

$$Z^{\rm bind}(m) = \sum_{n=0}^{n^{\rm max}(m)} Z^{\rm bind}(m,n). \tag{57}$$

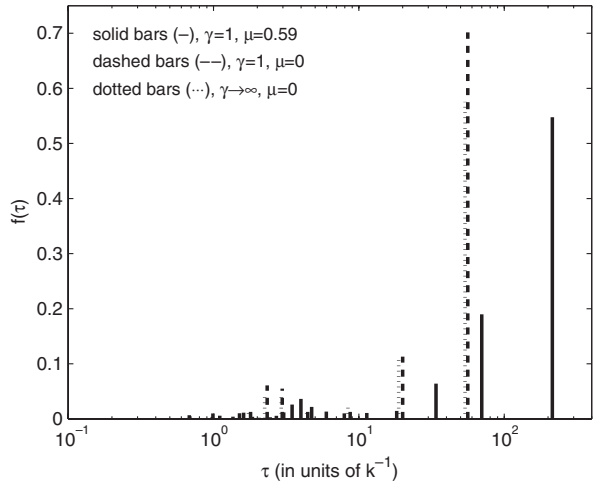

Figure 7. The spectral density $f(\tau)$ of relaxation times τ. The length of the DNA region was taken to be $M = 20$, the cooperativity parameter is $\sigma_0 = 10^{-3}$, the statistical weight $u = 0.6$, the binding strength $\kappa = 0.5$, and the SSB size $\lambda = 5$. The solid (dashed) bars corresponds to the case $\mu \neq 0$ ($\mu = 0$), respectively. The dotted bars denote the adiabatically eliminated ($\gamma \to \infty$) result for $\mu = 0$. Note that including the hook effect, $\mu \neq 0$, increases the characteristic relaxation times considerably.

The quantity $n^{\max}(m) = 2[m/\lambda]$ represents the maximum number of attached binding particles, as before, and the corresponding statistical weight $Z^{\text{bind}}(m, n)$ is given in equation (23). The boundary conditions become $\tilde{t}^{+}(-1) = \tilde{t}^{-}(0) = 0$ and $\tilde{t}^{-}(M+1) = \tilde{t}^{+}(M) = 0$ (i.e., reflecting boundary conditions at the ends). The eigenvalue equation corresponding to equation (53) is

$$\tilde{t}^{+}(m-1)\tilde{Q}_p(m-1) + \tilde{t}^{-}(m+1)\tilde{Q}_p(m+1) - (\tilde{t}^{+}(m) + \tilde{t}^{-}(m))\tilde{Q}_p(m) = -\eta_p \tilde{Q}_p(m), \quad (58)$$

where $p = 0, 1, \ldots, M$. Through the eigenvectors and eigenvalues of this equation, any quantity of interest may be constructed. In figure 3, the effective free energy, given by $F(m) = -k_B T \ln Z(m)$, with $Z(m)$ from equation (56), is shown as the solid lines. Note that the presence of SSBs in general has the effect of decreasing the slope of the free energy for large m, compared to the case in absence of SSBs; thus adding *fast binding* SSBs effectively lowers the melting temperature of the DNA. Depending on the parameters involved, SSBs can thus lead to full denaturation in an unclamped DNA molecule.

4.3. General two-dimensional case

In this subsection, we solve the general eigenvalue problem, equation (12), numerically (see appendix A for details) and compare to the limiting results of the preceding two subsections.

Thus, in figure 7 we plot the relaxation time spectrum $f(\tau)$, that is related to the autocorrelation function (13), which is an accessible experimental observable. Therefore, $f(\tau)$ is a convenient fingerprint of the combined DNA bubble and SSB dynamics. In the caption, we list the various parameters that were chosen to obtain this graph.

In figure 8, we show the relaxation time τ_{relax} as a function of the statistical weight u for (i) no SSBs, but including the loop closure; (ii) no SSBs and no loop closure; (iii) including SSBs and the loop closure; and (iv) including 'fast' SSBs, $\gamma \to \infty$. The relaxation time increases for increasing u (for larger temperature DNA bubbles stay open longer), and we see that by adding SSBs we increase the relaxation time further. The correction due to loop closure is

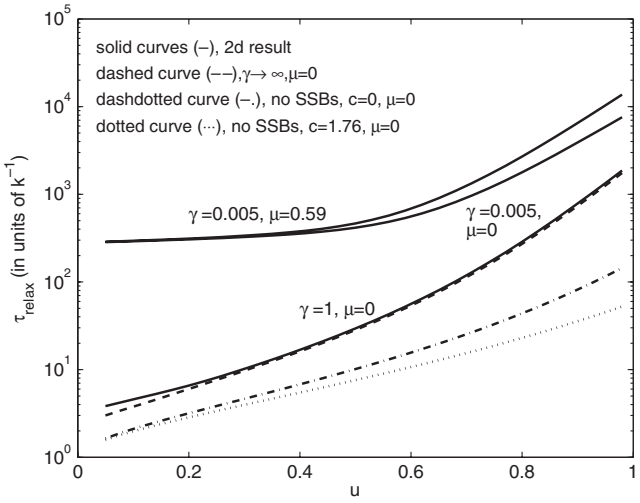

Figure 8. Plot of $\tau_{\rm relax}$ as a function of statistical weight u (which in turn depends on the temperature and the applied external torque), for (i) no SSBs, and including loop closure, dotted curve; (ii) no SSBs, no loop correction, dash–dotted curve; (iii) including SSBs and loop correction for different γ and μ, solid curves; and (iv) including SSBs and loop correction, using the fast binding result $\gamma \to \infty$. The following parameters were used: $\sigma_0 = 10^{-3}$, $M = 20$, $\kappa = 0.5$, and $\lambda = 5$.

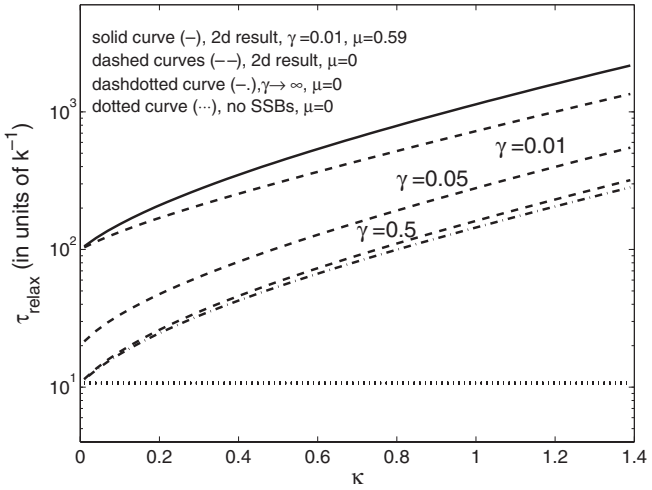

Figure 9. Plot of the relaxation time $\tau_{\rm relax}$ as a function of binding strength κ, for (i) no SSBs and no hook effect $\mu = 0$, dotted line; (ii) fast SSBs, $\gamma \to \infty$ and $\mu = 0$, dash–dotted curve; (iii) general 2D case for different γ and $\mu = 0$, dashed curves; (iv) general 2D case, including the hook effect ($\mu \neq 0$), solid line. Note that the relaxation time increases with increasing binding strength κ. The following parameters were used: $\sigma_0 = 10^{-3}$, $M = 20$, $u = 0.6$, and $\lambda = 5$.

more pronounced close to the melting temperature ($u \to 1$). Also note that the relaxation time $\tau_{\rm relax}$ can be orders of magnitude larger than the typical base-pair closing time (k^{-1}).

Finally, in figure 9, the relaxation time $\tau_{\rm relax}$ as a function of κ for different γ is displayed. The relaxation time increases with increasing binding strength κ, as it should. As γ becomes larger, the adiabatic result for fast SSB binding calculated in the previous subsection is

approached. Inclusion of the hook effect ($\mu \neq 0$) shifts the characteristic relaxation time to larger values.

5. Conclusions

We investigated the coupled dynamics of a single DNA-bubble in the presence of single-stranded DNA binding proteins (SSBs). This study was motivated by a recent experimental work in which a single DNA molecule was overstretched, to obtain information on the binding kinetics of SSBs. The observed kinetic block due to a timescale competition between SSB (un)binding kinetics and bubble lifetime provided an answer to some long-standing questions on the detailed interactions between locally denatured DNA and SSBs. The present investigation is meant to provide a theoretical model basis for the bubble–SSBs system, in particular, to obtain a quantitative interpretation of the process.

As a natural physical approach to the process, we developed a bivariate master equation for the probability distribution of having a bubble of size m with n bound SSBs for different times, for the case when m and n are the slowest variables in the system. The associated two-dimensional eigenvalue problem allowed us to compute the relaxation times for a DNA bubble. With the help of the detailed balance condition, the transfer coefficients in the master equation were explicitly constructed and expressed in terms of the fundamental physical parameters of the problem: the statistical weight u associated with breaking a base-pair interaction, the loop correction exponent c and 'hook exponent' μ, the cooperativity parameter σ_0, the SSB size λ, and binding strength κ. In addition, two rate constants enter the two subprocess governed by the master equation: the inverse zipping time k of a base-pair and the unbinding rate q of a single SSB from the DNA. These two fundamental scales, in combination with the other parameters, determine the competition between bubble and SSB dynamics. For the cases of vanishing and fast binding dynamics ($\gamma = q/k \gg 1$), the problem was reduced to one-dimensional master equations. In the latter case, we performed explicitly the adiabatic elimination of the fast variable n. We solved the general case numerically and compared to the no binding and slow binding limits. In particular, we found that the relaxation time (i) is large close to the melting temperature, and that (ii) the presence of SSBs in general increases the relaxation time, compared to the case when no binding proteins are present. Depending on the parameters, we can tune the competition between SSBs attempting to bind to the bubble and the bubble lifetime, that may eventually provoke full denaturation of the (unclamped) DNA molecule. It was furthermore suggested that the spectral density of relaxation times can be used as a fingerprint of the combined DNA/SSB system, containing experimentally accessible information about the physical parameters of the system.

Acknowledgments

We are happy to thank Mark Williams, Oleg Krichevsky, Andreas Isacsson and Suman Banik for very helpful discussions.

Appendix A. Numerical solution of the two-dimensional master equation on a triangular-like lattice

In this appendix, we present some details of the solution procedure for a two-dimensional eigenvalue problem of the type described by equation (12). One way to solve this problem numerically is to label each (m, n)-pair by a single running variable (here denoted by s), in

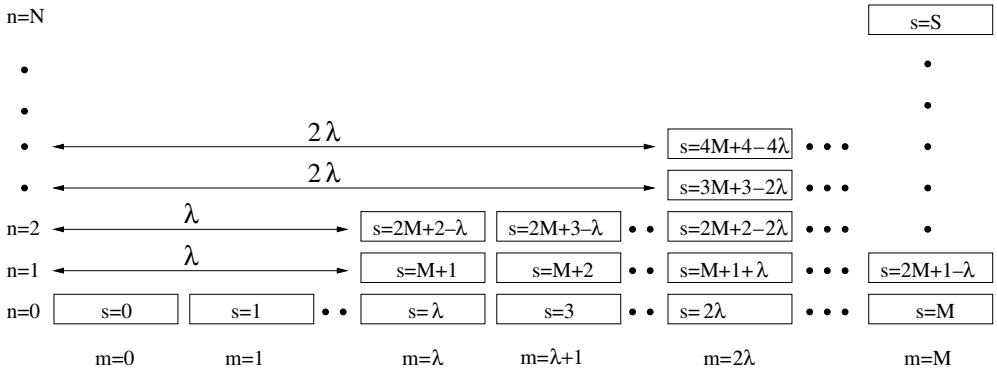

Figure A.1. The lattice, on which the general two-dimensional master equation and its associated eigenvalue problem are defined (see equations (1) and (12)). For numerical solutions the (m, n)-pairs are labelled by a single running variable s, as illustrated above.

order to effectively turn the problem into a 'one-dimensional' eigenvalue equation. The lattice, on which the eigenvalue problem is to be solved, is shown in figure A.1 (see also figure 2). From this figure, we notice that $0 \leqslant n \leqslant N$ (where $N = 2[M/\lambda]$), and that $[(n+1)/2] \leqslant m \leqslant M$. An arbitrary s-point can be obtained from a specific (m, n)-coordinate according to

$$s = n(M+1) - \lambda f(n) + m, \tag{A.1}$$

where

$$f(n) = \sum_{n'=0}^{n} \left[\frac{n'+1}{2} \right]. \tag{A.2}$$

From equation (A.1), we notice that the maximum s-value is

$$S = \max\{s\} = N(M+1) - \lambda f(N) + M. \tag{A.3}$$

Moreover, from equation (A.1) and figure A.1, we notice that a local jump in the m-direction is also a local jump in s-space, i.e., that

$$\begin{aligned} m \to m-1 &\Longleftrightarrow s \to s-1, \qquad \text{for } 1 \leqslant m \leqslant M \\ m \to m+1 &\Longleftrightarrow s \to s+1, \qquad \text{for } 0 \leqslant m \leqslant M-1. \end{aligned} \tag{A.4}$$

However, a jump in the n-direction is equal to a non-local jump in s-space:

$$\begin{aligned} n \to n-1 &\Longleftrightarrow s \to s - \{M+1-\lambda\,(f(n) - f(n-1))\}, \qquad \text{for } 1 \leqslant n \leqslant n^{\max} \\ n \to n+1 &\Longleftrightarrow s \to s + \{M+1-\lambda(f(n+1) - f(n))\}, \qquad \text{for } 0 \leqslant n \leqslant n^{\max} - 1. \end{aligned} \tag{A.5}$$

Using the result above, equation (12) can be written as

$$\begin{aligned} \mathsf{t}^+(s-1)Q_p(s-1) + \mathsf{t}^-(s+1)Q_p(s+1) - {} & (\mathsf{t}^+(s) + \mathsf{t}^-(s))Q_p(s) \\ & + \mathsf{r}^+(s - \{M+1-\lambda(f(n) - f(n-1))\}) \\ & \times Q_p(s - \{M+1-\lambda(f(n) - f(n-1))\}) \\ & + \mathsf{r}^-(s + \{M+1-\lambda(f(n+1) - f(n))\}) \\ & \times Q_p(s + \{M+1-\lambda(f(n+1) - f(n))\})) \\ & - (\mathsf{r}^+(s) + \mathsf{r}^-(s))Q_p(s) = -\eta_p Q_p(s), \end{aligned} \tag{A.6}$$

which, together with the boundary conditions equations (3)–(7), specify the problem: to determine the eigenvalues and eigenvectors of the $(S+1) \times (S+1)$-matrix above. Convenient

checks of the numerical results include: (i) one of the eigenvalues should be zero, and the corresponding eigenvector necessarily has to be the equilibrium distribution; (ii) all other eigenvalues are real and satisfy $\eta_p > 0$ (this guarantees that the equilibrium distribution is indeed reached for long times).

Appendix B. Spectral representation using an orthogonal polynomial approach

Reference [46] is an elaborate mathematical paper describing a polynomial technique of solving birth–death master equations. The paper is aimed at solving infinitely large matrix equations, but contains a number of results pertaining to finite matrices of the type considered here. In this appendix, we summarize some findings derived in [46], that are relevant for our study.

Consider a birth–death eigenvalue equation of the kind of equation (12). We introduce the eigenvector (a row-vector)

$$Q_p = (Q_p(0), \ldots, Q_p(M)) \tag{B.1}$$

and the matrix

$$W^{(M)} = \begin{pmatrix} -\mathrm{t}^+(0) & \mathrm{t}^+(0) & 0 & \cdot & \cdot & \cdot \\ \mathrm{t}^-(1) & -(\mathrm{t}^+(1)+\mathrm{t}^-(1)) & \mathrm{t}^+(1) & \cdot & \cdot & \cdot \\ 0 & \mathrm{t}^-(2) & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 & \mathrm{t}^-(M) & -\mathrm{t}^-(M) \end{pmatrix}. \tag{B.2}$$

Furthermore, introducing the $(M+1)\times(M+1)$ matrix

$$S^{(M)}(\eta) = \eta I^{(M)} + W^{(M)}, \tag{B.3}$$

where $I^{(M)}$ is the $(M+1)\times(M+1)$ unit matrix, we can rewrite equation (12) in the form

$$S^{(M)}(\eta_p) Q_p = 0. \tag{B.4}$$

We then proceed by defining the determinant

$$\Psi_m(\eta) = \det S^{(m)}(\eta), \tag{B.5}$$

which is a polynomial of degree m ($m = 0, 1, \ldots, M$). The $M+1$ eigenvalues $\{\eta_p\}_{p=0}^{M}$ are then determined by the equation

$$\Psi_M(\eta_p) = 0. \tag{B.6}$$

The spectral theory of a matrix of the form above guarantees that one of the eigenvalues is zero [27, 46]. It is therefore convenient to introduce

$$\bar{\Psi}_m(\eta) = \eta \Psi_m(\eta), \tag{B.7}$$

where in [46] it is shown that $\bar{\Psi}_m(\eta)$ satisfies the three-term recurrence relation

$$\bar{\Psi}_m(\eta) + (\eta - \mathrm{t}^+(m-1) - \mathrm{t}^-(m))\bar{\Psi}_{m-1}(\eta) + \mathrm{t}^+(m-1)\mathrm{t}^-(m-1)\bar{\Psi}_{m-2}(\eta) = 0 \tag{B.8}$$

together with

$$\begin{aligned} \bar{\Psi}_{-1}(\eta) &= 0, \\ \bar{\Psi}_0(\eta) &= 1. \end{aligned} \tag{B.9}$$

We note that the $\bar{\Psi}_m(\eta)$ defined by equation (B.8) are *orthogonal polynomials*, see [48], chapter 22. The eigenvectors can then be written [46] as

$$Q_p = (1, \chi_1(\eta_p), \ldots, \chi_M(\eta_p)), \tag{B.10}$$

with

$$\chi_m(\eta) = (\mathrm{t}^-(1)\cdots\mathrm{t}^-(m))^{-1}(\bar{\Psi}_m(\eta) - \mathrm{t}^-(m)\Psi_{m-1}(\eta)). \tag{B.11}$$

The general procedure for obtaining the eigenvalues and eigenvectors is hence: solve equations (B.8) and (B.9) in order to obtain the polynomials $\bar{\Psi}_0(\eta), \ldots, \bar{\Psi}_M(\eta)$. By thereafter solving equation (B.6), the $M+1$ eigenvalues $\eta_0, \ldots, \eta_M$ are obtained. Based on the set $\{\bar{\Psi}_m(\eta_p)\}_{m=0}^{M}$, the eigenvectors are constructed using equations (B.10) and (B.11).

In particular, it is straightforward to show, using the results above, that

$$\eta_0 = 0 \tag{B.12}$$

corresponds to the eigenvector

$$Q_0 = \left(1, \frac{\mathsf{t}^+(0)}{\mathsf{t}^-(1)}, \frac{\mathsf{t}^+(0)\mathsf{t}^+(1)}{\mathsf{t}^-(1)\mathsf{t}^-(2)}, \ldots \frac{\mathsf{t}^+(0)\cdots\mathsf{t}^+(M-1)}{\mathsf{t}^-(1)\cdots\mathsf{t}^-(M)}\right). \tag{B.13}$$

Assuming that the coefficients satisfy detailed balance, $\mathsf{t}^+(m)/\mathsf{t}^-(m+1) = Z(m+1)/Z(m)$, we have

$$Q_0 = \left(1, \frac{Z(1)}{Z(0)}, \frac{Z(2)}{Z(0)}, \ldots, \frac{Z(M)}{Z(0)}\right), \tag{B.14}$$

which is the proper equilibrium solution of the master equation, as it should be [27].

Appendix C. Adiabatic elimination of fast variables

In this appendix, we derive a reduced one-dimensional master equation starting from a two-dimensional master equation, of the form given in equation (1), under the assumption that the rate constant q for protein unbinding is much larger than the rate constant k for DNA zipping. That is, we assume that $\gamma = q/k \gg 1$. This adiabatic elimination procedure is performed on a 'triangular-like' lattice (as illustrated in figures 2 and A.1), which requires a certain care.

Let us consider equation (1) in the limit $\gamma \gg 1$, i.e., for the case that propagation in the 'n-direction' is much faster than the motion in the 'm-direction'. The aim is to eliminate the fast variable n, and derive a reduced master equation in terms of the slow variable m. In particular, we seek an equation for the probability distribution of having a bubble of size m at time t:

$$\tilde{P}(m,t) = \sum_{n=0}^{n^{\max}(m)} P(m,n,t). \tag{C.1}$$

The procedure we follow here is to be considered as the discrete counterpart of the elimination procedure outlined in [49] for the case of a continuous multivariate Fokker–Planck equation; we will point out differences between the discrete and continuous elimination procedure at the appropriate places. We start by introducing the eigenvalue equation associated with the fast variable n according to

$$\begin{aligned}\mathsf{r}^+(m,n-1)q_p(m,n-1) + \mathsf{r}^-(m,n+1)q_p(m,n+1) \\ - (\mathsf{r}^+(m,n) + \mathsf{r}^-(m,n))q_p(m,n) = -\Lambda_p(m)q_p(m,n),\end{aligned} \tag{C.2}$$

where the bubble size m is a fixed parameter. We label the eigenvalues such that $\Lambda_0 = 0 < \Lambda_1 < \cdots < \Lambda_{n^{\max}(m)}$. Equation (C.2) is a matrix equation, whose associated matrix is of size $(n^{\max}(m)+1) \times (n^{\max}(m)+1)$ (since $0 \leqslant n \leqslant n^{\max}(m)$). The eigenvector $q_0(m,n)$ corresponding to Λ_0 is the equilibrium probability distribution. The eigenvectors are orthogonal (in the van Kampen sense [27]), and we choose the normalization according to

$$\sum_{n=0}^{n^{\max}(m)} \frac{q_{p'}(m,n)q_p(m,n)}{q_0(m,n)} = \delta_{p'p}P^{\mathrm{eq}}(m), \tag{C.3}$$

where we introduced the equilibrium probability for a given m,

$$P_{\rm eq}(m) = \frac{Z(m)}{Z}. \tag{C.4}$$

Here,

$$Z(m) = \sum_{n=0}^{n^{\max}(m)} Z(m,n) \tag{C.5}$$

as before, and $Z = \sum_{m=0}^{M} Z(m)$. The reason for choosing the normalization constant as is done in equation (C.3) is that (letting $p = 0$ in this equation) we then obtain

$$\sum_{n=0}^{n^{\max}(m)} q_0(m,n) = P^{\rm eq}(m), \tag{C.6}$$

i.e., $q_0(m,n)$ is the *normalized* equilibrium distribution[14]. We proceed by expanding the probability distribution $P(m,n,t)$ in terms of the eigenvectors $q_p(m,n)$:

$$P(m,n,t) = \sum_{p=0}^{n^{\max}(m)} c_p(m,t) q_p(m,n). \tag{C.8}$$

Inserting this expansion into equation (1), multiplying by $q_{p'}(m,n)/q_0(m,n)$, summing the resulting equation over n, and making use of the orthogonality relation (C.3), we get

$$\left\{\frac{\partial}{\partial t} + \gamma\eta_{p'}(m)\right\} c_{p'}(m,t) P^{\rm eq}(m) = \sum_{p=0}^{n^{\max}(m)} \{L^{+}_{p'p}(m-1)c_p(m-1,t) + L^{-}_{p'p}(m+1)c_p(m+1,t) - L^{0}_{p'p}(m)c_p(m,t)\}, \tag{C.9}$$

where we introduced

$$\begin{aligned} L^{+}_{p'p}(m) &= \sum_{n=0}^{n^{\max}(m+1)} \frac{q_{p'}(m,n)}{q_0(m,n)} \mathsf{t}^{+}(m,n) q_p(m,n), \\ L^{-}_{p'p}(m) &= \sum_{n=0}^{n^{\max}(m-1)} \frac{q_{p'}(m,n)}{q_0(m,n)} \mathsf{t}^{-}(m,n) q_p(m,n), \\ L^{0}_{p'p}(m) &= \sum_{n=0}^{n^{\max}(m)} \frac{q_{p'}(m,n)}{q_0(m,n)} (\mathsf{t}^{+}(m,n) + \mathsf{t}^{-}(m,n)) q_p(m,n). \end{aligned} \tag{C.10}$$

We point out that equation (C.9) is exact. As we are only interested in processes occurring on a timescale larger than $(\gamma\Lambda_1(m))^{-1}$, we can therefore neglect the time-derivative in the equations with $p' \geqslant 1$. Then, following exactly the same steps as in [49], we find that the equation to lowest order in γ^{-1} becomes

$$\frac{\partial c_0(m,t)}{\partial t} P^{\rm eq}(m) = L^{+}_{00}(m-1)c_0(m-1,t) + L^{-}_{00}(m+1)c_0(m+1,t) - L^{0}_{00}(m)c_0(m,t). \tag{C.11}$$

[14] In [49], the following normalization is used instead:

$$\sum_{n=0}^{n^{\max}(m)} \frac{q_{p'}(m,n) q_p(m,n)}{q_0(m,n)} = \delta_{p'p}. \tag{C.7}$$

We notice that if $m < \lambda$ then $n^{\max} = 0$, and the left-hand side of equation (C.7) becomes $\sum_{n=0}^{0} q_0(m,n) = q_0(m, n = 0)$. However, the right-hand side of equation (C.7) equals one. Thus, for a triangular-like lattice (with $\lambda \geqslant 2$), the normalization of the type given in equation (C.3) is the most appropriate.

Making use of the definition (C.1) and equation (C.3), we find that we can write

$$\tilde{P}(m,t) = c_0(m,t)P^{\mathrm{eq}}(m), \tag{C.12}$$

which is the quantity of interest. Notice that we need to have $c_0(m,t) \to 1$ as $t \to \infty$. Using equation (C.12) and defining $\tilde{\mathrm{t}}^{\pm}(m) \equiv L_{00}^{\pm}(m)/P^{\mathrm{eq}}(m)$, as well as $\tilde{\mathrm{t}}^{0}(m) \equiv L_{00}^{0}(m)/P^{\mathrm{eq}}(m)$, we obtain the relation

$$\frac{\partial \tilde{P}(m,t)}{\partial t} = \tilde{\mathrm{t}}^{+}(m-1)\tilde{P}(m-1,t) + \tilde{\mathrm{t}}^{-}(m+1)\tilde{P}(m+1,t) - \tilde{\mathrm{t}}^{0}(m)\tilde{P}(m,t) \tag{C.13}$$

for the probability distribution of having a bubble of size m in the presence of fast binding SSBs at time t. Notice that the results above are general and independent of any particular m-dependence on $n^{\max}(m)$ (i.e., the results above apply to any kind of lattice boundary). Using the boundary conditions from equations (5) and (9), we can change the upper limit in the expression for $\tilde{\mathrm{t}}^{+}(m)$ and $\tilde{\mathrm{t}}^{-}(m)$ such that

$$\begin{aligned}
\tilde{\mathrm{t}}^{+}(m) &= \sum_{n=0}^{n^{\max}(m)} \mathrm{t}^{+}(m,n)\frac{Z(m,n)}{Z(m)}, \\
\tilde{\mathrm{t}}^{-}(m) &= \sum_{n=0}^{n^{\max}(m)} \mathrm{t}^{-}(m,n)\frac{Z(m,n)}{Z(m)}, \\
\tilde{\mathrm{t}}^{0}(m) &= \tilde{\mathrm{t}}^{+}(m) + \tilde{\mathrm{t}}^{-}(m).
\end{aligned} \tag{C.14}$$

We point out that imposing the boundary condition (5) is crucial in order to make sure that the upper summation limits in the expression for $\tilde{\mathrm{t}}^{+}(m)$ and $\tilde{\mathrm{t}}^{-}(m)$ are the same; these boundary conditions therefore also guarantee that the $\tilde{\mathrm{t}}^{0}(m)$ can be written as the sum of the $\tilde{\mathrm{t}}^{+}(m)$ and the $\tilde{\mathrm{t}}^{-}(m)$. In the continuum derivation given in [49], no such explicit incorporation of the boundary conditions were needed. Equation (C.13) together with equation (C.14) is the final equation for the time evolution of $\tilde{P}(m,t)$.

We stop to point out that boundary conditions (5) and (9) are also necessary for the detailed balance condition

$$\tilde{\mathrm{t}}^{+}(m-1)Z(m-1) = \tilde{\mathrm{t}}^{-}(m)Z(m) \tag{C.15}$$

to be satisfied, as is straightforward to show using equation (C.14).

Appendix D. Partition function for a clamped dsDNA

In this final appendix, we outline some details of the calculation of the partition function for a dsDNA, that is clamped at both ends. The melting of a finite-sized DNA usually occurs by melting from the ends, simply because of the less severe restrictions on the random walk at the ends, where no entropy loss due to loop closure occurs (see, for instance, [11]). Thus, the properties of a finite-sized dsDNA, that is clamped at its ends, is therefore different than for unclamped DNA molecules. We here give the details of the calculation of the partition function.

In general, the partition function for a DNA-segment of length M can be written as

$$\begin{aligned}
Z(M) = 1 &+ \sum(\text{statistical weight for 1 bubble}) \\
&+ \sum(\text{statistical weight for 2 bubbles}) + \cdots.
\end{aligned} \tag{D.1}$$

The statistical weight for a bubble consisting of m broken bonds is

$$S(m) = \sigma_0 g(m), \tag{D.2}$$

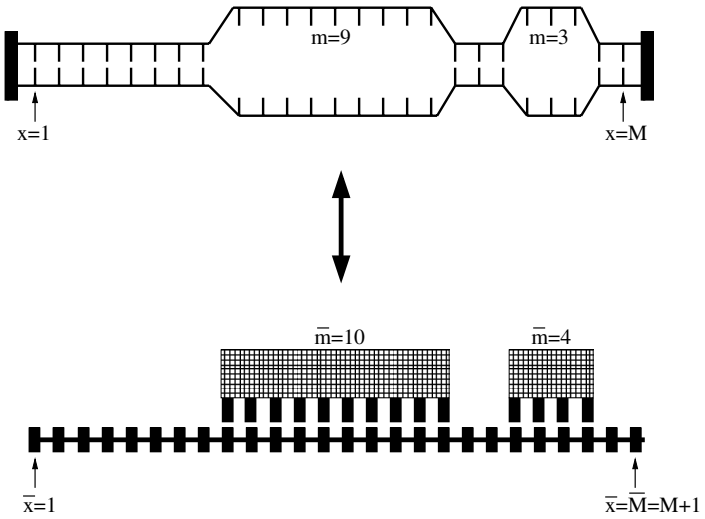

Figure D.1. Correspondence between DNA bubble counting problem and that of binding onto a finite lattice.

with

$$g(m) = u^m (1+m)^{-c}. \tag{D.3}$$

In order to calculate the partition function, equation (D.1), we count in how many ways one or several bubbles of different sizes can be placed along the DNA: consider the case where there are q bubbles along the chain. There are k_i bubbles of type i [$i = 1, \ldots, q$], where a bubble of type i is defined by its size m_i (i.e., consisting of m_i broken base-pair bonds). The k_i have to satisfy the relation[15]

$$q = \sum_{i=1}^{q} k_i. \tag{D.4}$$

We now require that $m_1 < m_2 < \cdots < m_N$, in order to avoid double counting below. Using equations (D.1) and (D.2), we can rewrite the partition function as

$$\begin{aligned} Z(M) = 1 &+ \sigma_0 \sum_{k_1,m_1} g(m_1)^{k_1} Q_M(k_1; m_1) \\ &+ \sigma_0^2 \sum_{k_1,m_1} \sum_{k_2,m_2} g(m_1)^{k_1} g(m_2)^{k_2} Q_M(k_1, k_2; m_1, m_2) + \cdots \\ &+ \sigma_0^q \sum_{k_1,m_1} \cdots \sum_{k_q,m_q} g(m_1)^{k_1} \cdots g(m_q)^{k_q} Q_M(k_1, \ldots, k_q; m_1, \ldots, m_q) + \cdots, \end{aligned} \tag{D.5}$$

where $Q_M(k_1, \ldots, k_q; m_1, \ldots, m_q)$ is the number of configurations for given sets $\{k_i\}$ and $\{m_i\}$. In order to proceed, we note that the present problem of calculating $Q_M(k_1, \ldots, k_q; m_1, \ldots, m_q)$ can be mapped onto the problem of binding to a finite lattice; the problem of placing bubbles of size m_i (i.e., bubbles consisting of m_i broken base-pairs) on a region consisting of M internal base-pairs is in fact identical to the one of placing binding

[15] For instance, in the third term of equation (D.5) proportional to σ_0^2, we can have the three combinations ($k_1 = 1$, $k_2 = 1$), ($k_1 = 2$, $k_2 = 0$) and ($k_1 = 0$, $k_2 = 2$).

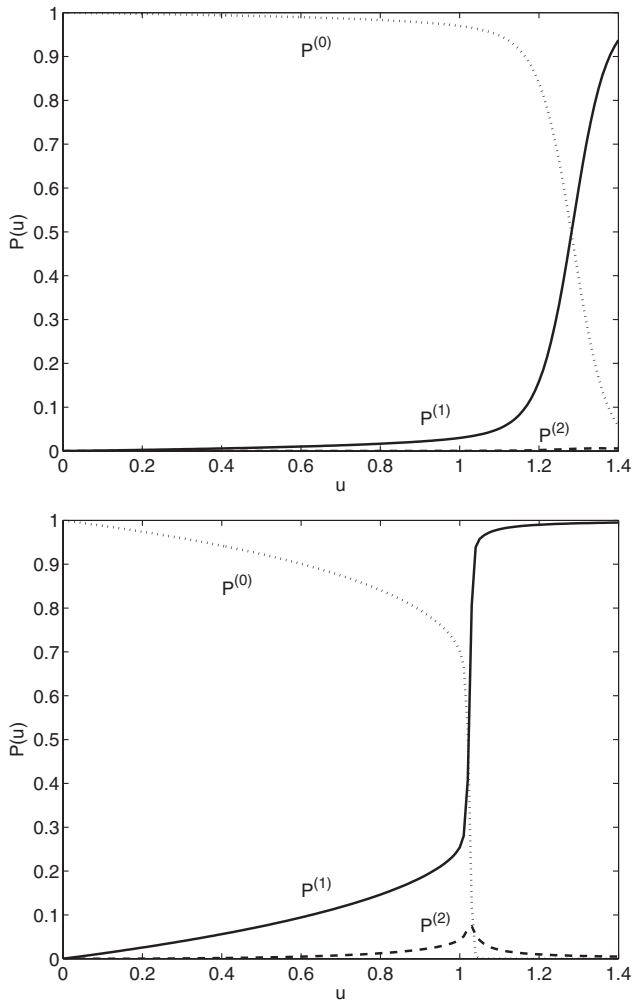

Figure D.2. Equilibrium probability of having 0, 1, or 2 bubbles as a function of u for two different DNA region sizes. Top: the chain length is $M = 40$. Bottom: the chain length was taken to be $M = 400$. In both graphs, the cooperativity parameter was $\sigma_0 = 10^{-3}$ and the loop correction exponent $c = 1.76$.

particles of size $\bar{m}_i = m_i + 1$ onto a lattice consisting of $\bar{M} = M + 1$ lattice sites, as is illustrated in figure D.1. The latter problem was worked out in [43]. One should remember the constraint that the number of occupied lattice units cannot exceed the lattice size, thus

$$\sum_{i=1}^{q} k_i(m_i + 1) \leqslant M + 1. \tag{D.6}$$

Let us recapitulate the results from [43]. We introduce q binding particles (i.e., q bubbles) with k_i particles of type i (where a type-i particle has size $\bar{m}_i = m_i + 1$) onto a lattice of size $\bar{M} = M + 1$. The number of unoccupied binding sites is then

$$I_{\mathrm{u}} = M + 1 - \sum_{i=1}^{q} k_i(m_i + 1), \tag{D.7}$$

and the number of binding particles is $q = \sum_{i=1}^{q} k_i$. The number of 'individuals' to be permuted becomes $I_u + q$, and hence for given sets $\{k_i\}$ and $\{m_i\}$, the number of configurations is

$$Q_M(k_1, \ldots, k_q; m_1, \ldots, m_q) = \frac{(I_u + q)!}{I_u! \prod_{i=1}^{q} k_i!} = \frac{[M + 1 - \sum_{i=1}^{q} m_i k_i]!}{[M + 1 - \sum_{i=1}^{q} (m_i + 1)k_i]! \prod_{i=1}^{q} k_i!}. \quad \text{(D.8)}$$

The partition function is thus given by equations (D.2), (D.5) and (D.8) in general, where the summation limits are determined by the constraints contained in equations (D.4) and (D.6).

Using the results above, we straightforwardly find that the partition function to second order in the number of loops becomes

$$Z(M) = 1 + \sigma_0 Z^{(1)}(M) + \sigma_0^2 Z^{(2)}(M) + \cdots, \quad \text{(D.9)}$$

with

$$\begin{aligned} Z^{(1)}(M) &= \sum_{m_1=1}^{M} (M - m_1 + 1) g(m_1), \\ Z^{(2)}(M) &= \tfrac{1}{2} \sum_{m_1=1}^{M-1} g(m_1) I(m_1), \\ I(m_1) &= \sum_{m_2=1}^{M-m_1-1} (M - m_1 - m_2 + 1)(M - m_1 - m_2) g(m_2). \end{aligned} \quad \text{(D.10)}$$

The probability of having zero bubbles is $P^{(0)} = 1/Z(M)$. The probability of having one bubble corresponds to $P^{(1)} = \sigma_0 Z^{(1)}(M)/Z(M)$, and the probability for the simultaneous occurrence of two bubbles becomes, similarly, $P^{(2)} = \sigma_0^2 Z^{(2)}(M)/Z(M)$. These quantities are plotted as a function of u in figure D.2 for chain lengths $M = 40$ and 400. Notice that the probability of having two bubbles is relatively small due to $\sigma_0 \ll 1$, but increases with the chain length (i.e., as we go from $M = 40$ to 400). It naturally rises closer to the melting temperature.

References

[1] Kornberg A 1974 *DNA Synthesis* (San Francisco, CA: Freeman)
[2] Delcourt S G and Blake R D 1991 *J. Biol. Chem.* **266** 15160
[3] Revzin A (ed) 1990 *The Biology of Non-Specific DNA–Protein Interactions* (Boca Raton, FL: CRC Press)
[4] Watson J D and Crick F H C 1953 *Cold Spring Harbor Symposia on Quantitative Biology* vol 18, p 123
[5] Wartell R M and Benight A S 1985 *Phys. Rep.* **126** 67
[6] 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
[7] Blossey R and Carlon E 2003 *Phys. Rev.* E **68** 061911
[8] Yeramian E 2000 *Gene* **255** 139
[9] Yeramian E 2000 *Gene* **255** 151
[10] Carlon E, Malki M L and Blossey R 2004 *Preprint* q-bio/0409034
[11] Poland D and Scheraga H 1970 *Theory of Helix–Coil Transitions in Biopolymers* (New York: Academic)
[12] Guéron M, Kochoyan M and Leroy J-L 1987 *Nature* **328** 89
[13] Kittel C 1969 *Am. J. Phys.* **37** 917
[14] Altan-Bonnet G, Libchaber A and Krichevsky O 2003 *Phys. Rev. Lett.* **90** 138101
[15] Zeng Y, Montrichok A and Zocchi G 2004 *J. Mol. Biol.* **339** 67
[16] Kornberg A and Baker T A 1992 *DNA Replication* (New York: Freeman)
[17] Karpel R L 1990 *The Biology of Non-Specific DNA–Protein Interactions* ed A Revzin (Boca Raton, FL: CRC Press)
[18] Jensen D E and von Hippel P H 1976 *J. Biol. Chem.* **251** 7198
[19] Jensen D E, Kelly R C and von Hippel P H 1976 *J. Biol. Chem.* **251** 7215

[20] Karpel R L 2002 *IUBMB Life* **53** 161
[21] Pant K, Karpel R L and Williams M C 2003 *J. Mol. Biol.* **327** 571
[22] Pant K, Karpel R L, Rouzina I and Williams M C 2004 *J. Mol. Biol.* **336** 851
[23] Hwa T, Marinari E, Sneppen K and Tang L-H 2003 *Proc. Natl Acad. Sci. USA* **100** 4411
[24] Hanke A and Metzler R 2003 *J. Phys. A: Math. Gen.* **36** L473
[25] Altan-Bonnet G, Libchaber A and Krichevsky O 2003 *Phys. Rev. Lett.* **90** 138101
[26] Ambjörnsson T and Metzler R 2005 Breathing dynamics in double-stranded DNA with sequence heterogeneity, to be submitted
[27] van Kampen N G 1992 *Stochastic Processes in Physics and Chemistry* 2nd edn (Amsterdam: North-Holland)
[28] Richard C and Guttmann A J 2004 *J. Stat. Phys.* **115** 925
[29] Zimm B H 1960 *J. Chem. Phys.* **33** 1349
[30] de Gennes P-G 1979 *Scaling Concepts in Polymer Physics* (Ithaca, NY: Cornell University Press)
[31] Fixman M and Freire J J 1977 *Biopolymers* **16** 2693
[32] Fisher M E 1966 *J. Chem. Phys.* **45** 1469
[33] Kafri Y, Mukamel D and Peliti L 2000 *Phys. Rev. Lett.* **85** 4988
[34] Garel T, Monthus C and Orland H 2001 *Europhys. Lett.* **55** 132
[35] Hanke A and Metzler R 2003 *Phys. Rev. Lett.* **90** 159801
[36] Kafri Y, Mukamel D and Peliti L 2003 *Phys. Rev. Lett.* **90** 159802
[37] Azbel M Ya 2004 *Preprint* physics/0403106
[38] Kafri Y, Mukamel D and Peliti L 2004 *Preprint* cond-mat/0406562
[39] Bhattacharjee S M 2002 *Europhys. Lett.* **57** 772
[40] Garel T, Monthus C and Orland H 2002 *Europhys. Lett.* **57** 774
[41] Hwa T, Marinari E, Sneppen K and Tang L H 2003 *Proc. Natl Acad. Sci.* **100** 4411
[42] McQuistan R B 1968 *Nuovo Cimento* **58** 86
[43] Epstein I R 1978 *Biophys. Chem.* **8** 327
[44] Ambjörnsson T and Metzler R 2004 *Phys. Biol.* **1** 77
[45] Di Marzio E A, Guttman C M and Hoffman J D 1979 *Faraday Discuss.* **68** 210
[46] Ledermann W and Reuter G E H 1954 *Phil. Trans. R. Soc.* A **246** 321
[47] Gradshteyn I S and Ryzhik I M 2000 *Table of Integrals, Series and Products* (San Diego, CA: Academic)
[48] Abramowitz M and Stegun I A 1972 *Handbook of Mathematical Functions* (New York: Dover Publications)
[49] Risken H 1989 *The Fokker–Planck Equation* (Berlin: Springer)