Dynamics of Force-Induced DNA Slippage

Richard A. Neher* and Ulrich Gerland†

Department of Physics and CENS, LMU München, Theresienstrasse 37, 80333 München, Germany

(Received 7 July 2004; published 1 November 2004)

We study the base pairing dynamics of DNA with repetitive sequences where local strand slippage can create, annihilate, and move bulge loops. Using an explicit theoretical model, we find a rich dynamical behavior as a function of an applied shear force f: reptationlike dynamics at $f = f_c$ with a rupture time τ scaling as N^3 with its length N, drift-diffusion dynamics for $f_c < f < f^*$, and a *dynamical* transition to an unraveling mode of strand separation at $f = f^*$. We predict a viscoelastic behavior for periodic DNA with time and force scales that can be *programmed* into its sequence.

DOI: 10.1103/PhysRevLett.93.198102 PACS numbers: 87.15.–v, 05.10.Gg, 87.80.Fe

The dynamics of base pairing in DNA and RNA molecules plays an important role in biological processes such as DNA replication, transcription, and RNA folding [1]. These dynamics can be probed in detail with modern single molecule techniques to exert and measure piconewton forces with nanometer spatial resolution [2]. For instance, double-stranded DNA (dsDNA) can be forced to open either by pulling on the two strands from the same end of the dsDNA ("unzipping") [3–5] or from opposite ends ("shearing") [6]. In the case of unzipping, the dynamics involves the consecutive opening of *native* base pairs, i.e., those present in the ground state of the molecule, and is well understood theoretically [7]. Here, we consider instead the shearing of dsDNA and focus specifically on *periodic* DNA sequences. This case is particularly interesting both from a physical and a biological point of view, since (i) periodic sequences have many non-native base pairing conformations where one strand is shifted with respect to the other, (ii) shearing probes the transitions between such states, i.e., the dynamics of DNA *slippage*; see Fig. 1, and (iii) DNA slippage during genome replication allows the expansion of nucleotide repeats, and, for certain repeats inside genes, triggers a variety of diseases including Huntington's disease [8].

The mechanism for DNA slippage has already been suggested by Pörschke [9]; see Fig. 1(a): small bulge loops can form at one end of the molecule when a few bases spontaneously unbind and rebind shifted by one or several repeat units. Once formed, a bulge loop may diffuse along the molecule and anneal at the other end, effectively sliding the two strands against each other by a length equal to the size of the bulge loop. This mechanism involves only small energetic barriers compared to the large barrier for complete unbinding and reassociation. Here, we present a detailed theoretical study of *force-induced* DNA slippage, which has so far not been studied experimentally. We show that this system displays a rich dynamical behavior that can be controlled experimentally by adjusting the force, sequence length, and sequence composition.

Model.— We consider a dsDNA of two perfectly complementary periodic sequences with N repeat units, each consisting of m nucleotides (for simplicity, we refer to repeat units also as "bases"). Assuming that base pairing within a strand is negligible, a base pairing configuration is specified by the set of the $n \le N$ interstrand base pairs $S = \{(u_i, l_i)\}$ with $1 \le u_1 < u_2 < \ldots < u_n \le N$ for the "upper" strand and analogously for the l_i in the "lower" strand. We assign a binding energy $-\varepsilon_b < 0$ to each base pair and a loop cost $E_\ell(j) > 0$ when there are $j > 0$ unpaired bases (total on both strands) between two consecutive base pairs. With a constant shear force f, see Fig. 1(b), the energy of a configuration S is

$$E[S] = -\varepsilon_b n[S] + \sum_{i=2}^{n[S]} E_\ell(\Delta u_i + \Delta l_i - 2) - fL[S], \quad (1)$$

where $\Delta u_i = u_i - u_{i-1}$ and $\Delta l_i = l_i - l_{i-1}$. The loop cost $E_\ell(j)$ increases with the loop length, starting from $E_\ell(0) = 0$. Free DNA ($f = 0$) is described by $E_\ell(j) = \varepsilon_\ell + 3\nu k_B T \ln(j)$, with a loop initiation cost $\varepsilon_\ell > 0$ and a logarithmic asymptotic behavior derived from polymer theory ($\nu \approx 0.6$ is the Flory exponent) [10]. An applied force can affect $E_\ell(j)$; however, our qualitative results are insensitive to its precise form [11]. Unless stated otherwise, we keep only the constant term, $E_\ell(j > 0) = \varepsilon_\ell$, for simplicity. The total extension L is

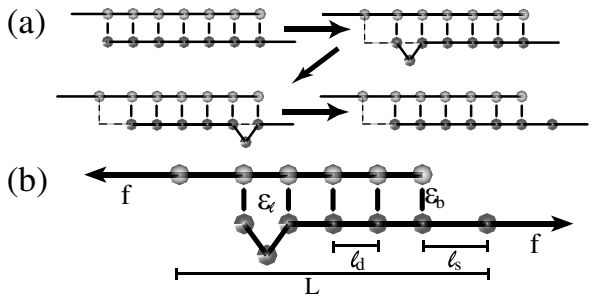

FIG. 1. Sketch of periodic dsDNA, where each bead represents one repeat unit consisting of one or several bases. (a) Many microscopic slippage events can lead to macroscopic sliding. (b) An applied shear force.

$$L[S] = \ell_s(u_1 - 1 + N - l_n) + \ell_d \sum_{i=2}^{n} \min(\Delta u_i, \Delta l_i), \quad (2)$$

where ℓ_d and $\ell_s > \ell_d$ are the effective lengths (in the direction of the force) per single and double-stranded unit, respectively. The entropic elasticity of DNA [12] causes both ℓ_d and ℓ_s to depend on the applied force; however, since the DNA is almost fully stretched at the forces of interest here, we use the constant values $\ell_d/m =$ 3.4 Å and $\ell_s/m = 7$ Å for simplicity [13].

We study the dynamics of our model with analytical methods (see below) and a Monte Carlo approach using three elementary moves [14]: opening, closing, and slippage of a base pair; i.e., a pair (u_i, l_i) is removed from the set S or added to it, or, if the base pair is adjacent to a loop, either u_i or l_i can be changed to another base inside the loop. The absolute time scale of these dynamics is hard to predict, but comparison with bulk reannealing experiments [9] suggests that our simulation time step is on the order of μs in real time.

Scaling of mean rupture times.— With a constant applied force $f > 0$, eventually every finite dsDNA will rupture, since complete separation of the strands ($L \to \infty$) is the state of minimal free energy. However, both the time scale and the nature of the rupture dynamics depend drastically on the force. Figure 2 displays the scaling of the mean rupture time $\langle\tau\rangle$ with the number of bases N for a number of different forces (see caption for parameters). We observe four distinct asymptotic behaviors: an exponential increase with N for small forces, a cubic scaling

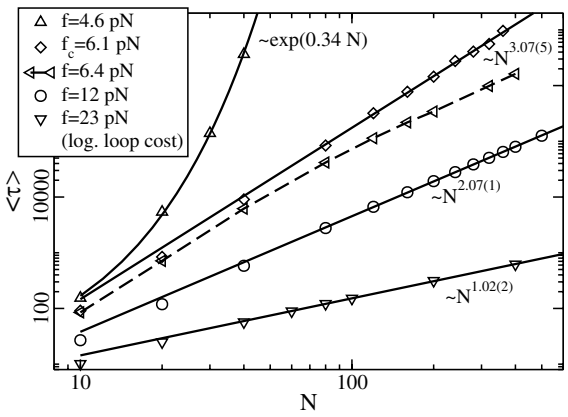

FIG. 2. Scaling of the mean rupture time $\langle\tau\rangle$ with the number of bases N for different shear forces (with $\varepsilon_b = 1.11$, $\varepsilon_\ell = 2.8$, which roughly corresponds to adenine-thymine (AT) sequences at 50 °C; see Fig. 5). The symbols represent Monte Carlo data (error less than symbol size). The solid lines for $f \geq f_c$ are power law fits (exponent with error in least significant digit is given; data with $N \leq 40$ show significant finite size deviations and are excluded). For $f < f_c$ the rupture time increases exponentially. The data for $f = 6.4$ pN $\gtrsim f_c$ (connected by the dashed line) demonstrate the crossover from diffusive to drift behavior; see main text . The data for $f = 23$ pN are calculated including the logarithmic loop cost, which becomes relevant at large forces [11].

with N at a certain force f_c, a nearly quadratic scaling above f_c but below a second threshold f^*, and linear scaling above f^*. The behavior in the two extremes is easily interpreted: for small f, rupture is driven by thermal fluctuations across a large free energy barrier with an associated Kramers time that scales exponentially with N, and linear scaling at large f is expected when individual bonds break sequentially at a constant rate. We now characterize the rich behavior in the intermediate force regime, including the nature of the two transitions.

The thermodynamic energy barrier disappears at a force f_c, which can be estimated by balancing the binding energy per base pair with the mechanical work exerted when sliding both strands against each other by one step,

$$f_c \approx \varepsilon_b/(2\ell_s - \ell_d). \quad (3)$$

f_c is a critical force in the thermodynamic sense, if the state of complete rupture is excluded (see below for the exact calculation including all base pairing configurations). At $f = f_c$, the rupture dynamics is best understood by analogy with the reptation problem [15], since bulge loops in the DNA structure behave similarly to the "stored length" excitations of a single chain in a polymer network: these excitations are generated at the ends of the polymer with constant rate independent of N, diffuse along the polymer, and reach the other end with a probability $\sim N^{-1}$. Therefore, the macroscopic diffusion constant for the relative motion of the two DNA strands should scale as $D \sim N^{-1}$ and the time for diffusion over distance N is $\sim N^3$.

For $f > f_c$, strand separation is energetically a downhill process, which induces a drift velocity v between the two strands. In linear response, we expect $v = \mu\Delta f$ for small $\Delta f = f - f_c$ with a mobility mediated by bulge loop diffusion, $\mu = D/k_BT \sim N^{-1}$ (from the Einstein relation), leading to $\langle\tau\rangle \sim N^2$. Why does this behavior not persist for large forces? The second transition in the scaling behavior is due to a change in the rupture *mode*: at forces larger than $f^* \approx \varepsilon_b/(\ell_s - \ell_d)$, the double strand can open by *unraveling* from both ends; i.e., the energy cost ε_b of opening a base pair at the end is outweighed by the gain $f(\ell_s - \ell_d)$ from a longer base-to-base distance in the single strand. In this unraveling mode, the rupture time scales linearly with N. The dynamical transition from sliding to unraveling is clearly reflected in the length at rupture, $L[S(\tau)]$; see Fig. 3(a), which is roughly a factor of 2 larger for sliding.

Rupture time distributions.— Single molecule setups are ideally suited to record the full distribution of rupture times, $P(\tau)$, which is a sensitive characteristic of the dynamics and permits a close examination of the physical picture introduced above. The histograms in Fig. 4 show $P(\tau)$ from simulations at $f = f_c$ and a larger force $f_c < f < f^*$; see caption for parameters. We observe that fluctuations play a dominant role at $f = f_c$, i.e., the width of $P(\tau)$ is comparable to the mean, while the rupture

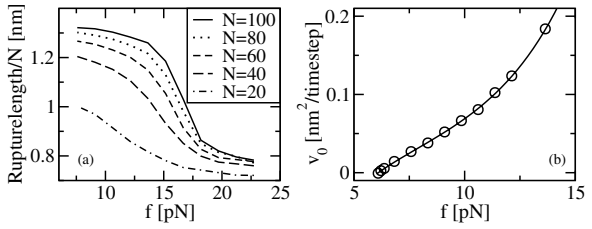

FIG. 3. (a) Rupture length as a function of applied force f (parameters as in Fig. 2). (b) Drift coefficient $v_0(f)$ extracted from simulations with $N = 150$ (circles) and analytical curve (solid line, $k_0 = 1.87$); see main text.

dynamics is drift dominated at the larger force, with a localized peak in $P(\tau)$.

To formulate the drift-diffusion dynamics quantitatively, we treat the number of bases in the double-stranded region as a continuum variable x with $0 < x < N$, and consider the probability distribution $\mathcal{P}(x, t)$, which satisfies the continuity equation $\partial_t \mathcal{P}(x, t) = -\partial_x j(x, t)$ with a force-dependent current

$$j(x, t) = -D(f, x)\partial_x \mathcal{P}(x, t) - v(f, x)\mathcal{P}(x, t). \quad (4)$$

The above discussion suggests a diffusion coefficient of the form $D(f, x) = D_0(f)/x$ and similarly a drift $v(f, x) = v_0(f)/x$. We have an absorbing boundary at $x = 0$ and it is natural to choose a reflecting boundary at $x = N$ and a delta peak at $x = N$ as initial condition. The solution $\mathcal{P}(x, t)$, which must in general be obtained numerically, determines the rupture time distribution through $P(\tau) = j(0, \tau)$.

We can determine the force dependence of the diffusion coefficient and drift empirically by fitting the calculated $P(\tau)$ to the simulation data using D_0 and v_0 as adjustable parameters. The solid lines in Fig. 4 show that the drift-diffusion theory describes the simulation data well. Figure 3(b) shows the fitted v_0 as a function of f (circles). The drift vanishes at the critical force, $v_0(f_c) = 0$, confirming the physical picture. The drift-diffusion theory also explains the crossover behavior in the vicinity of $f = f_c$; see Fig. 2: the drift is significant only when the system size N is larger than the diffusive length D_0/v_0 [16]. Hence, with $v_0 \sim \Delta f$, reptationlike dynamics is expected in a force interval $\delta f \sim N^{-1}$ around f_c.

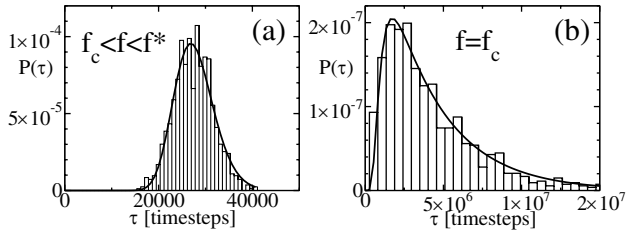

FIG. 4. Histogram of rupture times for two different forces, (a) 34.4 pN and (b) $f_c = 16.6$ pN, but the same set of DNA parameters, $N = 80$, $\varepsilon_b = 3.75$, $\varepsilon_\ell = 2.6$, which roughly correspond to a CG sequence at room temperature; see Fig. 5.

Microscopic dynamics.— Next, we study how the macroscopic drift in Eq. (4) emerges from the microscopic bulge loop dynamics and determine $v_0(f)$ in terms of our system parameters. Since bulge loops on opposite strands annihilate each other when they meet, the bulge loop dynamics is equivalent to a reaction-diffusion system of particles and antiparticles in one dimension. Both particles and antiparticles are created at each end, however, with different rates determined by the applied force. From the underlying master equation for these processes one obtains the mean-field equations [11]

$$\begin{aligned} \partial_t u(y, t) &= k_0 \partial_y^2 u(y, t) - k_1 u(y, t) l(y, t) + k_2, \\ \partial_t l(y, t) &= k_0 \partial_y^2 l(y, t) - k_1 u(y, t) l(y, t) + k_2, \end{aligned} \quad (5)$$

where $u(y, t)$ and $l(y, t)$ denote the bulge loop density on the upper/lower strand, $y \in [0, x]$ is the position within the double-stranded region, and k_0, k_1, k_2 are the rates for hopping, annihilation, and pair creation, respectively. At the boundaries, the densities take on constant values, $u(0, t) = l(x, t) = \rho_<$ and $u(x, t) = l(0, t) = \rho_>$, where $\rho_<(f)$ and $\rho_>(f)$ are calculated below by assuming a local equilibrium of the DNA at the edges. The macroscopic drift is determined by the stationary solution and depends only on the difference between the loop densities on the upper/lower strand, $v(f, x) = k_0 \partial_y [u(y) - l(y)]$. Using Eq. (5), this yields $v_0(f) = 2k_0[\rho_>(f) - \rho_<(f)]$. Figure 3(b) shows that this result is in excellent agreement with the empirical $v_0(f)$ obtained above.

Since the loop cost $E_\ell(j)$ is larger for two separate loops than for a single one of the combined length, bulge loops on the same strand feel a short-range attraction. However, the interaction is not strong enough to cause a significant aggregation of the loops in our Monte Carlo simulations. This is consistent with the observation that with our DNA parameters, the interaction energy ε_ℓ is never significantly larger than the entropic cost $\sim \log\rho$ of colocalization at loop density ρ. While $v_0(f)$ is apparently robust to interaction effects, the diffusion coefficient $D_0(f)$ is sensitive to interactions as well as correlations. Both are neglected in Eq. (5), leaving the calculation of $D_0(f)$ as a challenge for the future.

Critical force.— To obtain the exact critical force, we need the partition function $Z = \sum e^{-E[S]/k_B T}$ summed over all configurations S with at least one base pair. It is useful to allow for different numbers of bases in the two strands, e.g., $1 \le u_i \le N$ and $1 \le l_i \le M$, with a corresponding partition function

$$Z(N, M) = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} b_s^{i+j} \sum_{n=1}^{N-i} \sum_{m=1}^{M-j} Z_p(n, m), \quad (6)$$

where $b_s = e^{f\ell_s/k_B T}$ is the Boltzmann factor for a stretched base, and $Z_p(n, m)$ is the partition function for the central, double-stranded section starting with the first and ending with the last base pair cf. Figure 1(b). We calculate $Z_p(n, m)$ recursively by introducing a comple-

(a)

Sequence		20 °C	37 °C	50 °C
AA....	ϵ_b	2.37	1.62	1.11
TT....	ϵ_l	4.16	3.1	2.8
CC....	ϵ_b	3.75	2.98	2.46
GG....	ϵ_l	2.6	3.5	4.16
CAGCAG....	ϵ_b	10.5	8	6.36
GTCGTC....	ϵ_l	11.8	10.9	10.2

(b)

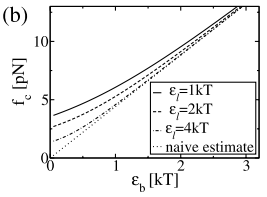

FIG. 5. (a) Model parameters for different DNA sequences and temperatures as obtained by fitting to a detailed thermodynamic model [11,18] (all energies in units of kT). (b) The exact critical force compared to the estimate of Eq. (3).

mentary partition function $Z_u(n, m)$ containing only structures where the last of the n upper bases is *not* bound to the last of the m lower bases:

$$Z_p(n+1, m+1) = qb_{\mathrm{d}} Z_p(n, m) + qb_{\mathrm{d}} g Z_u(n, m),$$

$$Z_u(n+1, m+1) = g \sum_{k=1}^{n} Z_p(k, m+1) + g \sum_{k=1}^{m} Z_p(n+1, k) + g b_{\mathrm{d}} Z_p(n, m) + b_{\mathrm{d}} Z_u(n, m). \quad (7)$$

Here, the Boltzmann factors $q = e^{\varepsilon_{\mathrm{b}}/k_B T}$, $g = e^{-\varepsilon_\ell/2k_B T}$, and $b_{\mathrm{d}} = e^{f\ell_{\mathrm{d}}/k_B T}$ account for base pairing, loop costs, and stretching of double strand, respectively. To obtain the critical behavior for $N \to \infty$, we take the z transform $\hat{Z}(z, y) = \sum_{N,M} Z(N, M) z^N y^M$. The inverse z transform is then determined by the simultaneous poles of $\hat{Z}(z, y)$ in z and y. For large N, the pair of poles with the smallest $|zy|$ dominates. A detailed analysis of the critical behavior will be presented elsewhere [11]; here we are interested in f_c, the force where the dominant pole switches. We find that f_c is the nontrivial root of

$$\left(\frac{b_{\mathrm{s}}^2}{b_{\mathrm{d}}} - q\right)\left(\frac{b_{\mathrm{s}}^2}{b_{\mathrm{d}}} - 1\right) - g^2 q\left(\frac{2}{b_{\mathrm{s}} - 1}\frac{b_{\mathrm{s}}^2}{b_{\mathrm{d}}} + 1\right) = 0. \quad (8)$$

When ε_{b} or $\varepsilon_\ell \gg k_B T$, the second term is negligible and the nontrivial root of (8) is $b_{\mathrm{s}}^2/b_{\mathrm{d}} = q$, recovering the naive estimate (3). However, for smaller ε_{b}, ε_ℓ one finds significant deviations from (3); see Fig. 5.

Loop densities.— Using the same approach as above, we can calculate the loop densities $\rho_<$, $\rho_>$ introduced above. Assuming equilibration between all possible conformations of the two strands with a fixed central base pair, we find $\rho_< = \sum_{a,b} P(a, b) a/\lambda$ and $\rho_> = \sum_{a,b} P(a, b) b/\lambda$, where $\lambda = \min(a, b) + 1$ and $P(a, b) = b_{\mathrm{s}}^{b-a} b_{\mathrm{d}}^{\lambda} q g Z_p(N - b - 1, N - b - 1)/Z_p(N, N)$. The sums can be evaluated exactly for large N [11].

Conclusions.— We find a response of periodic dsDNA to shear forces that is very distinct from that for nonperiodic sequences. Above a thermodynamic critical force f_c, but below a *dynamic* critical force f^*, bulge loop diffusion allows periodic DNA to open by *sliding*. This mechanism leads to a much lower thermodynamic critical force than the *unraveling* mechanism by which nonperiodic DNA opens. Within our model, we have calculated f_c exactly and characterized the associated dynamics, which is effectively *viscoelastic* with a creep compliance $\sim N^{-1}$ for $f_c < f < f^*$. Above f^*, periodic dsDNA also opens predominantly by unraveling (this dynamical transition may be regarded as a remnant of the thermodynamic transition for nonperiodic sequences). Interestingly, periodic DNA could be used as a viscoelastic nanomechanical element with properties that are *programmable* by choosing sequence length and composition. This may lead to applications in microstructured devices, similar to the programmable DNA-based force sensors reported in Ref. [17].

We thank T. Hwa, F. Kühner, and M. Rief for fruitful discussions and the DFG for financial support.

*Electronic address: Richard.Neher@physik.lmu.de
†Electronic address: Ulrich.Gerland@physik.lmu.de

[1] B. Alberts *et al.*, *Molecular Biology of the Cell* (Garland, New York, 2002).
[2] C. Bustamante, J. C. Macosko, and G. J. Wuite, Nat. Rev. Mol. Cell Biol. **1**, 130 (2000); H. Clausen-Schaumann, M. Seitz, R. Krautbauer, and H. E. Gaub, Curr. Opin. Chem. Biol. **4**, 524 (2000); R. Merkel, Phys. Rep. **346**, 343 (2001); R. Lavery, A. Lebrun, J.-F. Allemand, D. Bensimon, and V. Croquette, J. Phys. Condens. Matter **14**, R383 (2002).
[3] C. Danilowicz *et al.*, Proc. Natl. Acad. Sci. U.S.A. **100**, 1694 (2003).
[4] R. Krautbauer, M. Rief, and H. E. Gaub, Nano Lett. **3**, 493 (2003).
[5] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, Phys. Rev. E **58**, 2386 (1998).
[6] T. Strunz, K. Oroszlan, R. Schäfer, and H.-J. Güntherodt, Proc. Natl. Acad. Sci. U.S.A. **96**, 11277 (1999).
[7] D. K. Lubensky and D. R. Nelson, Phys. Rev. Lett. **85**, 1572 (2000); Phys. Rev. E **65**, 031917 (2002).
[8] S. T. Lovett, Mol. Microbiol. **52**, 1243 (2004); L. S. Kappen *et al.*, Biochemistry **42**, 2166 (2003).
[9] D. Pörschke, Biophysical Chemistry **2**, 83 (1974).
[10] T. Hwa, E. Marinari, K. Sneppen, and L. Tang, Proc. Natl. Acad. Sci. U.S.A. **100**, 4411 (2003).
[11] R. Neher and U. Gerland (to be published).
[12] S. B. Smith, Y. Cui, and C. Bustamante, Science **271**, 795 (1996).
[13] Here we assume the B-DNA form. Before melting, DNA with a high cytosine-guanine (CG) content undergoes a transition to a stretched S-DNA form at large forces. This transition was found to be rapid compared to melting [H. Clausen-Schaumann *et al.*, Biophys. J. **78**, 1997 (2000)], so that S-DNA could be treated by different effective parameters.
[14] C. Flamm, W. Fontana, I. Hofacker, and P. Schuster, RNA **6**, 325 (2000).
[15] P.-G. de Gennes, J. Chem. Phys. **55**, 572 (1971).
[16] D. K. Lubensky and D. R. Nelson, Biophys. J. **77**, 1824 (1999).
[17] C. Albrecht *et al.*, Science **301**, 367 (2003).
[18] M. Zuker, Nucleic Acids Res. **31**, 3406 (2003).