Sequence Dependence of DNA Translocation through a Nanopore

Kaifu Luo,[1,*] Tapio Ala-Nissila,[1,2] See-Chen Ying,[2] and Aniket Bhattacharya[3]

[1]*Laboratory of Physics, Helsinki University of Technology, P.O. Box 1100, FIN-02015 TKK, Espoo, Finland*
[2]*Department of Physics, Box 1843, Brown University, Providence, Rhode Island 02912-1843, USA*
[3]*Department of Physics, University of Central Florida, Orlando, Florida 32816-2385, USA*

(Received 12 October 2007; published 5 February 2008)

We investigate the dynamics of DNA translocation through a nanopore using 2D Langevin dynamics simulations, focusing on the dependence of the translocation dynamics on the details of DNA sequences. The DNA molecules studied in this work are built from two types of bases *A* and *C*, which have been shown previously to have different interactions with the pore. We study DNA with repeating blocks A_nC_n for various values of n and find that the translocation time depends strongly on the *block length* $2n$ as well as on the *orientation* of which base enters the pore first. Thus, we demonstrate that the measurement of translocation dynamics of DNA through a nanopore can yield detailed information about its structure. We have also found that the periodicity of the block sequences is contained in the periodicity of the residence time of the individual nucleotides inside the pore.

DOI: 10.1103/PhysRevLett.100.058101 PACS numbers: 87.15.A−, 87.15.H−

The translocation of biopolymers across membranes is ubiquitous in biological systems, such as DNA and RNA translocation across nuclear pores, protein transport through membrane channels, and virus injection into cells. In a seminal experimental paper, Kasianowicz *et al.* [1] demonstrated that an electric field can drive single-stranded DNA and RNA molecules through the water-filled α-hemolysin channel and that the passage of each molecule is signaled by a blockade in the channel current, whose magnitude and duration depend on the structure of the DNA or RNA molecule. Similar experiments have been done recently using solid state nanopores with more precisely controlled dimensions. Triggered by these experiments and potential technological applications [1,2], such as rapid DNA sequencing, gene therapy, and controlled drug delivery, the translocation of biopolymers through a nanopore has become the subject of intensive experimental [3–8] and theoretical [8–21] studies. A particular important question is if DNA translocation through a nanopore can be used to determine the detailed sequence structure of the molecule [2–4,18].

It has been demonstrated in experiments [4,5] that translocation through an α-hemolysin pore can be used to discriminate between polydeoxyadenylic acid [poly(dA)] and polydeoxycytidylic acid [poly(dC)] molecules of the same chain length. The translocation time of poly(dA) is found to be longer, and its distribution is wider with a longer tail compared with the corresponding data for poly(dC). The different behavior was attributed to different interactions of the nucleotides with the pore, with the base *A* having a stronger attractive interaction with the pore than the base *C*. These experimental findings and conclusions were quantitatively supported by recent Langevin dynamics (LD) simulations [19] with a model for the DNA molecules incorporating different base-pore interactions.

Inspired by the ability to discriminate poly(dA) and poly(dC) with the same chain length, Meller *et al.* [4,5] have studied different behavior of the translocation time distribution for the hetero-DNA molecules poly($dA_{50}dC_{50}$) and poly$(dAdC)_{50}$. The result suggests that translocation through a nanopore can distinguish between DNA polynucleotides of similar length with compositions that differ only in detailed sequence structure. In this Letter we seek to shed light on this important question. We adopt a model for the hetero-DNA molecules with different base-pore interactions and investigate the sequence dependence of their translocation dynamics using LD simulations.

In our model, a single-stranded DNA molecule is represented as a bead-spring chain. A short range repulsive Lennard-Jones (LJ) potential $U_{\mathrm{LJ}}(r) = 4\varepsilon[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^{6}] + \varepsilon$ for $r \le 2^{1/6}\sigma$ and 0 for $r > 2^{1/6}\sigma$ exists between all beads leading to excluded volume interaction. Here, σ is the diameter of a bead, and ε is the depth of the potential. Neighboring beads are connected by a finite extension nonlinear elastic (FENE) spring with interaction energy $U_{\mathrm{FENE}}(r) = -\frac{1}{2}kR_0^2 \ln(1 - r^2/R_0^2)$, where r is the distance between consecutive monomers, k is the spring constant, and R_0 is the maximum allowed separation between connected monomers. We consider a 2D geometry as shown in Fig. 1, where the wall of thickness L is formed by columns of stationary particles. A pore of length L and width W in the center of the wall connects the *cis* and the *trans* sides, and a voltage is applied across the pore to drive the negatively charged DNA through the pore. Between all bead-wall particle pairs, there exists the same short range repulsive LJ interaction as described above. The pore-bead interaction is modeled by a LJ potential with a cutoff of 2.5σ and interaction strength ε_{pA} for base *A* and ε_{pC} for base *C*. Each bead is subjected to conservative, frictional, and random forces, respectively, leading to the equation of motion $m\ddot{\mathbf{r}}_i = -\nabla(U_{\mathrm{LJ}} + U_{\mathrm{FENE}}) + \mathbf{F}_{\mathrm{ext}} - \xi\mathbf{v}_i + \mathbf{F}_i^R$, where m is the bead's mass, ξ is the friction coefficient, $\mathbf{v}_i$ is the bead's velocity, and $\mathbf{F}_i^R$ is the random force which satisfies the fluctuation-dissipation theorem [22]. The ex-

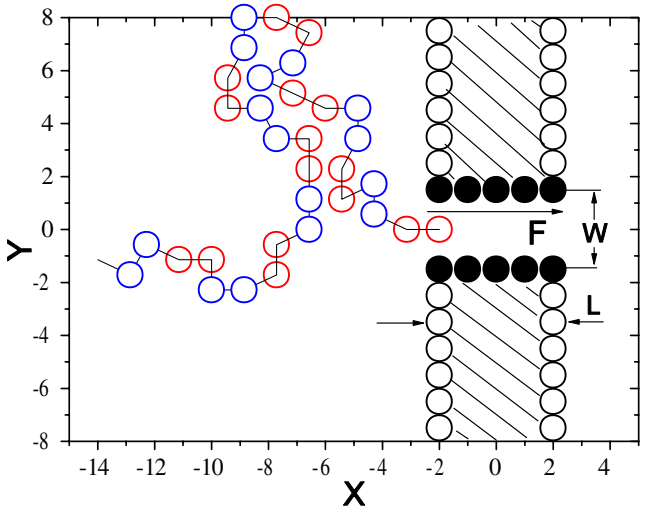

FIG. 1 (color online). A schematic representation of the system. The pore length $L = 5$ and the pore width $W = 3$.

ternal force due to the applied voltage is represented by $\mathbf{F}_{\text{ext}} = F\hat{\mathbf{x}}$.

The LJ parameters ε and σ and the bead mass fix the system energy, length, and mass units, respectively, leading to the corresponding time scale $t_{\text{LJ}} = (m\sigma^2/\varepsilon)^{1/2}$ and force scale ε/σ. In our model, each bead corresponds to a Kuhn length of a single-stranded DNA containing approximately three nucleotide bases, so the value of $\sigma \sim 1.5$ nm [23]. The average mass of a base in DNA is about 312 amu, so the bead mass $m \approx 936$ amu. We set $k_BT = 1.2\varepsilon$, which means the interaction strength ε is 3.39×10^{-21} J at actual temperature 295 K. This leads to a time scale of 32.1 ps and a force scale of 2.3 pN. The dimensionless parameters in the model are then chosen to be $R_0 = 2$, $k = 7$, $\xi = 0.7$, $L = 5$, $W = 3$, and $F = 0.5$. Each base (nucleotide) is estimated to have an effective charge of $0.094e$ from Ref. [7], leading to an effective charge of a bead being $0.282e$. Thus, $F = 0.5$ corresponds to a voltage of about 187.9 mV across the pore within the range of experimental parameters [1,2,4–6]. The choice of $W = 3$ ensures that the average interaction of both bases A and C with the pore are attractive. The pore-base interactions $\varepsilon_{pA} = 3.0$ and $\varepsilon_{pC} = 1.0$ are chosen based on comparison of the theoretical results [19] with the experimental data [4] for the translocation time distribution histogram of the homo-DNA molecules poly(dC)$_{100}$ and poly(dA)$_{100}$. The Langevin equation is integrated in time by a method described by Ermak and Buckholz [24] in 2D. Initially, the first monomer of the chain is placed in the entrance of the pore, while the remaining monomers are under thermal collisions described by the Langevin thermostat to obtain an equilibrium configuration.

We consider the sequence dependent translocation results for DNA of chain length $N = 128$ with the symmetric blocks A_nC_n having block length $M = 2n$, with minimum value of $n = 1$ for poly(dAdC)$_{64}$ and maximum value of $n = N/2$ for poly(dA$_{64}$dC$_{64}$). Figure 2 shows the translocation time τ as a function of the block length. The translocation time is obtained as the time interval between the entrance of the first bead into the pore and the exit of the last bead [25]. Typically, we average our data over 2000 independent runs. The horizontal dashed, dotted, and dash-dotted lines correspond to τ_A, $(\tau_A + \tau_C)/2$, and τ_C, respectively. Here, τ_A and τ_C are the translocation times for poly(dA)$_{128}$ and poly(dC)$_{128}$, respectively. For $M < 8$, τ is close to τ_C and much lower than $(\tau_A + \tau_C)/2$, with very weak dependence on the block length and the orientation of which the monomer enters the pore first.

However, τ begins to increase rapidly with M for $M \geq 8$, approaches a maximum between $M = 16$ and 32, and finally decreases slowly with increasing M. In addition, τ also depends strongly on the polymer orientation. It is always much longer for the base A entering the pore first than the other orientation. Quantitatively, we define r as the ratio of translocation times for base C entering the pore first and base A entering the pore first. The inset of Fig. 2 shows r as a function of the block length. For $M \leq 4$, $r = 1$, but for longer M, r increases exponentially with M. For poly(dA$_{64}$dC$_{64}$), $r \approx 5$ [26].

Qualitatively, the large block length results can be understood by examining the extreme case of the largest block $M = 128$. When the C_{64} block is translocated first through the pore, subsequent forward motion is energetically unfavored because of the strong attraction between the pore and base A. As a result, frequent backward transitions occur which slows down the overall translocation process. In the opposite orientation when the A_{64} block goes through the pore first, the difference of probabilities

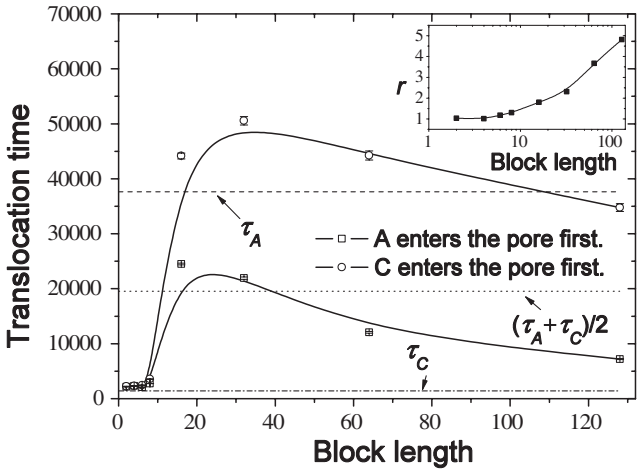

FIG. 2. Translocation time as a function of the block length for multiblock DNA with symmetric repeat units A_nC_n. Here the block length $M = 2n$, $\varepsilon_{pA} = 3.0$, $\varepsilon_{pC} = 1.0$, $F = 0.5$, and the chain length $N = 128$. The insert shows r as a function of the block length for multiblock DNA with symmetric repeat units A_nC_n. Here, we define r as the ratio of translocation times for the base C entering the pore first and the base A entering the pore first.

between the forward and the backward steps of the remaining motion is smaller, leading to a value $r > 1$. Similar behaviors have been recently analyzed by Kotsev and Kolomeisky [20] where the translocation of polymers consisting of a double-stranded block and a single-stranded block is considered, and by Tsuchiya and Matsuyama [21] where they studied the translocation of an amphiphilic polymer.

For base C entering the pore first, $\tau \gg (\tau_A + \tau_C)/2$ for $M \geq 16$. It is a surprise that $\tau > \tau_A$ for $16 \leq M \leq 64$. For poly(dA), the frequency of backward and forward motion is much slower than that for poly(dC). Incorporating base C with a suitable block length into poly(dA) will increase the frequency of backward and forward motion when the C block is in the pore. As a result, the translocation time is larger than τ_A. For base A entering the pore first, $\tau > \tau_C$ and for $16 \leq M \leq 32$, $\tau > (\tau_A + \tau_C)/2$.

Meller *et al.* [4] have studied blockade signals for the hetero-DNAs poly($dA_{50}dC_{50}$) and poly$(dAdC)_{50}$. The translocation events are organized into two well-localized groups with different blockage currents, the origin of which is yet uncertain [27]. Direct comparison with the present study is further complicated by the fact that poly(dA) molecules have a higher tendency to form single-stranded base-stacked helices as compared with poly(dC) [2,5], although the base-pore interaction effect is still expected to be dominant. The group 2 data at low temperatures show that the translocation time for poly($dA_{50}dC_{50}$) is longer than that for poly$(dAdC)_{50}$, in agreement with our finding here that larger block length of repeat unit leads to a greater translocation time, as shown in Fig. 2. It would be desirable to have future experimental tests for the orientation dependence of the translocation for these heteropolymers as well. We have also studied the histograms of translocation time for the hetero-DNAs poly(dA_ndC_n). For short block length $M = 2n$, the histograms depend only weakly on the orientation, shown in Fig. 3(a), and the behavior is close to that of poly(dC). However, for longer block lengths, the histogram deviates markedly from a Gaussian with a long exponential tail as shown in Fig. 3(b). This behavior is in agreement with the experimental observation [4]. There is also a strong orientational dependence, with the histogram for C entering first shifted to longer translocation times.

We have also investigated the distribution for waiting (residence) time of base s defined as the time between the events that the base s and the base $s + 1$ exit the pore. We find that the residence times for the ordered DNA with repeat units A_nC_n (for $n > 2$) exhibit "fringes" reminiscent of an optical interference pattern, as shown in Fig. 4. The number of peaks is exactly equal to $N/2n$. The periodicity of the waiting time not only depends on the block length but also on the orientational property of the chain as well. For the sequence A_nC_n with A entering the pore first ($\varepsilon_A > \varepsilon_C$), the residence time is symmetric with respect to the center of the chain containing exactly $N/4n$ maxima on either side, whereas with C entering the pore first it has

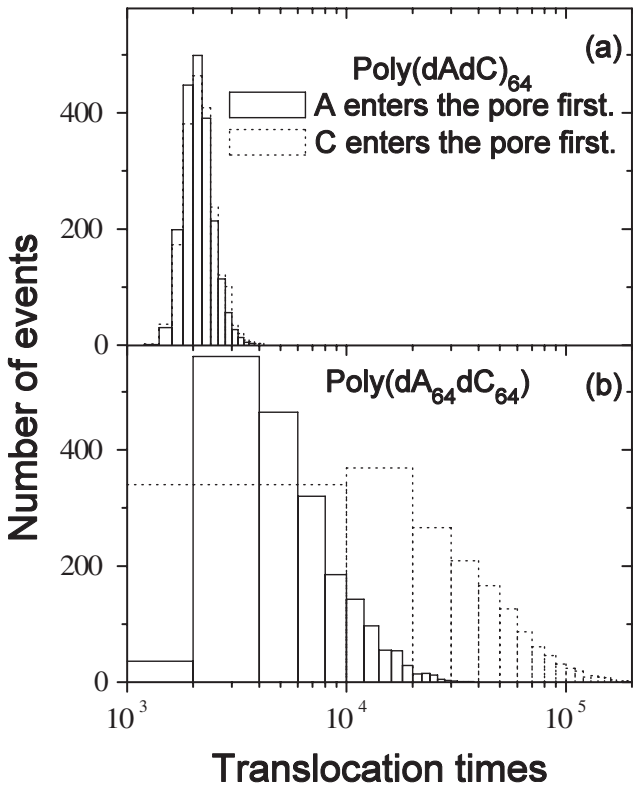

FIG. 3. Histogram of the translocation times for (a) poly$(dAdC)_{64}$ and (b) poly($dA_{64}dC_{64}$) under $F = 0.5$.

$(N/2n - 1)$ maxima with the last maximum ending with the largest waiting time.

The sequence dependence of the waiting time distribution yields a better understanding for the sequence dependence of τ shown in Fig. 2. The translocation time can be written as $\tau \sim \tau_1 + \tau_2 + \tau_3$, where τ_1, τ_2, and τ_3 correspond to the initial filling of the pore, transfer of the base from the *cis* side to the *trans* side, and finally the emptying of the pore, respectively. For the present case, $\tau_1 \ll \tau_2, \tau_3$. For $M \leq 8$, τ_2 dominates, and it has no strong dependence on the detailed sequence or the orientation of the chain. When the base C enters the pore first, τ_3 increases rapidly with increasing M for $8 \leq M \leq 16$ and then saturates to a constant value for $32 \leq M \leq 128$. On the other hand, τ_2 is related to both the number of the fringes and the corresponding maximum time. With increasing M, the former decreases and the latter increases. The interplay of all these factors leads to a maximum for τ as a function of the block length M. Similar consideration applies for base A entering the pore first except that here τ_2 dominates over τ_3.

To summarize, we have demonstrated that sequences of a driven DNA can be identified from its translocation specific characteristics driven through a nanopore that has different affinity for each base. Simulation studies based on this *attractive nanopore* model are in accord with the existing experimental data. A stronger attraction for the polynucleotide A inside the nanopore leads to a much longer translocation time for $(polydA)_{100}$ as com-

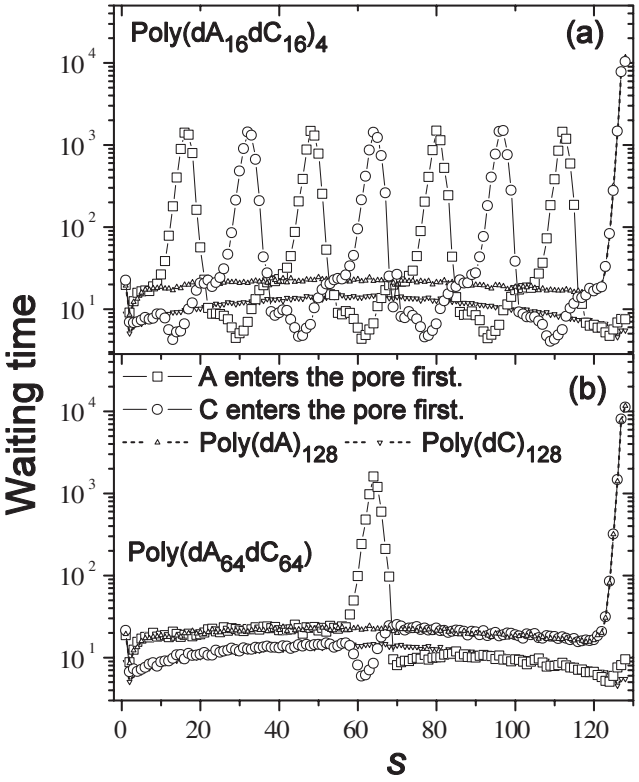

FIG. 4. Waiting times for (a) poly($dA_{16}dC_{16}$)$_4$ and (b) poly($dA_{64}dC_{64}$) under $F = 0.5$.

pared to (polydC)$_{100}$. Further analysis explains the shape of the histogram of the first passage time, provides an understanding of how translocation time depends on a specific sequence, and explains the experimental data of longer translocation time for (polydA)$_{50}$(polydC)$_{50}$ compared to (polydAdC)$_{50}$. Our simulation studies also reveal a novel phenomenon that the information for the periodicity of the block sequences is contained in the periodicity of the residence time of the individual nucleotides.

This work has been supported in part by The Academy of Finland through its Center of Excellence (COMP) and TransPoly Consortium grants.

*To whom all correspondence should be addressed.
luokaifu@yahoo.com

[1] J. J. Kasianowicz, E. Brandin, D. Branton, and D. W. Deamer, Proc. Natl. Acad. Sci. U.S.A. **93**, 13 770 (1996).
[2] A. Meller, J. Phys. Condens. Matter **15**, R581 (2003).
[3] M. Akeson, D. Branton, J. J. Kasianowicz, E. Brandin, and D. W. Deamer, Biophys. J. **77**, 3227 (1999).
[4] A. Meller, L. Nivon, E. Brandin, J. A. Golovchenko, and D. Branton, Proc. Natl. Acad. Sci. U.S.A. **97**, 1079 (2000).
[5] A. Meller and D. Branton, Electrophoresis **23**, 2583 (2002).
[6] A. Meller, L. Nivon, and D. Branton, Phys. Rev. Lett. **86**, 3435 (2001).
[7] A. F. Sauer-Budge, J. A. Nyamwanda, D. K. Lubensky, and D. Branton, Phys. Rev. Lett. **90**, 238101 (2003); J. Mathe, H. Visram, V. Viasnoff, Y. Rabin, and A. Meller, Biophys. J. **87**, 3205 (2004).
[8] A. J. Storm, C. Storm, J. Chen, H. Zandbergen, J.-F. Joanny, and C. Dekker, Nano Lett. **5**, 1193 (2005).
[9] W. Sung and P. J. Park, Phys. Rev. Lett. **77**, 783 (1996).
[10] M. Muthukumar, J. Chem. Phys. **111**, 10 371 (1999).
[11] D. K. Lubensky and D. R. Nelson, Biophys. J. **77**, 1824 (1999).
[12] Y. Kafri, D. K. Lubensky, and D. R. Nelson, Biophys. J. **86**, 3373 (2004).
[13] R. Metzler and J. Klafter, Biophys. J. **85**, 2776 (2003); T. Ambjornsson, M. A. Lomholt, and R. Metzler J. Phys. Condens. Matter **17**, S3945 (2005).
[14] J. Chuang, Y. Kantor, and M. Kardar, Phys. Rev. E **65**, 011802 (2002); Y. Kantor and M. Kardar, Phys. Rev. E **69**, 021806 (2004).
[15] A. Milchev, K. Binder, and A. Bhattacharya, J. Chem. Phys. **121**, 6042 (2004).
[16] K. F. Luo, T. Ala-Nissila, and S. C. Ying, J. Chem. Phys. **124**, 034714 (2006); K. F. Luo, I. Huopaniemi, T. Ala-Nissila, and S. C. Ying, J. Chem. Phys. **124**, 114704 (2006).
[17] I. Huopaniemi, K. F. Luo, T. Ala-Nissila, and S. C. Ying, J. Chem. Phys. **125**, 124901 (2006); Phys. Rev. E **75**, 061912 (2007).
[18] K. F. Luo, T. Ala-Nissila, S. C. Ying, and A. Bhattacharya, J. Chem. Phys. **126**, 145101 (2007).
[19] K. F. Luo, T. Ala-Nissila, S. C. Ying, and A. Bhattacharya, Phys. Rev. Lett. **99**, 148102 (2007).
[20] S. Kotsev and A. B. Kolomeisky, J. Chem. Phys. **125**, 084906 (2006).
[21] S. Tsuchiya and A. Matsuyama, Phys. Rev. E **76**, 011801 (2007).
[22] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1987).
[23] S. B. Smith, Y. Cui, and C. Bustamante, Science **271**, 795 (1996); M. C. Murphy, I. Rasnik, W. Cheng, T. M. Lohman, and T. Ha, Biophys. J. **86**, 2530 (2004).
[24] D. L. Ermak and H. Buckholz J. Comput. Phys. **35**, 169 (1980).
[25] It is worthwhile to note that the translocation time depends strongly on the pore length L. We have checked that $L \sim 10$ nm produces average translocation times $\tau \sim 100$ μs in accord with Meller's experiments [4] for $L \sim 10$ nm. For computational efficiency we present results for $L = 5$ here.
[26] We checked r as a function of ε_{pA} for $F = 0.5$ and $\varepsilon_{pC} = 1$ and found that the dependence of τ on chain orientation starts to vanish below $\varepsilon_{pA}/\varepsilon_{pC} \approx 2$.
[27] We interpret that the 2nd peak in Meller *et al.*'s experiment [4] corresponds to pure translocation while peak 1 may have an unwanted component of DNA residing at the pore to block the channel current and finally exiting from the same side it came from.