Heteropolymer translocation through nanopores

Kaifu Luoa
Laboratory of Physics, Helsinki University of Technology, P.O. Box 1100, FIN-02015 TKK Espoo, Finland

Tapio Ala-Nissila
Laboratory of Physics, Helsinki University of Technology, P.O. Box 1100, FIN-02015 TKK Espoo, Finland and Department of Physics, Brown University, P.O. Box 1843, Providence, Rhode Island 02912-1843

See-Chen Ying
Department of Physics, Brown University, P.O. Box 1843, Providence, Rhode Island 02912-1843

Aniket Bhattacharya
Department of Physics, University of Central Florida, Orlando, Florida 32816-2385

(Received 4 January 2007; accepted 1 March 2007; published online 10 April 2007)

The authors investigate the translocation dynamics of heteropolymers driven through a nanopore using a constant temperature Langevin thermostat. Specifically, they consider heteropolymers consisting of two types of monomers labeled A and B, which are distinguished by the magnitude of the driving force that they experience inside the pore. From a series of studies on polymers with sequences $A_mB_n$, the authors identify both universal as well as specific sequence properties of the translocating chains. They find that the scaling of the average translocation time as a function of the chain length $N$ remains unaffected by the heterogeneity, while the residence time of each bead is a strong function of the sequence for short repeat units. They further discover that for a symmetric heteropolymer $A_mB_n$ of fixed length, the pattern exhibited by the residence times of the individual monomers has striking similarity with a double slit interference pattern where the total number of repeat units $N/2n$ controls the number of interference fringes. These results are relevant for designing nanopore based sequencing techniques. © 2007 American Institute of Physics.

I. INTRODUCTION

The transport of a polymer through a nanopore plays a critical role in numerous biological processes, such as DNA and RNA translocation across nuclear pores, protein transport through membrane channels, and virus injection. Due to various potential technological applications, such as rapid DNA sequencing, gene therapy, and controlled drug delivery, polymer translocation has become a subject of intensive experimental, theoretical, and numerical studies.

For a polymer threading through a nanopore, loss of available configurations due to the geometric constriction leads to an effective entropic barrier. Kasianowicz et al. 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. These observations can directly be used to characterize the polymer length, and with further improvements the technique could be used to read off the nucleotide sequence of DNA or RNA. Other recent experiments also show that an orientation asymmetry plays an important role in the translocation due to the complex interaction of DNA nucleotides with the protein nanopore. In addition to the α-hemolysin channel, experiments have been done with solid state nanopores, whose advantages are tunable pore size and stability over changes in external conditions including voltage, temperature, salinity, and pH.

Theoretically, translocation dynamics was initially examined by considering equilibrium entropy of the polymer as a function of the position of the polymer through the nanopore. This has been indicated to be inappropriate for the study of translocation dynamics. Most recently, we have investigated both free and forced translocation using both the two-dimensional fluctuating bond model with single-segment Monte Carlo moves and Langevin dynamic simulations. For the free translocation, we numerically verified that the translocation time $\tau \sim N^{1+2\nu}$, where $N$ is the chain length and $\nu$ is the Flory exponent. For forced translocation, we found a crossover scaling from $\tau \sim N^{2\nu}$ for relatively short polymers to $\tau \sim N^{1+\nu}$ for longer chains.

However, most of the theoretical treatments of translocation have assumed a homogeneous polymer, although several experiments show that in the real biological systems inhomogeneities in the structure and interactions between polymer and other molecules might have a significant effect on the overall dynamics. Based on the equilibrium entropy of the polymer, Muthukumar theoretically discussed some effects of sequence on translocation by considering diblock copolymers. It was found that for the weak driving forces...
the translocation time depends on which polymer end begins the translocation process. In addition, Kafri et al.\(^2\) also studied the effect of sequence heterogeneity in polymer translocation and found that the heterogeneity might lead to anomalous dynamics at some conditions.

The purpose of this paper is to investigate systematically how the pore can sense the effect of heterogeneity of the chain manifested through quantities related to its translocation dynamics. To this end, we use Langevin dynamics simulations in two dimensions (2D). We consider a coarse-grained model heteropolymers composed of two kinds of monomers labeled A and B. Under an applied external field, the driving forces for A and B are different. With this in mind we launch a series of investigation on chains with repeat unit \(A_nB_n\) and study various properties as a function of such units. In Sec. II, we briefly describe our model and the simulation technique. In Sec. III, we present our results. Finally, the discussion and conclusions are in Sec. IV.

### II. MODEL AND METHOD

In the simulations, the polymer chains are modeled as bead-spring chains of Lennard-Jones (LJ) particles with the finite extension nonlinear elastic (FENE) potential. Both excluded volume and van der Waals interactions between beads are modeled by a repulsive LJ potential between all bead pairs,

\[
U_{LJ}(r) = \begin{cases} 
4\varepsilon[(\sigma/r)^{12} - (\sigma/r)^{6}] + \epsilon, & r \leq 2^{1/6}\sigma \\
0, & r > 2^{1/6}\sigma,
\end{cases}
\]

where \(\sigma\) is the diameter of a bead and \(\epsilon\) is the parameter adjusting the depth of the potential.

The connectivity between the beads is modeled as a FENE spring,

\[
U_{FENE}(r) = -\frac{1}{2}kR_0^2 \ln \left(1 - \frac{r^2}{R_0^2}\right),
\]

where \(r\) is the separation between consecutive beads, \(k\) is the spring constant, and \(R_0\) is the maximum allowed separation between connected beads.

In the Langevin dynamics method, each bead obeys the following equation of motion: \(^38\)

\[
m\ddot{\mathbf{r}}_i = -\nabla U_{LJ} + U_{FENE} + \mathbf{F}_{\text{ext}} - \xi \dot{\mathbf{r}}_i + f_i(t),
\]

where \(m\) is the monomer’s mass, \(\xi\) is the friction coefficient, and \(\mathbf{v}_i\) is the monomer’s velocity. \(\mathbf{F}_{\text{ext}}\) denotes the external force due to the applied external field. \(f_i(t)\) is the random force and satisfies the fluctuation dissipation theorem,

\[
\langle f_i(t) \rangle = 0,
\]

\[
\langle f_i(t) \cdot f_j(t') \rangle = 6k_BT\xi\delta_{ij}\delta(t-t').
\]

The wall is described as \(L\) columns of stationary particles within distance \(\delta\) from one another, which interact with the beads by the repulsive Lennard-Jones potential. The wall particle positions are not changed in the simulations. The pore is introduced in the wall by simply removing \(W\) beads from each column. The external field exists only inside the pore corresponding to an external voltage applied across both ends of the pore.

In our coarse-grained model, a polymer is composed of two kinds of monomers A and B, which may have different charges. Under the applied electric field, the driving force is proportional to the net charge on the monomer. Within the pore the external forces take the value \(\mathbf{F}_{\text{ext}}=(F+\Delta)i\) for monomer A and \(\mathbf{F}_{\text{ext}}=(F-\Delta)i\) for monomer B, with \(i\) being the unit vector in the direction along the pore. For a homopolymer, \(\Delta=0\). This model is not intended for a quantitative study of any real biopolymer such as DNA, where no charge difference between the monomers exists, but rather for a study of the generic features induced by heterogeneity, such as scaling behavior of various physical quantities.

In our simulations, the LJ parameters \(\varepsilon\) and \(\sigma\) fix the system energy and length units, respectively. The time scale is given by \(t_{LJ}=(m\sigma^2/\varepsilon)^{1/2}\). The parameters are \(\sigma=1\), \(R_0=2\sigma\), \(k=7\varepsilon\), \(k_BT=1.2\varepsilon\), and reduced friction \(\xi=0.7\). In the simulations, \(L=1\) and \(W=2\) unless otherwise stated. The Langevin equation is integrated in time by a method described by Ermak and Buckholz\(^39\) in 2D. To create the initial configuration, the first monomer of the chain is placed in the entrance of the pore. The polymer is then allowed to relax to obtain an equilibrium configuration such that the first monomer is fixed at the entrance and the other monomers are under thermal collisions described by the Langevin thermostat. In all of our simulations we did a number of runs with uncorrelated initial states and random numbers describing the random collisions. The translocation time is defined as the duration of time it takes for the chain to move through the pore in the direction of the driving force.

As to heteropolymer translocation, the reduced driving force is set as \(F=5\varepsilon/\sigma\) and \(\Delta=\pm 2.5\varepsilon/\sigma\) unless otherwise stated. The corresponding voltages of these parameters are in the range of real values used in the experiments.\(^7\)

### III. RESULTS AND DISCUSSION

Since our primary objective is to study the effect of heterogeneous sequences, it will be useful to present the simulation results into four main categories. As mentioned earlier, in our model the specific details of the beads are distinguished only when they are inside the pore where each type of bead experiences force \(F_A=(F+\Delta)\) and \(F_B=(F-\Delta)\). This gross simplification is deliberately done to identify specific pore effects on the chain which is assumed to be the most dominant interaction between the two dissimilar species of monomers (notice that we do not differentiate the two species outside the pore).

#### A. Effect of sequence on heteropolymer translocation with a fixed \(N\)

As mentioned in the Introduction, one of the main objectives of studying the translocation dynamics of a single molecule through a nanopore is to develop methods where the pore can distinguish polynucleotides not only of different chain lengths but also of a fixed length with different sequences. With that in mind we present in the following sec-
tion some of the specific sequence results. Without any loss of generality here we present results for a chain of length \( N = 128 \), but most of our conclusions are valid for longer chains as well.

In Table I, we show the average translocation times for multiblock copolymers with repeat units \( A_m B_n \), where \( m \) and \( n \leq 3 \). Increasing row index corresponds to increasing relative volume fraction for species \( A \). Likewise, for a given row, an increasing column index raises the volume fraction of species \( B \). Obviously, the indices \( n = 0 \) and \( m = 0 \) represent homopolymers \( A \) and \( B \), respectively. We define \( m+n \) as the block length, and the volume fraction of the \( B \) component in the polymer \( f_B = n/(m+n) \). As intuitively expected, \( \tau \) increases with increasing \( f_B \) because the driving force for the \( B \) segments is smaller than that for the \( A \) species. A three-dimensional surface plot of the average translocation time is also shown in Fig. 1.

As an example, let us look at the histograms of translocation times for multiblock copolymers with repeat units \( A A A B, A A B, \) and \( A B \) more closely. Here, the values for \( f_B \) are 1/4, 1/3, and 1/2, respectively. As shown in Figs. 2(a)–2(c), the histograms for different polymers all obey Gaussian distribution. With increasing \( f_B \), the distribution of translocation times is wider. If we assume that the relative increase in \( \tau \) is simply proportional to the increase in \( f_B \), \( \tau \) has an exponential dependence on the volume fraction [Fig. 2(d)]. From our data, we indeed find that both the translocation time and the standard deviation can be fitted by \( \tau = \tau_A e^{0.96 \pm 0.03} f_B \) and \( \sigma \tau = \sigma_A e^{0.18 \pm 0.06} f_B \), respectively, where \( \tau_A \) and \( \sigma_A \) denote the corresponding translocation time and its standard deviation for homopolymer \( A \).

Interestingly, we also notice that for many short repeat units \( A_m B_n \), where \( m \) and \( n \leq 8 \), the sequence dependence of the translocation time exhibits a very rich pattern and, in principle, the result can be used to predict other unknown sequences, as shown in Fig. 3, where the translocation times are plotted as a function of \( f_B \). Here the solid line represents \( \tau = \tau_A (1-f_B) \) and the dashed line corresponds to \( \tau = \tau_B e^{f_B} \). A careful look at the location of the translocation time reveals intriguing features. First, one notices that the translocation times for all sequences with either \( n = 1 \) and/or \( m = 1 \) are distributed along the curve \( \tau = \tau^*_A e^{f_B} \). This curve ends at the pure species \( A \) and \( B \). For all other sequences \( A_m B_n \), where \( m \) and \( n \) are not equal to unity, the translocation times lie close to the straight line, \( \tau = \tau_A (1-f_B) + f_B \tau_B \). Therefore, this type of a plot could be useful for reading an unknown sequence by looking at its location on the plot.

### B. Universal features of heteropolymer translocation

In addition to exhibiting sequence dependent rich structures, the translocation of heteropolymers also exhibits certain universal aspects similar to those for the translocation of homopolymers as found in our previous studies\(^ {33,44} \) and by other groups.\(^ {21} \) Previously, we observed that the translocation time and velocity of the center of mass scale as \( \tau \sim N^{2 \nu} \) and \( v \sim N^{-\nu} \) for relatively short chains and for longer chains \( (N \gg 200) \), \( \tau \sim N^{1+\nu} \) and \( v \sim N^{-1} \), where \( \nu = 0.75 \) is the Flory exponent in 2D. We find that these scaling properties remain valid for heteropolymers also with arbitrary repeat units \( A_m B_n \), as shown in Fig. 4. This could be easily understood by noting that at a higher level of coarse graining, the microstructure of the chain is irrelevant as far as universal scaling properties are concerned. We have checked this behavior by simulating chains of different lengths and repeat units \( A_m B_n \).

### C. Sequence dependent features of the symmetric blocks

Having shown the sequence dependence characteristics of the heteropolymers as a function of \( f_B \), we now further analyze in detail the sequence dependent results specifically for the symmetric blocks \( A_m B_n \), i.e., for \( f_B = 0.5 \). Here, the block length \( M = 2n \), where the minimum value of \( n = 1 \) for the repeat unit \( A B \) and its maximum value is equal to \( N/2 \) for a chain length \( N \). Figure 5 shows the translocation time as a function of the block length. The horizontal dotted line corresponds to \( (\tau_A + \tau_B)/2 \), where \( \tau_A \) and \( \tau_B \) are translocation times for homopolymers \( A \) and \( B \), respectively. We consider two different cases where monomer \( A \) or monomer \( B \) enters the pore first. When \( A \) enters the pore first, for \( M \leq 4 \), \( \tau \) is lower than the value \( (\tau_A + \tau_B)/2 \), then overshoots it with a maximum at \( M \approx 16 \), and finally approaches this value asymptotically for long blocks. When monomer \( B \) enters the pore first, the qualitative behavior is very similar.
It is noteworthy in Fig. 5 that for $M/H_{11021}^{16}/H_{20849}^{16}/H_{20850}^{16}$ the translocation time for both cases is a strong function of the block length. With increasing block length $M$ the dependence is small. This could be attributed to a decoherence effect or loss of memory for the large $A$ or $B$ segments as they go through the pore. The persistence length of the chain in the vicinity of the pore is expected to be different. Let us suppose that at a certain time the $i^{th}$ chain is inside the pore. Evidently, one expects that the $(i-p)^{th}$ monomer will feel this effect beyond which the chain will recover its bulk characteristics. Likewise, the $(i+q)^{th}$ monomer ahead of this $i^{th}$ chain will also be correlated. As a result, up to the repeat unit length $(p+q)$ the dependence will be strong. If one calculates the correlation function as a function of distance from the monomer located at the center of the pore.

D. Waiting time distributions for multiblock copolymers

Since the detection of the sequence from the translocation dynamics is among the central issues in this paper, we have further investigated the increase of $\tau$ with $f_B$ in terms of the waiting (residence) time distribution as we studied previously for coarse-grained homopolymers.\textsuperscript{43,44} We define the waiting or residence time of monomer $s$ as the average time between the events that monomer $s$ and monomer $s+1$ exit the pore. The nonequilibrium nature of translocation has a considerable influence on this variable. We have numerically calculated the waiting times for each monomer passing through the pore.

Evidently, one expects that the $(i-p)^{th}$ monomer will feel this effect beyond which the chain will recover its bulk characteristics. Likewise, the $(i+q)^{th}$ monomer ahead of this $i^{th}$ chain will also be correlated. As a result, up to the repeat unit length $(p+q)$ the dependence will be strong. If one calculates the correlation function as a function of distance $\delta$ from the monomer $i$ located at the center of the pore.

FIG. 2. [(a)–(c)] Histogram of translocation times for multiblock copolymers $A_1B_1, A_2B_1$, and $A_3B_1$, which correspond to $f_B=0.5, 1/3$, and $1/4$, respectively. (d) Standard deviation and translocation time as functions of the volume fraction of monomer $B$. The driving force $F=5.0, \Delta=2.5$, and the chain length $N=128$.

FIG. 3. Translocation time for heteropolymers with repeat units $A_mB_n$ for different values of $m$ and $n$ as a function of the volume fraction of the monomer $B$. The driving force $F=5.0, \Delta=2.5$, and the chain length $N=128$. 

D. Waiting time distributions for multiblock copolymers

Since the detection of the sequence from the translocation dynamics is among the central issues in this paper, we have further investigated the increase of $\tau$ with $f_B$ in terms of the waiting (residence) time distribution as we studied previously for coarse-grained homopolymers.\textsuperscript{43,44} We define the waiting or residence time of monomer $s$ as the average time between the events that monomer $s$ and monomer $s+1$ exit the pore. The nonequilibrium nature of translocation has a considerable influence on this variable. We have numerically calculated the waiting times for each monomer passing through the pore.

Figure 6 shows the waiting times for multiblock copolymers with different block sizes. As a reference, we also show waiting times for homopolymers $A$ and $B$. As noted by our previous work,\textsuperscript{43,44} the waiting time depends strongly on the monomer position in the chain. For a homopolymer of length $N=128$ the waiting time distribution has a maximum at a
value $s$ close to the middle monomer in the chain (Fig. 6). We note that for a longer chain, this maximum shifts towards the end of the chain.43,44

For the symmetric heteropolymer $AB$ we notice that the residence time portrays characters of the individual monomers across the entire chain, as shown in Fig. 6. An interesting result is that the residence times for the ordered heteropolymers $A_mB_n$ (for $n > 2$) exhibit “fringes” similar to double slit optical grating. To draw the analogy, we note that in a double slit experiment the diffraction envelope is limited by the width of the individual slit. Likewise, here we note that it is the residence times of the individual homopolymers $A$ and $B$ that serve as the envelope (maximum and minimum). The block length brings in additional features that oscillate between these two bounds. It is worth noting that in a double slit experiment the ratio $d/b$, where $d$ is the distance between the slits and $b$ is the width of the individual slit, dictates the number of fringes. Here, this ratio $d/b$ translates to $N/2n$. We also notice that the waiting time for the homopolymer $A$ or $B$ (around each maximum or minimum) develops additional fine structures. Evidently, these additional fine structures of residence times of the individual monomers as a function of their relative position in each block are a natural consequence of the presence of a dissimilar block on either side of a given block. One immediately notices that the value of the repeat unit $n$ can easily be read from the number of points between two minima, enabling us to calculate the length of each block. The number of peaks is exactly equal to $N/(2n)$ [128/8=16 in Fig. 6(c), 128/16=8 in Fig. 6(d), and 128/128=1 in Figs. 6(e) and 6(f)]. Thus, if the residence time can be measured experimentally, information about the block length is immediately accessible. The dependence of the translocation time on the block length, as shown in Fig. 5, can be understood according to the waiting time distribution. As shown in Fig. 6(a), the waiting times of the monomer $B$ in heteropolymer $AB$ are much shorter than that of its homopolymer, while for the monomer $A$ they are slightly longer. For the $B$ monomer the last and next monomers are $A$, which leads to less backward events, resulting in faster translocation. As a result, $\tau$ is less than $(\tau_A + \tau_B)/2$. In Fig. 6(b) and 6(c), for $A_2B_2$ and $A_3B_4$ heteropolymers, all the monomers in the basic block show different waiting time behaviors. The first $A$ and the first $B$ monomers in the repeat unit show slightly longer and shorter waiting times compared with their homopolymers, respectively. However, for the other $A$ monomers, their waiting times are shorter than that of homopolymer $A$, while the opposite is true for the other $B$ monomers. The latter dominates the final outcome for translocation time, which leads to a rapid increase of translocation time with block size.

IV. SUMMARY AND CONCLUSIONS

In this work, we have investigated the dynamics of heteropolymer translocation through a nanopore driven by an external force using 2D Langevin dynamics simulations with an aim to characterize both specific and universal sequence aspects of translocation. We find that scaling exponents of the chain length dependence of the translocation time and the velocity of the center of mass are the same as that of a homopolymer. This can be easily reconciled with further coarse graining of the chain in terms of the individual blocks of length $m+n$ for $A_mB_n$. The translocation times plotted as a function of $m$ and $n$ reveal novel features thus far not reported. The plots show that the translocation time $\tau = \tau(m,n)$ is unique for small values of the block $m+n$. We
also notice that the average translocation time $\tau$ falls under different curves dependent on the specific patterns of the sequence. Moreover, a detailed analysis of the symmetric multiblock copolymers reveals unique and intriguing features. We find that the residence times of the individual monomers act as a fingerprint of the sequence. In particular, there is a striking similarity between the patterns observed for the distribution of the residence time for heteropolymers $A_nB_m$ of length $N$ and the interference pattern obtained from a double slit with $N/(2n)$ interference fringes. This mapping can possibly be extended from symmetric blocks to arbitrary repeat units $A_nB_m$, which then may lead to a better understanding of interpreting nanopore based sequence detection.

ACKNOWLEDGMENTS

This work has been supported in part by the COMP Center of Excellence grant from the Academy of Finland and the TransPoly Consortium grant. The authors also wish to thank the Center for Scientific Computing Ltd. for the allocation of computer time. One of the authors (A.B.) gratefully acknowledges the local hospitality of professor Tapio Alaniisila and the travel support from the Helsinki University of Technology.

1 B. Alberts and D. Bray, Molecular Biology of the Cell (Garland, New York, 1994).
2 J. Darnell, H. Lodish, and D. Baltimore, Molecular Cell Biology (Scien-
Heteropolymer translocation


145101-7