

## Bubble Dynamics in Double-Stranded DNA

Grégoire Altan-Bonnet,<sup>1</sup> Albert Libchaber,<sup>1</sup> and Oleg Krichevsky<sup>1,2,\*</sup>

<sup>1</sup>Center for Studies in Physics and Biology, The Rockefeller University, New York, New York 10021

<sup>2</sup>Department of Physics, Ben-Gurion University, BeerSheva 84105, Israel

(Received 11 November 2001; published 1 April 2003)

We report the first measurement of the dynamics of bubble formation in double-stranded DNA. Fluctuations of fluorescence of a synthetic DNA construct, internally tagged with a fluorophore and a quencher, are monitored by fluorescence correlation spectroscopy. The relaxation dynamics follow a multistate relaxation kinetics, with a characteristic time scale of 50  $\mu$ s. A simple model of bubble dynamics based on constant zipping-unzipping rates is proposed to account for our experimental data. The role of different secondary structures stabilizing the open bubble is tested.

DOI: 10.1103/PhysRevLett.90.138101

PACS numbers: 87.15.Ya, 87.14.Gg, 87.15.He, 87.64.Ni

The structure of double-stranded DNA (dsDNA) is strongly stable due to the self-assembly of its many base pairs. Yet the interactions within each base pair are relatively weak (free energy less than  $2k_B T$  [1]), so that thermal excitations lead to DNA breathing, i.e., local denaturation and reclosing of the double-stranded structure [2,3]. Breathing fluctuations are intriguing from the physics point of view as an example of fluctuations in a quasi one-dimensional system [4,5] as well as from a biology perspective as limiting steps to DNA replication [6], transcription, denaturation [7], and protein binding [8]. While all of these processes imply simultaneous opening of many base pairs, the only available experimental technique to monitor DNA breathing dynamics (NMR of imino-proton exchange) measures the lifetime of a single base pair only [9]. Here we present a new approach combining fluorescence correlation spectroscopy (FCS) and fluorescence quenching in synthetic DNA molecules, and report the first measurements of the relaxational kinetics of the breathing modes. We observe multiexponential kinetics, well accounted for by a wide distribution of excited modes, with a typical relaxation time scale around 50  $\mu$ s at 25 °C.

Our samples are synthetic DNA constructs, containing two modified bases tagged with a fluorophore and a quencher (Fig. 1). These modifications are specifically designed so that, when the DNA structure is closed, fluorophore and quencher are in close proximity and the fluorescence is quenched; when the structure opens up, fluorophore and quencher are pulled apart and the fluorescence is restored [10]. Thus the base-pair fluctuations translate into fluorescence fluctuations. The correlation spectrum of the fluctuations, monitored by FCS [11,12], reveals their characteristic dynamics of relaxation, as demonstrated previously with hairpin conformational fluctuations [13].

All the constructs are hairpin loops, to forbid complete separation of the two complementary strands (Fig. 1): a 4-Thymidine (T) loop concatenates the two complementary strands into a 29-base-pair stem. The stem's double strand

contains one breathing domain made of adenosine-thymidine (AT) base pairs flanked by two regions composed of guanosine (G) and cytosine (C) bases, whose stability constrains the thermal modes to the less-stable AT region (Fig. 1). Having in mind that open bubbles might be stabilized by alternative secondary structures, we studied three constructs having the same GC-rich regions, but different 18-AT-base-pair regions. The first one (named  $M_{18}$ , of sequence 5'-GGCGCCCAA TATAAAATATTAATAATGCGCTTTTGGCGATTTTAA-TATTTTATTTGGGCGCC-3') contains a mixed AT sequence that does not readily produce secondary structures. The second one (named  $A_{18}$  of sequence 5'-GGCGCCCAAAAAAAAAATAAAAAAAAAAGCGCTTTT-GCGCTTTTTTTTATTTTTTTTTGGGCGCC-3') has a number of low energy states resulting from the different shifts of 5'- $A_{18}$ -3' strand with respect to 3'- $T_{18}$ -5'. The third one [named (AT)<sub>9</sub>, of sequence 5'-GGCGCCCATATATATATATATATATATGCGCTTTT-GCGCATATATATATATATATATATATATATGGGCGCC-3'] can in addition self-hybridize in the open bubble, to form cruciforms. We also synthesized the same constructs with end tagging: the fluorophore is coupled at the 5' end of the stem, the quencher being introduced at the 3' end. This end tagging is used as a control to check that the end-GC-clamp remains closed at the temperatures where AT region melts (Fig. 2).

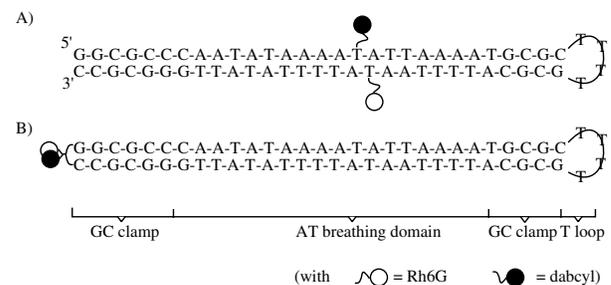


FIG. 1. Sketch of the ( $M_{18}$ ) DNA constructs: (A) with internal tagging; (B) with end tagging.

All DNA constructs were synthesized using standard cyanoethyl phosphoramidite chemistry (Midland Certified Reagent, TX), with a 4-((4-(dimethyl amino)-phenyl)azo) benzoic acid (DABCYL) either on a modified thymine base (**T** for the internally tagged constructs) or on the 3' end (for the end-tagged constructs), and a primary amine via a spacer of six carbons either on a modified thymine base (**T** for the internally tagged constructs) or on the 5' end (for the end-tagged constructs). The fluorophore was conjugated by reaction of a succinidymil ester of Carboxy-Rhodamine 6G (Molecular Probes, OR) onto the primary amines, followed by gel filtration, reverse-phase high-pressure liquid chromatography and ethanol precipitation [13].

We first measured the melting curve of these constructs by monitoring their fluorescence as a function of temperature. The DNA constructs were resuspended at a concentration of 10 nM, in 0.1 M sodium chloride, 10 mM sodium cacodylate, 1 mM ethylene diamine tetra acetate pH 8.0. Their melting curve was recorded according to the method presented in [13] (Fig. 2). The comparison of the melting curves show that, between 20 °C and 70 °C, the AT domains are melting (inducing a separation of fluorophore and quencher, and a high fluorescence in the internally tagged construct), while the GC clamp remains closed (low fluorescence of the end-tagged construct). The GC clamps melt above 75 °C, restoring the full fluorescence of the fluorophore. Thus, in a wide range of temperatures the melting dynamics is confined to the AT region of the molecules, as designed.

In order to verify that the dyes do not induce significant disruption of the constructs' secondary structure, we checked that the melting curves of nonlabeled constructs measured through UV-absorption are similar to those of dye-tagged constructs, measured by fluores-

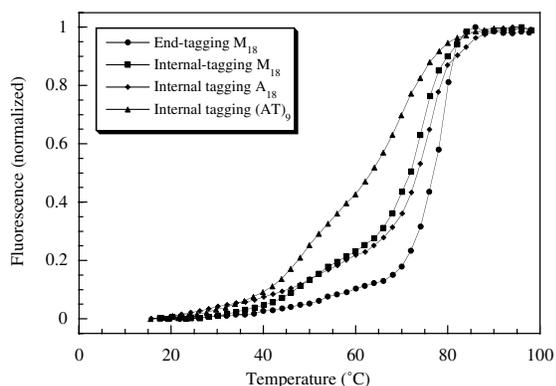


FIG. 2. Normalized fluorescence melting curves of the DNA constructs with different AT-rich breathing domains. The melting curve for the end-tagged construct is very sharp, corresponding to the all-or-none melting of the GC clamp. The melting curves for the internally tagged constructs display two transitions, as measured by the dequenching of the internal dye.

cence dequenching, consistent with measurements presented in [16].

To measure the opening-closing dynamics of dsDNA, we then carried out FCS measurements on the internally tagged constructs, with a setup and method introduced elsewhere [13]. The correlation functions  $G_{it}(t) = (\langle I(0)I(t) \rangle_t - \langle I(0)^2 \rangle) / \langle I(0) \rangle^2$  of the fluorescence intensity  $I(t)$  collected from a solution of internally tagged constructs, were divided by the correlation function  $G_{\text{control}}$  measured on molecules lacking the quencher DABCYL.  $G(t) = G_{it}(t) / G_{\text{control}}(t)$  eliminates the diffusion contribution in  $G_{it}(t)$ , and is a measure of the correlation function of the fluorescence-quenching fluctuations, associated with the sole conformational fluctuation of our DNA molecules. For a more complete description of our method, and the use of a control construct, please refer to [15].

We notice that generally, the shape of  $G(t)$  is governed by the faster of the two processes, opening or closing. The experiments were performed between 20 and 50 °C, i.e., below the melting temperature  $T_m$  of the AT region (see Fig. 2). In this range of temperature, the kinetics of closing of bubbles is faster than that of opening, and therefore  $G(t)$  reflects mainly the closing kinetics.

Typical correlation functions of the fluctuations are presented in Fig. 3, where the background level is subtracted from  $G(t)$  and their amplitude normalized to 1. We would like to emphasize the three salient features of these correlation functions: (i) the overall characteristic relaxation time scales between 30 and 100  $\mu\text{s}$  are 3–4 orders of magnitude higher than estimations by NMR for the open base-pair lifetime, (ii) the correlation function is obviously not a single exponential: it is a superposition of many relaxation modes from 1  $\mu\text{s}$  up to 1 ms, and (iii) the correlation functions for all the constructs, at all the temperatures follow the same universal temporal behavior: presented as a function of rescaled time they all collapse into a single universal curve  $g(u) = G(t/t_{1/2})$ , where  $t_{1/2}$  is such that  $G(t_{1/2}) = 0.5$  [Fig. 3(b)]. This experimental result indicates that a single mechanism must account for the generic shape of this multiexponential relaxation, in all of the samples for all temperatures.

A two-state system (state being either open or closed) would have been characterized by a single-exponential correlation function, and our experimental method has been tested in this situation extensively [13]. Thus, the breathing fluctuations are not limited to a single base pair, and bubbles of various sizes are activated. This is consistent with  $T < T_m$ , if the energy  $E_{\text{init}}$  to initiate a bubble is much larger than  $k_B T$  (a well-known fact [15,16]), while the energy  $\epsilon$  to extend the existing bubble by one base pair is smaller than  $k_B T$ . Thus most of the time the DNA structure is kept closed due to the high  $E_{\text{init}}$ , but once the bubbles are formed they are not limited to a single base pair due to the low  $\epsilon$ . In fact, a very simple model of the bubbles' dynamics based on constant zipping

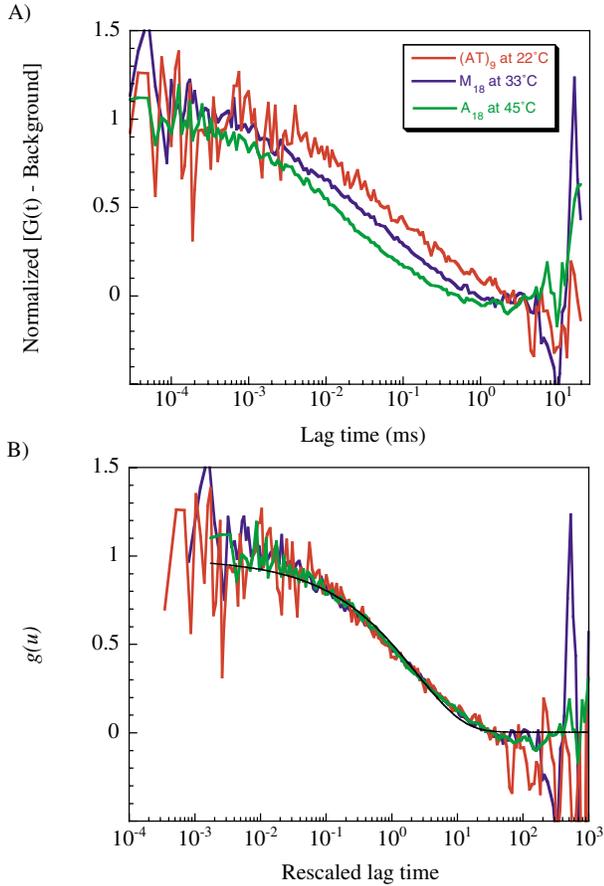


FIG. 3 (color). (A) Autocorrelation function  $G(t)$  for internally tagged constructs: (green)  $A_{18}$  construct at 45°C, (blue)  $M_{18}$  construct at 33°C and (red)  $(AT)_9$  construct at 22°C. (B) Rescaled autocorrelation functions  $g(u)$ . The black line represents  $g(t)$  [Eq. (3) derived from our model].

and unzipping rates can account for the shape of the correlation function in Fig. 3. We present here the outline of the derivation, while the detailed calculation is presented in the supplementary material [17].

We assume that the bubbles can appear only in the  $AT$  region of the constructs, and we consider the distribution of bubbles  $\{b_n\}$ , where  $n$  is the number of open bases. Once formed, a bubble of  $n$  bases can grow into  $(n+1)$  or shrink into  $(n-1)$ , with the respective rates  $k_+$  and  $k_-$ . Then  $b_n$  follows the classical rate equation: for  $0 < n < N$ ,

$$\frac{db_n}{dt} = k_+ b_{n-1} - (k_- + k_+) b_n + k_- b_{n+1}, \quad (1)$$

where  $N = 18$  is the number of  $AT$  base pairs.

$G(t)$  describes the kinetics of fluctuations in the number of open bubbles, which is determined by the two parallel processes, opening or closing. Below the melting temperature, the kinetics of closing is faster than opening, and thus, the decay of bubbles governs the behavior of  $G(t)$ :  $G(t) \propto B(t)$ , where  $B(t) = \sum b_n(t)$  is the number of open bubbles which existed at  $t = 0$  and survived until

time  $t$ . To obtain  $B(t)$  we solve Eq. (1) with the equilibrium distribution as initial condition  $b_n(0) \propto (k_+/k_-)^n \propto e^{-\epsilon n/(k_B T)}$  and  $b_0(t) = b_{N+1}(t) = 0$  as boundary condition. Assuming low extension energy  $\epsilon/k_B T = \ln(k_-/k_+) < 1$ , we can use the continuous limit. Assuming also that bubbles on average are much smaller than  $N$ , i.e.,  $\epsilon N > k_B T$ , we can set  $N \rightarrow \infty$  in the boundary conditions. Then  $b_n(t)$  becomes  $b(n, t)$ , which satisfies:

$$\frac{\partial b}{\partial t} = \frac{k_- + k_+}{2} \frac{\partial^2 b}{\partial n^2} + (k_- - k_+) \frac{\partial b}{\partial n}, \quad (2)$$

$$b(0, t) = b(\infty, t) = 0,$$

with the stationary initial condition  $b(n, 0) \propto \exp[-2(k_- - k_+)n/(k_- + k_+)]$ .

As the set of solutions of this drift-diffusion equation is known [18], the dependence  $b(x, t)$  can be obtained by standard methods [17]. Then,

$$G(t) \propto B(t) = B(0) - \frac{k_- + k_+}{2} \int_0^t \left( \frac{\partial b}{\partial n} \right)_{n=0} dt',$$

$$G(t) \propto \left( 1 + \frac{t}{2\tau} \right) \times \operatorname{erfc} \left( \sqrt{\frac{t}{4\tau}} \right) - \sqrt{\frac{t}{\pi\tau}} e^{-t/4\tau}, \quad (3)$$

with  $\operatorname{erfc}(u) = 1 - \frac{2}{\sqrt{\pi}} \int_0^u e^{-x^2} dx$ , and  $\tau = [(k_- + k_+)/2(k_- - k_+)]^2$ .

This model of fluctuations of dsDNA accounts well for the "universal" shape of  $G(t)$ , as shown with the match of this junction with our data in Fig. 3(b).

Going from (1) to (2) we used two conditions which set limits on  $\epsilon$ : to ensure the coexistence of bubbles of various sizes,  $\epsilon < k_B T$  and to confine the bubbles within the  $AT$ -rich domain,  $\epsilon N > k_B T$ . Thus, for  $N = 18$ ,  $\epsilon$  must be in the range between 0.05 and 1  $k_B T$ . As no experiment has measured this extension energy  $\epsilon$ , we compare it with the literature data for the denaturation energy of a single base pair: these data give somewhat higher values around 1–2  $k_B T$  [1]. However, these values have been classically extracted from melting curves of short DNA oligonucleotides, monitoring the complete denaturation of dsDNA at high temperature ( $> 70^\circ\text{C}$ ). Our experimental approach is sensitive to different modes of DNA deformation, where the strand separation in the open bubble remains constrained by the adjacent double-stranded regions. These modes do not require complete DNA denaturation and hence require less energy. They are more relevant to a biological situation where bubbles open within large dsDNA, at temperatures where the full denaturation of the complementary strands is not complete.

Likewise there is no direct contradiction between NMR estimations [9], the rate for the closing of an unbounded base pair ( $\sim 10^8 \text{ s}^{-1}$ ), and our results ( $\sim 10^4$ – $10^5 \text{ s}^{-1}$ ): NMR distinguishes the open state through the feasibility of imino-proton exchange. This might require a much smaller actual conformational change (and hence higher respective rates) as compared

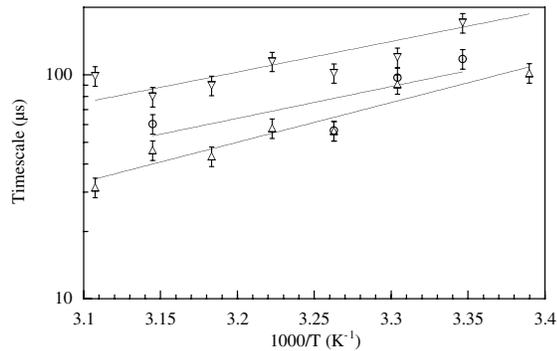


FIG. 4. Arrhenius plot of the breathing time scales  $\tau$  for the constructs with different breathing domains: constructs (○)  $M_{18}$ ; (△)  $A_{18}$ ; (▽)  $(AT)_9$ . Error bars were derived from the fit of the correlation function with errors estimated from at least 60 independent measurements.

to the change needed to separate fluorophore and quencher. Also, NMR measures the average base-pair closing rates only, which must be strongly biased by the fast fluctuations at the level of individual bases. Contrary to NMR, our method is mostly sensitive to the relaxation of large and long-lived bubbles. The slow relaxation of the bubbles as well as the difference between the denaturation energy and  $\epsilon$  imply that the open regions are not completely denatured but rather have some underlying structure which stabilizes the large bubbles.

Three possible mechanisms can stabilize a bubble in DNA: (a) stacking of bases in the single-stranded domain, (b) mismatched reclosing of a double-stranded domain, or (c) formation of hairpin loops [19]. To assess their respective relevance, we specifically designed our three constructs to be prone to undergo only mechanism (a) for construct  $M_{18}$ , mechanisms (a),(b) for construct  $A_{18}$  and mechanisms (a),(b),(c) for construct  $(AT)_9$ . All the relaxation dynamics of the three constructs can be fitted well with Eq. (3) and yield comparable relaxation time scales (Fig. 4). Moreover, the temperature dependence of the characteristic fluctuation time scale obeys an Arrhenius law with a similar activation enthalpy of  $\sim 7$  kcal/mol (Fig. 4). Thus, even if the stability of hairpin-loop structures [expected in the construct  $(AT)_9$ ] may explain a slightly more stable open state (Figs. 2 and 4), the relaxation process in the three constructs must be essentially limited by the same physical barrier, i.e., the base destacking in the open domain.

In conclusion, we have presented the first measurements of the fluctuation dynamics of DNA breathing modes. The most striking feature of this dynamics is its long characteristic time scale, in the 20–100  $\mu$ s range. The relaxation follows a multiexponential kinetics in a wide range of temperatures, which implies that bubbles of many sizes are formed. The shape of their relaxation is consistent with a constant zipping rate and a small ex-

tension energy [estimated to be in  $(0.05-1.0)k_B T$  range]. The relaxation time scales are only weakly sensitive to the formation of hairpin loops, but are dominated by the stabilization induced by the base stacking in the open bubble. To sum up into a simple picture, bubbles of 2 to 10 base pairs with lifetimes in the 50  $\mu$ s range spontaneously open in dsDNA at 37 °C, under low salt condition (0.1 M). The existence of these long-lived fluctuating bubbles adds a new and interesting dimension to the dynamical picture of DNA behavior and of DNA-protein interactions.

This work is supported by the Mathers Foundation. G.B. and O.K. were supported by the Burrough-Wellcome Foundation. We thank Amit Meller and Didier Chatenay for discussion and comments on the manuscript.

\*Corresponding author.

- [1] J. SantaLucia, Jr., Proc. Natl. Acad. Sci. U.S.A. **95**, 1460 (1998).
- [2] M. Peyrard, Europhys. Lett. **44**, 271 (1998).
- [3] A. Campa, Phys. Rev. E **63**, 21901 (2001).
- [4] T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **47**, R44 (1993).
- [5] R. M. Fye and C. J. Benham, Phys. Rev. E **59**, 3408 (1999).
- [6] A. Kornberg and T. A. Baker, *DNA Replication* (W. H. Freeman, New York, 1992).
- [7] Y. Kafri, D. Mukamel, and L. Peliti, Phys. Rev. Lett. **85**, 4988 (2000).
- [8] J. F. Léger *et al.*, Proc. Natl. Acad. Sci. U.S.A. **95**, 12295 (1998).
- [9] M. Guéron and J. L. Leroy, Methods Enzymol. **261**, 383 (1995).
- [10] S. Tyagi and F. R. Kramer, Nat. Biotechnol. **14**, 303 (1996).
- [11] D. Magde, E. Elson, and W. W. Webb, Phys. Rev. Lett. **29**, 705 (1972).
- [12] R. Rigler *et al.*, Eur. Biophys. J. **22**, 169 (1993).
- [13] G. Bonnet, O. Krichevsky, and A. Libchaber, Proc. Natl. Acad. Sci. U.S.A. **95**, 8602 (1998).
- [14] J. Telser, K. A. Cruickshank, L. E. Morrison, and T. L. Netzel, J. Am. Chem. Soc. **111**, 6966 (1989).
- [15] B. R. Amirikyan, A. V. Vologodskii, and Y. Lyubchenko, Nucleic Acids Res. **9**, 5469 (1981).
- [16] C. J. Benham, J. Mol. Biol. **225**, 835 (1992).
- [17] See EPAPS Document No. E-PRLTAO-90-028310 for full analytical derivation of formula (3). A direct link to this document may be found in the online article's HTML reference section. The document may also be reached via the EPAPS homepage (<http://www.aip.org/pubservs/epaps.html>) or from <ftp.aip.org> in the directory `/epaps/`. See the EPAPS homepage for more information.
- [18] Z. Farkas and T. Fülöp, J. Phys. A **34**, 3191 (2001).
- [19] V. A. Bloomfield, D. M. Crothers, and I. Tinoco, *Nucleic Acids: Structures, Properties, and Functions* (University Science Books, Sausalito, CA, 2000).