

Dynamic Approach to DNA Breathing

RALF METZLER* and TOBIAS AMBJÖRNSSON

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

*(*Author for correspondence, e-mail: metz@nordita.dk)*

Abstract. Even under physiological conditions, the DNA double-helix spontaneously denatures locally, opening up fluctuating, flexible, single-stranded zones called DNA-bubbles. We present a dynamical description of this DNA-bubble breathing in terms of a Fokker-Planck equation for the bubble size, based on the Poland-Scheraga free energy for DNA denaturation. From this description, we can obtain basic quantities such as the lifetime, an important measure for the description of the interaction of a breathing DNA molecule and selectively single-stranded DNA binding proteins. Our approach is consistent with recent single molecule measurements of bubble fluctuation. We also introduce a master equation approach to model DNA breathing, and discuss its differences from the continuous Fokker-Planck description.

Key words: DNA denaturation, DNA-bubbles, single molecule dynamics, Poland-Scheraga free energy

Abbreviations: dsDNA – double-stranded DNA, ssDNA – single-stranded DNA, SSB – single-stranded DNA binding protein

PACS: 87.15.-v, 82.37.-j, 87.14.G

1. Introduction

The equilibrium structure of a DNA molecule under physiological conditions is the Watson-Crick double helix. Its stability is effected by Watson-Crick H-bonding and base-stacking [1]. While Watson-Crick base-pairing contributes only little to the stability, it guarantees the high specificity under physiological processes such as transcription and replication due to the key-lock principle according to which the base A exclusively binds to T, and G to C. Base-stacking, the major contribution, introduces hydrophobic interactions between adjacent, planar pairs of bases [1, 2]. By variation of temperature or pH-value in solution, double-stranded DNA (dsDNA) progressively denatures, producing regions of single-stranded DNA (ssDNA), until the double-strand is fully molten. The melting temperature T_m defines the temperature at which half of the DNA molecule has denatured [1, 3]. In Figure 1, we show a sketch of the progressive melting of dsDNA.

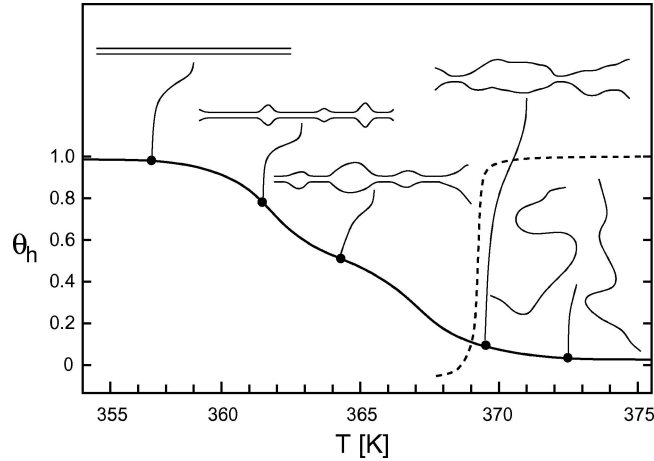


Figure 1. Fraction θ_h of double-helical domains within the DNA as a function of temperature. Schematic representation of $\theta_h(T)$, showing the increased formation of bubbles and unzipping from the ends at increased temperature, until full denaturation has been reached.

Due to ambient thermal fluctuations, even under physiological conditions, ds-DNA can spontaneously unzip locally, producing comparatively flexible, single-stranded DNA-bubbles that are of a few tens of base-pairs in size [4]. The size of a DNA-bubble changes constantly by successive unzipping or zipping of base-pairs at the zipping forks where the single-stranded bubble connects to the intact double-helix. As we assume that the (un)zipping dynamics (single zipping events running off on the scale of a few tens of μsec [5]) is slower than the Rouse relaxation time $\sim n^2$ of the rather short ssDNA strand making up the bubble [6] (i.e., we neglect dynamic feedback from the rest of the DNA molecule), this *DNA breathing* can be interpreted as a random walk in the one-dimensional coordinate n , the number of denatured base-pairs. In what follows, we first propose a continuum description to bubble dynamics in terms of a Fokker-Planck equation, that describes a stochastic process for the bubble size n in a potential given by the Poland-Scheraga free energy. After introducing some essential features of this model, we compare the continuum approach to the probably more natural choice of a discrete approach in terms of a master equation. The latter offers the advantage of directly including the loop initiation barrier.

2. Poland-Scheraga Free Energy, One-Bubble Case

The Poland-Scheraga model for DNA melting, in an Ising-type approach based on earlier ideas by Kittel [7] and Zimm [8], captures the competition between energetic stacking interactions of the intact double-stranded portion of the DNA molecule with the entropy gain from the more flexible ssDNA loops [3, 9–13].¹

The statistical weight for the dissociation of base-pairs includes an enthalpic contribution H_{ij} from unstacking base-pair i from base-pair j and an entropic contribution S_{ij} for positioning of unbound bases. Both give rise to the Gibbs free energy $G_{ij} = H_{ij} - TS_{ij}$, that is tabled for all possible combinations of neighbouring pairs of base-pairs [14]. Melting profiles are usually measured through UV-absorption, revealing quite precise data that can even be used to distinguish coding from non-coding regions in a genome [15, 16].

For a homopolymer, we define $\gamma \equiv \beta G_{ii}/2$ measuring the Gibbs free energy per broken base-pair in units of $k_B T$ ($\equiv \beta^{-1}$). The melting temperature T_m is then defined through $\gamma(T_m) = 0$.² To interrupt the double helix to initiate a bubble, that is bordered by helical domains (compare Figure 2), co-operative interactions have to be overcome, to release base-stacked and paired nucleotides. This is expressed by an additional weight $\sigma_0 \equiv \exp(-\gamma_0)$, and varies between a few 10^{-2} to 10^{-5} , for different temperatures [11, 12, 17, 18]. Finally, the formation of a flexible bubble includes an entropy loss corresponding to the returning probability of a random walk [3, 11–12]. For DNA-bubbles, one usually assumes the form $F(n) = (n + D)^{-c}$ where the offset D accounts for persistence length effects [19], and c is the loop closure exponent [3, 11, 10]. For D , a standard choice is $D = 1$ [14], while the value of c depends on the boundary conditions; for a dilute solution, $c \approx 1.76$ is assumed [14], while other values have been suggested [20, 21]. For the dynamics of bubbles the exact value of c is indeed of lesser importance [17, 22]. Collecting all contributions, we obtain for the statistical weight $\mathcal{L}(n)$ of a single bubble of size n in a homopolymer the expression

$$\mathcal{L}(n) = \sigma_0 e^{-n\gamma} (n + 1)^{-c}. \quad (1)$$

Consequently, the total free energy of the bubble is given by

$$\beta \mathcal{F}(n) = -\log \mathcal{L}(n) = n\gamma(T) + \gamma_0 + c \log(n + 1), \quad (2)$$

in units of $k_B T$. The free energy landscape is portrayed in Figure 3 for various temperatures for the case of an A-T-bubble.

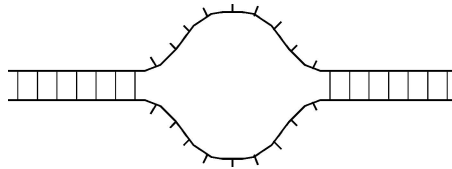


Figure 2. Denaturation bubble of size $n = 9$ bordered by helical segments.

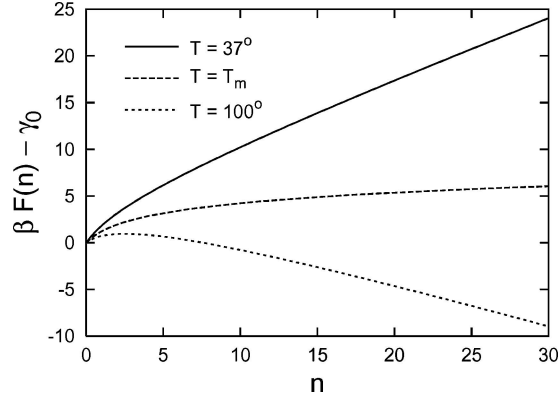


Figure 3. The variable part of the bubble free energy (2) as a function of the bubble size n , $T = 37^\circ\text{C}$ ($\gamma \approx 0.6$), $T_m = 66^\circ\text{C}$ ($\gamma = 0$), and $T = 100^\circ\text{C}$ ($\gamma \approx -0.5$). In the latter case, a small barrier precedes the negative drift towards bubble opening as dominated by the $\gamma < 0$ contribution.

3. Continuous Fokker-Planck Equation Approach

Following the continuum approach suggested in reference [22], we combine the continuity equation

$$\frac{\partial P(n, t)}{\partial t} + \frac{\partial j(n, t)}{\partial n} = 0 \quad (3)$$

for the probability density function $P(n, t)$ to find the bubble with size n at time t and the probability current j , with the constitutive expression for the associated current (compare the discussion in reference [23]),

$$j(n, t) = -K \left(\frac{\partial P(n, t)}{\partial n} + \frac{P(n, t)}{k_B T} \frac{\partial \mathcal{F}}{\partial n} \right). \quad (4)$$

Here, it is assumed that the potential exerting the drift is given by the bubble free energy (2), and we imposed an Einstein relation of the form $K = k_B T \mu$, where the mobility μ has dimensions $[\mu] = \text{sec}/(\text{g} \cdot \text{cm}^2)$, and therefore $[K] = \text{sec}^{-1}$ represents an inverse time scale.³ Combining Eqs. (2)–(4), we obtain the one-dimensional Fokker-Planck equation

$$\frac{\partial P(n, t)}{\partial t} = K \left(\frac{\partial}{\partial n} \left\{ \gamma + \frac{c}{n+1} \right\} + \frac{\partial^2}{\partial n^2} \right) P(n, t). \quad (5)$$

The contribution from the loop closure, $c/(n+1)$, decreases quickly with n , and can thus be viewed a correction term to the constant drift exerted by the free energy term γ per unzipped base-pair. In fact, without the loop closure term,

Eq. (5) exactly equals the expression used to fit the fluorescence correlation data in reference [5].

The Fokker-Planck Eq. (5) describes the relative diffusion of the two zipper forks where the ssDNA bubble connects to the dsDNA helical segments. At temperatures $T < T_m$ below the melting temperature, the drift from the free energy biases this diffusion towards $n = 0$, that can be included as an absorbing boundary condition. The lifetime of a bubble can then be obtained from the corresponding first passage time problem. In Figure 4, we plot the typical lifetime of a bubble as function of the initial bubble size n_0 , both for physiological temperature and close to T_m . In the physiological regime, the bubble lifetime grows approximately linearly with n_0 . This can be understood from the first passage time density

$$f(0, t) = \frac{n_0}{\sqrt{4\pi K t^3}} \exp \left\{ -\frac{(n_0 - K\gamma t)^2}{4Dt} \right\}, \quad (6)$$

that is valid for the homopolymer bubble when the loop closure factor is neglected in comparison to the constant drift γ . The corresponding mean first passage time $\mathcal{T} = \int_0^\infty f(0, t) t dt = n_0/[K\gamma]$ indeed scales linearly in n_0 . Conversely, close to T_m and neglecting the loop closure, the drift-free first passage calculation leads to $\mathcal{T} \simeq n_0^2$, corresponding to the roughly parabolic form for the lifetime in Figure 4.

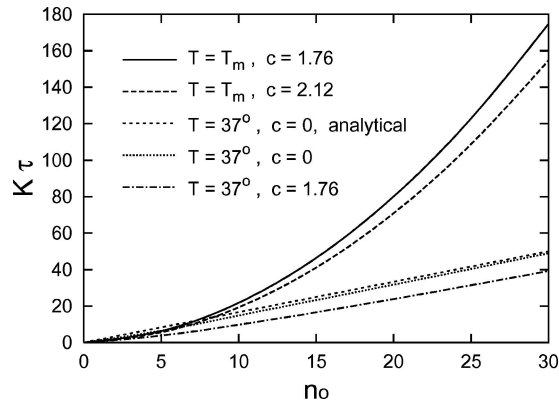


Figure 4. Characteristic bubble closing times τ as a function of initial bubble size n_0 for an AT-homopolymer, obtained from the Fokker-Planck Eq. (5) by numerical integration. At $T = 37^\circ\text{C}$, the result for $c = 1.76$ is compared to the approximation $c = 0$ which leads to somewhat larger closing times. The analytical solution for $c = 0$ compares well with the numerical result, the slight discrepancy being due to the reflecting boundary condition applied in the numerics, in comparison to the natural boundary condition at $n \rightarrow \infty$ used to derive Eq. (6). At the melting temperature $T_m = 66^\circ\text{C}$, the closing times for the values $c = 1.76$ and $c = 2.12$ can be distinguished.

4. Discrete Master Equation Approach: Including the Bubble Initiation

Let us now compare the continuum approach from the previous section to a discrete master equation approach. A priori, there are two major advantages to a discrete formulation: (i) it is closer to the actual physical situation, as the co-ordinate n of (un)zipping base-pairs is in fact discrete; (ii) the master equation allows us to explicitly include the loop initiation factor σ_0 , in contrast to the continuous Fokker-Planck Eq. (5), in which the gradient of the free energy appears, and the constant σ_0 -factor drops out. Thus, in Figure 5, we display the free energy including the rather sharp jump between $n = 0$ and 1.

The probability of finding a bubble of size n at some time t , $P(n, t)$, is governed by the master equation

$$\frac{\partial}{\partial t} P(n, t) = \mathfrak{t}^+(n-1)P(n-1, t) + \mathfrak{t}^-(n+1)P(n+1, t) - (\mathfrak{t}^+(n) + \mathfrak{t}^-(n))P(n, t), \quad (7)$$

whose forward and backward transfer coefficients $\mathfrak{t}^\pm(n)$ (opening and closing rates) are defined in terms of the partition function (1) through

$$\mathfrak{t}^+(n) = k \frac{\mathcal{L}(n+1)}{\mathcal{L}(n)} = ke^{-\gamma} \left(\frac{1+n}{2+n} \right)^c, \quad n \geq 1; \quad (8)$$

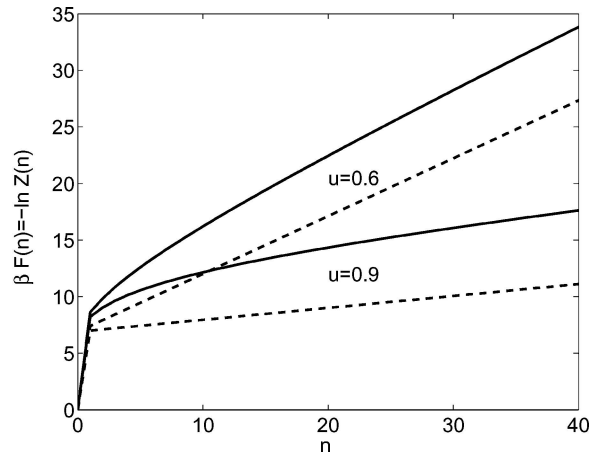


Figure 5. Full free energy $\mathcal{L}(n) = -\ln \mathcal{L}(n)$ including the loop initiation energy σ_0 as function of bubble size n , for two different temperatures, i.e., $u = \exp(-\gamma)$. The full lines correspond to the full expression with loop closure factor, whereas the dashed lines neglect the latter. We used the loop initiation factor $\sigma_0 = 10^{-3}$.

for the forward rate (including the rate $t^+(0) = 2^{-c}k\sigma_0e^{-\gamma}$ for bubble opening), while the backward transfer coefficients are chosen in the form

$$t^-(n) = k. \quad (9)$$

The choice of relations (8) and (9) ensures that the dynamics fulfils the detailed balance condition and the system eventually reaches equilibrium. To include the bubble boundaries appropriately, we impose the reflecting boundary conditions

$$t^+(m = M) = 0; \quad t^-(m = 0) = 0, \quad (10)$$

guaranteeing that the bubble size does not exceed the value M nor takes on negative values.

As derived in reference [24] in detail, the master Eq. (7) can be solved conveniently by the eigenmode expansion

$$P(n, t) = \sum_{p=0}^M c_p Q_p(n) \exp(-t/\tau_p). \quad (11)$$

Neglecting the loop closure factor, an exact solution can be found, whereas in the general case, a numerical solution, for instance with MatLab, is straightforward. The zeroth term of the eigenmode expansion (11) represents the equilibrium distribution

$$P_{\text{eq}}(n) = Q_0(n), \quad (12)$$

that is time-independent. All higher order contributions ($p = 1, 2, \dots$) describe the exponential relaxation of modes whose relative contribution c_p is determined by the initial condition.

An important experimental quantity is the autocorrelation function $A(t) = \langle m(t)m(0) \rangle$ [5]. It can be written in the form [25]

$$A(t) = \int_0^\infty \exp\left(-\frac{t}{\tau}\right) f(\tau) d\tau, \quad (13)$$

where we use the spectral density of relaxation times in discrete form, $f(\tau) = \sum_{p=1}^M A_p \delta(\tau - \tau_p)$, corresponding to the discrete set of time-eigenvalues τ_p^{-1} ; and the result

$$A_p = \left(\sum_n n Q_p(n) \right)^2 \quad (14)$$

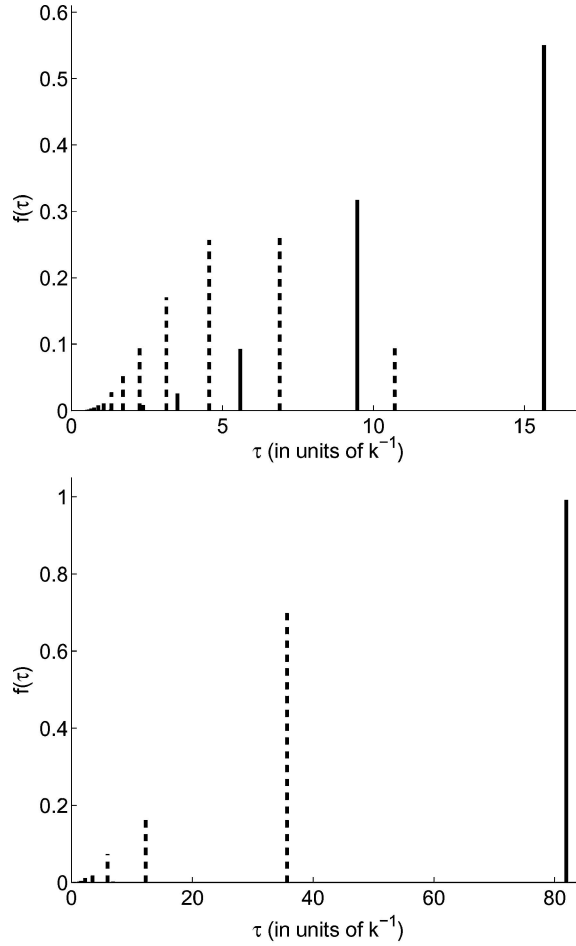


Figure 6. Spectral density $f(\tau)$ of relaxation times τ for two different $u = \exp(-\gamma)$. Top: $u = 0.6$ well below T_m exhibiting a pronounced multi-exponential behaviour; bottom: $u = 0.9$ close to T_m with clear dominance of one (the largest) relaxation time mirroring a single exponential behaviour typical for a two-state system. Notice the different scales on the abscissae. The solid lines correspond to the case including the loop closure factor, while the dashed lines neglect the entropy penalty from loop closure. We chose $M = 20$ and $\sigma_0 = 10^{-3}$.

for the amplitudes A_p . In Figure 6, we show the spectral density $f(\tau)$ for two different temperatures, demonstrating the pronounced multi-exponential behaviour well below T_m , whereas close to T_m , the spectrum is approximately that of a two-state system. Note the change brought about by neglecting the loop closure factor.

A related quantity of interest is the longest relaxation time

$$\tau_{\text{relax}} = \tau_1, \quad (15)$$

that determines the characteristic time the system takes to equilibrate. This τ_{relax} can be used to characterise the behaviour of the bubble as function of the system parameters. For instance, we display τ_{relax} in Figure 7 as function of γ , and for two different sizes of the bubble zone.

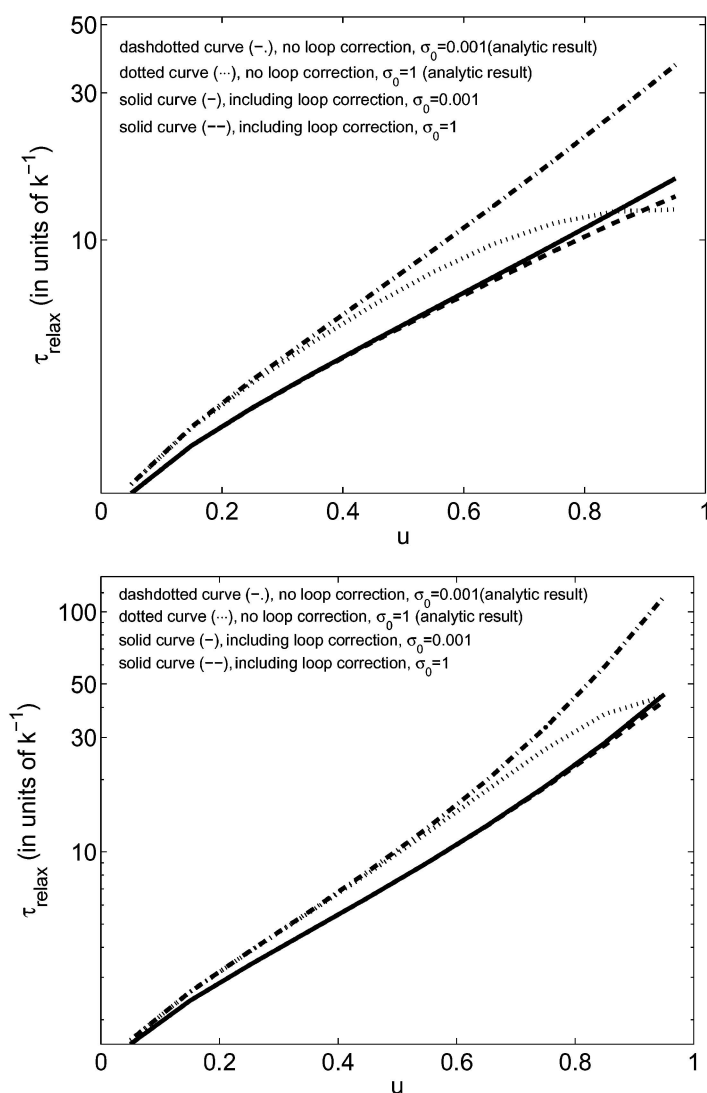


Figure 7. Longest relaxation time τ_{relax} as function of statistical weight $u = \exp(-\gamma)$. Top: shorter bubble zone with $M = 10$; bottom: longer bubble zone, $M = 20$. Notice that the dependencies on σ_0 and the loop correction is weaker for the longer bubble zone. On approaching the melting temperature T_m , the relaxation time diverges.

5. Discussion

We presented two dynamical approaches to DNA breathing, the incessant unzipping and zipping of single-stranded DNA zones within a double-stranded DNA. Both models reduce the full three-dimensional dynamics to an effective, one-dimensional reaction coordinate, that being the bubble size n . This is legitimate, as, due to the relatively small characteristic bubble size combined with the relatively long time scales for (un)zipping of a base-pair, n becomes the slow variable of the dynamics, while the Rouse modes of the bubble (and, for the short DNA used in typical experiments, even the entire chain) equilibrate much faster. We confined our discussion to a single bubble, keeping in mind typical experiments on designed DNA. However, also for unconstrained DNA, this assumption is expected to hold well for temperatures below the melting temperature, as statistically the bubbles can be assumed to be independent due to σ_0 .

The first approach corresponds to a continuum Fokker-Planck equation governing the probability density function $P(n, t)$, corresponding to the probability $P(n, t)dn$ ($n \in \mathbb{R}_0^+$) for finding a bubble size in the interval $n, \dots, n + dn$ at time t . Mathematically, this Fokker-Planck equation can be treated easily, and the first passage time problem neglecting the loop closure correspond to the standard cases of constant drift and zero drift diffusion. The second approach reflects more the physical nature of the problem, i.e., the discrete steps of sequential base-pair (un)zipping. To some extent, this is at the expense of the mathematical handling. To allow for an easy numerical treatment of the master equation controlling the time evolution of the discrete probability $P(n, t)$ ($n \in \mathbb{N}_0$), we introduced an eigenmode expansion that can be implemented in programs like MatLab. Apart from being the more physical way to describe the bubble fluctuations, the master equation approach allows us to explicitly include the loop initiation factor σ_0 .

In the discrete model a base-pair is assumed to be in one of two states (open or closed); this assumption is justified for a rather short-ranged interaction between bases. In the continuum model the base-pair interaction is assumed to be spatially smooth and of somewhat longer range, which justifies the use of a continuum free energy landscape describing the breathing process.

We demonstrated the influence of typical system parameters such as temperature (that can be effectively changed by applying an external torque [17]), co-operativity factor of loop initiation, initial bubble size, and loop closure factor. To visualise these effects, we made use of the relaxation time spectrum and the longest relaxation time. We believe that this model is a physical but still transparent approach, elucidating some of the essential features of DNA breathing. While we chose to concentrate on the experimentally relevant, simplest case of a homopolymer one-bubble situation, the more general case can be included in a straightforward manner. We are currently expanding the model to include the interaction of breathing bubbles with proteins that selectively bind to single-strand DNA-binding proteins [24], that turn out to

compete in concern of their binding dynamics with the lifetime of DNA-bubbles [26].

Acknowledgment

We are happy to acknowledge very helpful discussions with Andreas Hanke, Yossi Klafter, Oleg Krichevsky, and Mark Williams.

Notes

1. The persistence length of dsDNA is roughly 50 nm (≈ 150 bp), in comparison to 2.5 nm for ssDNA.
2. For an A-T homopolymer at 37 °C and standard salt, $\gamma \approx 0.6$, while for a G-C homopolymer $\gamma \approx 1.46$. The corresponding melting temperatures are $T_m(AT) \approx 67$ °C and $T_m(GC) \approx 103$ °C.
3. Note that μ is the mobility of the zipper forks, and therefore not related to the bulk mobility of the bubble-ssDNA. In our model, this parameter sets the time scale of the dynamics, and has to be calibrated from experiments or quantum chemical calculations. Note also that we assume for simplicity that K is independent of n . This latter assumption will have to be changed if heteropolymer effects are considered. For a detailed treatment, see reference [27].

References

1. Kornberg, A.: *DNA Synthesis*, W.H. Freeman, San Francisco, CA, (1974).
2. Delcourt, S.G. and Blake R.D.: Stacking Energies in DNA, *J. Biol. Chem.* **266** (1991), 15160–15169.
3. Poland, D. and Scheraga, H.A.: *Theory of Helix-Coil Transitions in Biopolymers*, Academic Press, New York, NY, (1970).
4. Guéron, M., Kochoyan, M. and Leroy J.-L.: *Nature* **328** (1987), 89.
5. Altan-Bonnet, G., Libchaber, A. and Krichevsky, O.: *Phys. Rev. Lett.* **90** (2003), 138101.
6. de Gennes, P.G., *Scaling concepts in polymer physics*, Cornell University Press, Ithaca, New York, (1979).
7. Kittel, C.: *Am. J. Phys.* **37** (1969), 917.
8. Zimm, B.H.: *J. Chem. Phys.* **33** (1960), 1349.
9. Poland, D. and Scheraga, H.A.: *J. Chem. Phys.* **45** (1966), 1456; *ibid.* 1464.
10. Fisher, M.E.: *J. Chem. Phys.* **45** (1966), 1469.
11. Wartell, R.M. and Benight, A.S.: *Phys. Rep.* **126** (1983), 67.
12. Gotoh, O.: *Adv. Biophys.* **16** (1983), 1.
13. Richard, C. and Guttman, A.J.: *J. Stat. Phys.* **115** (2004), 925.
14. Blake, R.D., Bizzaro, J.W., Blake, J.D., Day, G.R., Delcourt, S.G., Knowles, J., Marx, K.A. and SantaLucia, Jr. J.: *Bioinformatics* **15** (1999), 370.
15. Yeramian, E.: *Gene* **255** (2000), 139; *ibid.* 151.
16. Carlon, E., Malki, M.L. and Blossey, R.: *Phys. Rev. Lett.* **94** (2005), 178101, *E-print* (2004), q-bio/0409034.
17. Hwa, T., Marinari, E., Sneppen K. and Tang, L.-H.: *Proc. Natl. Acad. Sci. USA* **100** (2003), 4411.
18. Blossey, R. and Carlon, E.: *Phys. Rev. E* **68** (2003), 061911.
19. Fixman, M. and Freire, J.J.: *Biopol.* **16** (1977), 2693.

20. Kafri, Y., Mukamel, D. and Peliti, L.: *Phys. Rev. Lett.* **85** (2000), 4988; *ibid.* **90** (2003), 159802.
21. Hanke, A. and Metzler, R.: *Phys. Rev. Lett.* **90** (2003), 159801.
22. Hanke, A. and Metzler, R.: Bubble dynamics in DNA, *J. Phys. A* **36** (2003), L473–L480.
23. Chuang, J., Kantor, Y. and Kardar, M.: *Phys. Rev. E* **65** (2002), 011802.
24. Ambjörnsson, T. and Metzler, R.: *Phys. Rev. E* **72** (2005), 030901 (R), *E-print* q-bio.BM/0411053.
25. van Kampen, N.G.: *Stochastic Processes in Physics and Chemistry*, North-Holland, Amsterdam, (1981).
26. Pant, K., Karpel, R.L. and Williams, M.C.: *J. Mol. Biol.* **327** (2003), 571.
27. Ambjörnsson, T. and Metzler, R.: *Dynamics of bubble breathing in heterogeneous DNA*, unpublished.