Sequence Sensitivity of Breathing Dynamics in Heteropolymer DNA

Tobias Ambjörnsson,[1] Suman K. Banik,[2] Oleg Krichevsky,[3] and Ralf Metzler[1]
[1]*NORDITA, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark*
[2]*Dept. of Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0435, USA*
[3]*Physics Department, Ben Gurion University, Be'er Sheva 84105, Israel*
(Received 2 May 2006; published 21 September 2006)

We study the fluctuation dynamics of localized denaturation bubbles in heteropolymer DNA with a master equation and complementary stochastic simulation based on novel DNA stability data. A significant dependence of opening probability and waiting time between bubble events on the local DNA sequence is revealed and quantified for a biological sequence of the T7 bacteriophage. Quantitative agreement with data from fluorescence correlation spectroscopy is demonstrated.

DOI: 10.1103/PhysRevLett.97.128105 PACS numbers: 87.15.−v, 02.50.−r, 05.40.−a, 82.37.−j

The biological function of DNA largely relies on its physical properties: Protein binding is sensitive to local DNA structure [1], DNA looping facilitates the search of binding proteins for their specific site [2], and DNA knots impair transcription or act as barriers between different genome regions [3]. Similarly, local denaturation of DNA is necessary for protein binding to a DNA single strand [4–6], and is implicated in transcription initiation [7,8]. DNA melting has a long tradition in statistical physics [9]. Its biological relevance is due to the fact that the free energy for breaking a single base pair (bp) at physiological temperature is $\sim k_B T$ [10,11]. Renewed interest in DNA melting, from a physics perspective, is nourished by the possibility to measure the fluctuation *dynamics* of local denaturation bubbles by single molecule fluorescence correlation spectroscopy (FCS) [12].

We present a master equation (ME) and complementary stochastic simulation that provide the time series of the bubble fluctuations. A full two-variable formulation in terms of bubble size m and left fork location x_L allows us to investigate an arbitrary sequence of bp's, beyond previous homopolymer [5] and random energy models [13]. In certain limits, the ME can be solved analytically. We employ DNA stability data from a novel approach measuring the ten stacking interactions separately and, *inter alia* predicting a distinct asymmetry between AT-AT and AT-TA nearest-neighbor bp's [11]. As proved on recent FCS experimental data our model describes well the bubble dynamics with only one free parameter. We demonstrate the delicate sensitivity of bubble dynamics to the local sequence of heterogeneous DNA on the promoter sequence of the T7 bacteriophage, and illustrate good potential for nanosensor applications.

Model.—With typical experimental setups [12] in mind, we consider a segment of double-stranded DNA with M internal bp's, that are clamped at both ends (Fig. 1). The full sequence of bp's enters via the position-dependence of the statistical weights $u_{\rm hb}(x) = \exp\{\epsilon_{\rm hb}(x)/[k_B T]\}$ for breaking the hydrogen bonds of the bp at position x, and $u_{\rm st}(x) = \exp\{\epsilon_{\rm st}(x)/[k_B T]\}$ for disrupting the stacking interactions between bp's $x-1$ and x. Because of the high free energy barrier for bubble initiation ($\xi \ll 1$), opening and merging of multiple bubbles are rare events, such that a one-bubble description is appropriate. The positions x_L and x_R of the zipper forks correspond to the right- and left-most closed bp of the bubble. x_L and x_R are stochastic variables, whose time evolution in the energy landscape defined by the partition factor ($m \geq 1$)

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

characterizes the bubble dynamics. $\mathcal{Z}$ is written in terms of x_L and bubble size $m = x_R - x_L - 1$, with $\mathcal{Z}(m=0) = 1$. Here, $\xi' = 2^c \xi$, where $\xi \approx 10^{-3}$ is the ring factor for bubble initiation from Ref. [11] that is related to the cooperativity parameter $\sigma_0 \approx 10^{-5}$ [9,10] by $\sigma_0 = \xi \exp\{\epsilon_{\rm st}\}$ [11]. For the entropy loss on forming a closed polymer loop we assign the factor $(1+m)^{-c}$ [10,14] and take $c = 1.76$ for the critical exponent [15]. Note that a bubble with m open bp's requires breaking of m hydrogen bonds and $m+1$ stacking interactions.

The zipper forks move stepwise $x_{L/R} \to x_{L/R} \pm 1$ with rates $\mathrm{t}^{\pm}_{L/R}(x_L, m)$. We define for bubble size decrease

$$\mathrm{t}^{+}_{L}(x_L, m) = \mathrm{t}^{-}_{R}(x_L, m) = k/2 \qquad (m \geq 2) \qquad (2)$$

for the two forks [16]. The rate k characterizes a single bp

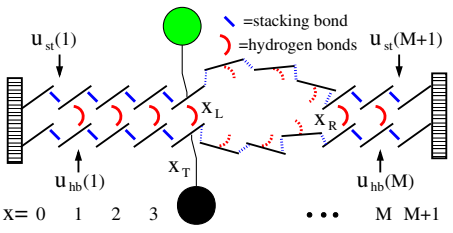

FIG. 1 (color online). Clamped bubble domain with internal bp's $x = 1$ to M, statistical weights $u_{\rm hb}(x)$, $u_{\rm st}(x)$, and tag position x_T.

zipping. Its independence of x corresponds to the view that bp closure requires the diffusional encounter of the two bases and bond formation; as sterically AT and GC bp's are very similar, k should not significantly vary with bp stacking. k is the only adjustable parameter of our model, and has to be determined from experiment or future MD simulations. The factor $1/2$ is introduced for consistency [5]. Bubble size increase is controlled by

$$\begin{aligned} \mathsf{t}_L^-(x_L, m) &= k u_{\mathrm{st}}(x_L) u_{\mathrm{hb}}(x_L) s(m)/2, \\ \mathsf{t}_R^+(x_L, m) &= k u_{\mathrm{st}}(x_R + 1) u_{\mathrm{hb}}(x_R) s(m)/2, \end{aligned} \tag{3}$$

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

$$\mathsf{t}_G^+(x_L) = k\xi' s(0) u_{\mathrm{st}}(x_L + 1) u_{\mathrm{hb}}(x_L + 1) u_{\mathrm{st}}(x_L + 2) \tag{4}$$

and $\mathsf{t}_G^-(x_L) = k$. The rates t fulfill detailed balance conditions. The annihilation rate $\mathsf{t}_G^-(x_L)$ is twice the zipping rate of a single fork, since the last open bp can close either from the left or right. Because of the clamping, $x_L \geq 0$ and $x_R \leq M + 1$, ensured by reflecting conditions $\mathsf{t}_L^-(0, m) = \mathsf{t}_R^+(x_L, M - x_L) = 0$. The rates t together with the boundary conditions fully determine the bubble dynamics.

In the FCS experiment fluorescence occurs if the bp's in a Δ neighborhood of the fluorophore position x_T are open [12]. Measured fluorescence time series thus correspond to the stochastic variable $I(t)$, that takes the value 1 if at least all bp's in $[x_T - \Delta, x_T + \Delta]$ are open, else it is 0. The time-averaged (overbar) fluorescence autocorrelation $A_t(x_T, t) = \overline{I(t)I(0)} - \overline{I(t)}^2$ for the sequence AT9 from [12] are rescaled in Fig. 2 [17].

ME.—DNA breathing is described by the probability distribution $P(x_L, m, t)$ to find a bubble of size m located at x_L whose time evolution follows the ME $\partial P(x_L, m, t)/\partial t = \mathbb{W} P(x_L, m, t)$. The transfer matrix $\mathbb{W}$ incorporates the rates t. Detailed balance guarantees equilibration toward $\lim_{t\to\infty} P(x_L, m, t) = \mathcal{Z}(x_L, m)/\mathcal{Z}$, with $\mathcal{Z} = \sum_{x_L, m} \mathcal{Z}(x_L, m)$ [5,18]. The ME and the explicit construction of $\mathbb{W}$ are discussed at length in Refs. [5,19]. Eigenmode analysis and matrix diagonalization produces all quantities of interest such as the ensemble averaged autocorrelation function $A(x_T, t) = \langle I(t)I(0)\rangle - (\langle I\rangle)^2$. $\langle I(t)I(0)\rangle$ is proportional to the survival density that the bp is open at t and that it was open initially [5,19].

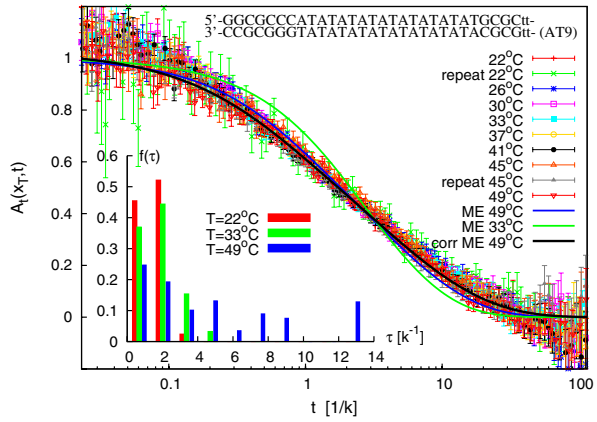

FIG. 2 (color online). Scaling plot of $A_t(x_T, t)$ at various T for the sequence AT9 from [12]. Inset: Relaxation time spectrum. See text.

In Fig. 2 the thick (blue online) curve shows the predicted behavior of $A(x_T, t)$, calculated for $T = 49\,^\circ\mathrm{C}$ with the parameters from Ref. [11]. As in the experiment we assumed that fluorophore and quencher attach to bp's x_T and $x_T + 1$, that both are required open to produce a fluorescence signal. From the scaling plot, we calibrate the zipping rate as $k = 7.1 \times 10^4/\mathrm{s}$, in good agreement with the findings from Refs. [12,20]. The calculated behavior reproduces the data within the error bars, while the model prediction at $T = 35\,^\circ\mathrm{C}$ shows more pronounced deviation. Potential causes are destabilizing effects of the fluorophore and quencher, and additional modes that broaden the decay of the autocorrelation. The latter is underlined by the fact that for lower temperatures the relaxation time distribution $f(\tau)$, defined by $A(x_T, t) = \int \exp(-t/\tau) f(\tau) d\tau$, becomes narrower (Fig. 2 inset). Deviations may also be associated with the correction for diffusional motion of the DNA construct, measured without quencher and neglecting contributions from internal dynamics [21]. Indeed, the black curve shown in Fig. 2 was obtained by a 3% reduction of the diffusion time [22]; see details in Ref. [19].

Stochastic simulation.—Based on the rates t, stochastic simulations give access to single bubble fluctuations [19,23]. Our customized Gillespie algorithm uses the joint probability density of waiting time τ and path $\mu = +/-$,

$$P(\tau, \mu, \nu) = \mathsf{t}_\nu^\mu(x_L, m) \exp\Big(-\tau \sum_{\mu,\nu} \mathsf{t}_\nu^\mu(x_L, m)\Big), \tag{5}$$

defining for given state (x_L, m) after what time τ the next step of fork $\nu \in G, L, R$ occurs. The formulation via the waiting time density $\sum_{\mu,\nu} P$ is economical computationally, avoiding a large number of unsuccessful opening attempts in traditional Langevin simulations. Using (5) we obtain the single bubble time series in Fig. 3.

Phage T7 analysis.—By ME and stochastic simulation we investigate the promoter sequence of the T7 phage,

```
    1                   20
    |                   |
5'-aTGACCAGTTGAAGGACTGGAAGTAATACGACTC
   AGTATAGGGACAATGCTTAAGGTCGCTCTCTAGGAg-3'
      | |                            |
      38 41                          68
```
(8)

whose TATA motif appears above the label "38" and is

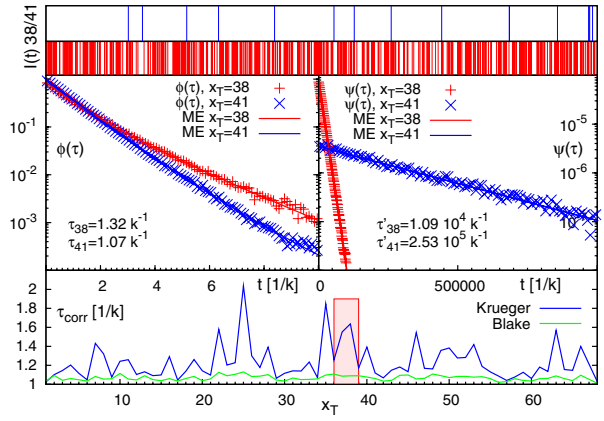

FIG. 3 (color online). Time series $I(t)$ for the T7 promoter, with $x_T = 38, 41$. Middle: Waiting time $[\psi(\tau)]$ and fluorescence time $[\phi(\tau)]$ densities. Bottom: Mean fluorescence time $\Delta = 0$.

marked gray [8]. Figure 3 shows the time series of $I(t)$ at 37 °C for the tag positions $x_T = 38$ in the core of TATA, and $x_T = 41$ at the second GC bp after TATA: Bubble events occur much more frequently in TATA (AT/TA bp's are particularly weak [11]). This is quantified by the density of waiting times $\psi(\tau)$ in the $I = 0$ state, whose characteristic time scale τ' is more than an order of magnitude longer at $x_T = 41$. In contrast, we observe similar behavior for the density $\phi(\tau)$ in the $I = 1$ state for $x_T = 38$ and 41. Both $\psi(\tau)$ and $\phi(\tau)$ decay exponentially for long t; the overlaid lines represent numerical evaluation of the ME; see Ref. [19]. As shown in the bottom for the parameters from Ref. [11], the variation of the mean correlation time $\tau_{\text{corr}} = \int A(x_T, t)dt$ obtained from the ME is small for the entire sequence, consistent with the low sequence sensitivity of $\phi(\tau)$. Note the even smaller variation predicted for the parameters of Ref. [10].

Figure 4 shows the equilibrium probability that the bp's $[x_T - \Delta, x_T + \Delta]$ are open, as necessary for fluorescence to occur. We plot data obtained from the zeroth mode of the ME together with the time average from the Gillespie algorithm (GA), finding excellent agreement. Whereas for $\Delta = 0$ several segments show increased tendency to opening, for the case $\Delta = 2$, one major peak is observed; the data from Ref. [11] coincide precisely with TATA, while the data from Ref. [10] peak upstream. Also shown is a comparison to the opening probability of a random sequence demonstrating that the enhanced opening probability at TATA is significant; compare Ref. [19]. Analysis for various Δ indicate best discrimination of the TATA sequence being open for $\Delta = 2$. For future FCS or energy transfer experiments, it therefore appears important to optimize the Δ dependence for best resolution, e.g., by adjusting the linker lengths of fluorophore and quencher.

Nanosensing.—Figure 5 shows the dependence of the mean correlation time of the AT9 sequence on salt concentration C and T. The variation with C and T is significant, pointing toward potential applications of DNA fluorescence constructs as nanosensors [24]. The triangles denote the melting concentration of infinitely long random AT and GC stretches, respectively, (see Ref. [11]). The maxima of the τ_{corr} curves hallmark the critical slowing down of the autocorrelation at the phase transition point beyond which the bubble is preferentially open; see also Ref. [25]. Note that the maxima coincide with the melting concentrations in the bottom panel. The dashed line ($\tau_{\max}$ 2D) corresponds to the longest relaxation time obtained numerically from the ME; it agrees well with τ_{corr} close to the maximum, analogously for the other T. The horizontal line ($\tau_{\max}$ 1D) represents the longest relaxation time $(2M + 1)^2/\pi^2 k^{-1}$ obtained from the homopolymer model of Ref. [5] in the limit $u \to 1$, $\sigma_0 \to 0$, and $c = 1$ ($M = 27$, length of the AT9 construct), with the same scaling as the first exit of unbiased diffusion.

Discussion.—Previous bulk melting studies provided DNA stability data [10,11], on whose basis the relation between local sequence stability and coding properties of the associated genes was shown [7,26]. However, it is single molecule experiments that permit to study the *dynamics* of DNA denaturation and renaturation [12]. We

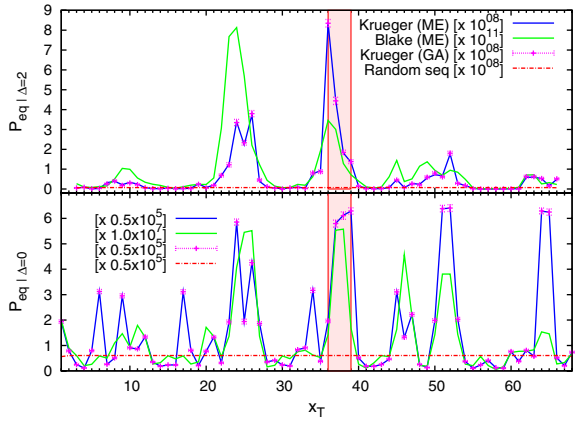

FIG. 4 (color online). Probability of having bp's $[x_T - \Delta, x_T + \Delta]$ open.

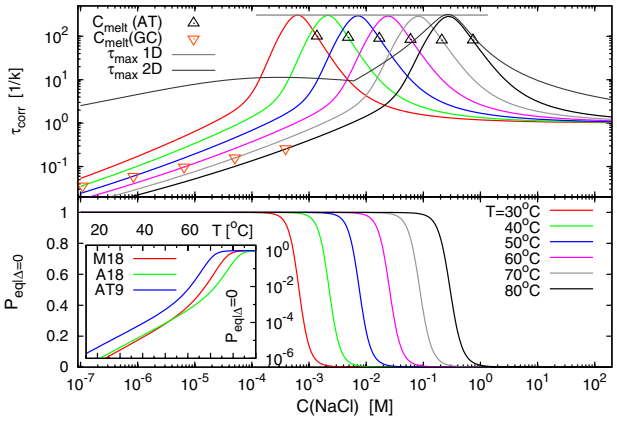

FIG. 5 (color online). Mean correlation time vs salt concentration C for various T (top), and melting curves vs C and T (bottom).

here derive a physical framework for the opening and closing fluctuations of intermittent DNA bubbles in an arbitrary sequence of bp's using the position of the two bubble zipper forks as fundamental coordinates. By comparison with previously unpublished FCS data we prove the predictive power of our model. As a complementary approach based on the same (un)zipping rates, we introduced the stochastic Gillespie simulation, that provides the time series of single bubble fluctuations. The time averages from the stochastic simulation agree well with the ensemble properties derived from the ME. By its computationally attractive formulation based on the waiting time the Gillespie approach allows to include additional effects such as protein binding dynamics, or to consider longer chains and multibubble states. For a long homopolymer our model is analytically tractable [5,19].

We used recent DNA stability data from Ref. [11] based on separation of hydrogen bond and stacking energies, a distinct feature being the low stacking in a TA/AT stack, translating into a pronounced instability of the TATA motif, as shown for the T7 promoter sequence. The relevance of stacking interactions is also shown in the inset in Fig. 5 exhibiting pronouncedly different melting behavior despite identical AT and GC contents for the constructs in Refs. [12,27]. Regarding the biological relevance of TATA, from our analysis it may be speculated that it is not primarily the bubble lifetime (typically shorter than the time scale of protein conformational changes) but the recurrence frequency of bubble events that triggers the initiation of transcription. Note that typical binding energies of TATA binding proteins exceed the free energy to break up TATA, while both energies are comparable for a random sequence of the same length [19].

Given the high sensitivity of bubble dynamics to the stability parameters it should be of interest to employ FCS on designed DNA constructs to more accurately obtain DNA stability data and to calibrate the (un)zipping rates.

We thank G. Altan-Bonnet and A. Libchaber for sharing the data for Fig. 2, M. Frank-Kamenetskii for discussion and access to the new stability data prior to publication, and M. A. Lomholt and K. Splitorff for discussion.

[1] B. Alberts *et al.*, *The Molecular Biology of the Cell* (Garland, New York, 1994).
[2] M. A. Lomholt, T. Ambjörnsson, and R. Metzler, Phys. Rev. Lett. **95**, 260603 (2005); G. Wuite (unpublished).
[3] T. J. Pollock and H. A. Nash, J. Mol. Biol. **170**, 1 (1983); P. Staczek and N. P. Higgins, Mol. Microbiol. **29**, 1435 (1998).
[4] K. Pant, R. L. Karpel, and M. C. Williams, J. Mol. Biol. **327**, 571 (2003).
[5] T. Ambjörnsson and R. Metzler, Phys. Rev. E **72**, 030901(R) (2005); J. Phys. Condens. Matter **17**, S1841 (2005).
[6] I. M. Sokolov, R. Metzler, K. Pant, and M. C. Williams, Biophys. J. **89**, 895 (2005); Phys. Rev. E **72**, 041102 (2005).
[7] E. Yeramian, Gene **255**, 139 (2000); **255**, 151 (2000).
[8] C. H. Choi *et al.*, Nucleic Acids Res. **32**, 1584 (2004); G. Kalosakas *et al.*, Europhys. Lett. **68**, 127 (2004).
[9] D. Poland and H. A. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers* (Academic, New York, 1970); R. M. Wartell and A. S. Benight, Phys. Rep. **126**, 67 (1985).
[10] R. D. Blake *et al.*, Bioinformatics **15**, 370 (1999); S. G. Delcourt and R. D. Blake, J. Biol. Chem. **266**, 15 160 (1991); J. SantaLucia, Jr., Proc. Natl. Acad. Sci. U.S.A. **95**, 1460 (1998).
[11] A. Krueger, E. Protozanova, and M. D. Frank-Kamenetskii, Biophys. J. **90**, 3091 (2006); E. Protozanova, P. Yakovchuk, and M. D. Frank-Kamenetskii, J. Mol. Biol. **342**, 775 (2004).
[12] G. Altan-Bonnet, A. Libchaber, and O. Krichevsky, Phys. Rev. Lett. **90**, 138101 (2003).
[13] T. Hwa *et al.*, Proc. Natl. Acad. Sci. U.S.A. **100**, 4411 (2003).
[14] M. Fixman and J. J. Freire, Biopolymers **16**, 2693 (1977).
[15] See C. Richard and A. J. Guttmann, J. Stat. Phys. **115**, 925 (2004), and references therein. The bubble dynamics is only marginally influenced by taking c slightly above 2.
[16] Because of intrachain coupling (e.g., Rouse), larger bubbles may involve an additional "hook factor" $m^{-\mu}$ [5].
[17] Except for Fig. 5, the parameters are for 100 mM NaCl.
[18] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1992).
[19] T. Ambjörnsson *et al.* (unpublished).
[20] Note that the zipping rate is the only free parameter in our model. Its value from the FCS studies is significantly lower than from NMR experiments [10^7–10^8/s, M. Guéron, M. Kochoyan, and J.-L. Leroy, Nature (London) **328**, 89 (1987)]. The difference may stem from the higher temperatures and longer AT sequences probed in the FCS experiments. However, a perturbing effect of the fluorophore-quencher pair cannot be excluded.
[21] O. Krichevsky and G. Bonnet, Rep. Prog. Phys. **65**, 251 (2002).
[22] For diffusion time $\tau_D = 150$ μs measured for an RNA construct of comparable length in Ref. [21].
[23] S. K. Banik, T. Ambjörnsson, and R. Metzler, Europhys. Lett. **71**, 852 (2005).
[24] T. Ambjörnsson and R. Metzler, J. Phys. Condens. Matter **17**, S4305 (2005); R. Metzler and T. Ambjörnsson, J. Comp. Theor. Nanosci. **2**, 389 (2005).
[25] D. J. Bicout and E. Kats, Phys. Rev. E **70**, 010902(R) (2004).
[26] E. Carlon, M. L. Malki, and R. Blossey, Phys. Rev. Lett. **94**, 178101 (2005).
[27] We also found from our analysis of the FCS data from Ref. [12] that the M18 and A18 compounds with B' domains of repeated AT bp's show different dynamics, similar to modified exchange rates of B' in bulk: C. Chen and I. M. Russu, Biophys. J. **87**, 2545 (2004).