Numerical simulation of conformational variability in biopolymer translocation through wide nanopores

J. Stat. Mech. (2009) P06009

(http://iopscience.iop.org/1742-5468/2009/06/P06009)

Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

Numerical simulation of conformational variability in biopolymer translocation through wide nanopores

Maria Fyta[1,6], **Simone Melchionna**[2,3], **Massimo Bernaschi**[4], **Efthimios Kaxiras**[1,2] **and Sauro Succi**[4,5]

[1] Department of Physics, Harvard University, Cambridge, MA, USA
[2] School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA
[3] INFM-SOFT, Department of Physics, Università di Roma, La Sapienza, Piazzale Aldo Moro 2, 00185 Rome, Italy
[4] Istituto Applicazioni Calcolo, CNR, Viale del Policlinico 13, 00161, Roma, Italy
[5] Initiative in Innovative Computing, Harvard University, Cambridge, MA, USA
E-mail: mfyta@seas.harvard.edu, simone.melchionna@roma1.infn.it, massimo@iac.cnr.it, kaxiras@physics.harvard.edu and succi@iac.rm.cnr.it

Received 27 October 2008
Accepted 3 December 2008
Published 15 June 2009

Online at stacks.iop.org/JSTAT/2009/P06009
doi:10.1088/1742-5468/2009/06/P06009

Abstract. Numerical results on the translocation of long biopolymers through mid-sized and wide pores are presented. The simulations are based on a novel methodology which couples molecular motion to a mesoscopic fluid solvent. Thousands of events of long polymers (up to 8000 monomers) are monitored as they pass through nanopores. Comparison between the different pore sizes shows that wide pores can host a larger number of multiple biopolymer segments, as compared to smaller pores. The simulations provide clear evidence of folding quantization in the translocation process as the biopolymers undertake multi-folded configurations, characterized by a well-defined integer number of folds. Accordingly, the translocation time is no longer represented by a single-exponent power-law dependence on the length, as is the case for single-file translocation through narrow pores. The folding quantization increases with the biopolymer length, while the rate of translocated beads at each time step is linearly correlated with the number of resident beads in the pore. Finally, analysis

[6] Present address: Department Physik, Technische Universität München, Garching, 85747, Germany.

1742-5468/09/P06009+10$30.00

of the statistics over the translocation work unravels the importance of the hydrodynamic interactions in the process.

Keywords: dynamics (theory), polymers, copolymers, polyelectrolytes and biomolecular solutions, lattice Boltzmann methods, molecular dynamics

Contents

1. Introduction **2**

2. Multiscale scheme **3**

3. Configurational analysis **4**

3.1. Quantization of the folding number 6

4. Forces influencing the translocation process **8**

5. Summary **9**

Acknowledgment **10**

References **10**

1. Introduction

Biological systems exhibit a complexity and diversity far richer than those of the simple solid or fluid systems traditionally studied in physics or chemistry. Advances in computer technology and breakthroughs in computational methods have been constantly reducing the gap between quantitative models and actual biological behavior. The main challenge remains the wide range of spatio-temporal scales involved in the dynamical evolution of complex biological systems. In response to this challenge, various strategies have been developed recently, based on composite computational schemes in which information is exchanged between the scales. Motivated by recent experimental studies [1, 2], we apply such a computational scheme to the translocation of a biopolymer through nanopores. The translocation of biopolymers plays a major role in many important biological processes, such as viral infection by phages, inter-bacterial DNA transduction, and gene therapy [3]. The ultimate goal of these studies is to open a path for ultra-fast DNA sequencing by sensing the base-sensitive electronic signal as the biopolymer passes through a nanopore with attached electrodes. The importance of this process has spawned a number of *in vitro* experiments, aimed at exploring the translocation process through microfabricated channels [4] under the effects of an external electric field, or through protein channels across cellular membranes [5, 6]. From a theoretical point of view, simplified schemes [7, 8], coarse-grained or microscopic models with and without hydrodynamic interactions [9]–[11] and mesoscopic approaches [12] are able to analyze universal features of the translocation process. However, a quantitative description of this complex process, which involves the competition between many-body interactions at the atomic or molecular scale, fluid–atom hydrodynamic coupling, as well as the interaction of the biopolymer with wall molecules

J. Stat. Mech. (2009) P06009

in the region of the pore, calls for state of the art modeling, towards which the results presented here are directed.

In a recent paper, the translocation of biopolymers through (relatively) large pores was reported to exhibit the intriguing phenomenon of current-blockade quantization [1]. This was interpreted as an *indirect* evidence that the polymer crosses the pore in the form of 'quantized' configurations, associated with integer values of the folding number, the number of strands simultaneously occupying the pore during the translocation. Such a behavior has been recently confirmed by the *direct* observation of multi-folded configurations in large-scale numerical simulations of biopolymer translocation by the present authors [13]. Here, we considerably extend the scope of such simulations, by increasing the polymer lengths up to 8000 monomers, about an order of magnitude above any previous simulation in the field. This unexplored regime of polymer lengths reveals an extremely rich configurational dynamics, especially for larger pores. In this work, we elaborate more on this enriched dynamics and analyze in detail the phenomenon of folding quantization.

2. Multiscale scheme

Our simulations are based on a multiscale methodology [14], which involves the coupling of a mesoscopic lattice Boltzmann (LB) [15] approach for the solvent degrees of freedom and molecular dynamics (MD) for the biopolymer motion. The comparison of our previous results to those of experiments on DNA translocation allows us to map the anonymous simulated polymers to actual biopolymers [16]. A three-dimensional box of size $(N_x \times N_y \times N_z)\Delta x$ lattice units, with Δx the spacing between lattice points, surrounds both the solvent and the polymer. We take $N_x = 2N_y$, $N_y = N_z$ and $N_x = 128$ and biopolymers with $N_0 = 100$–8000 beads, spanning nearly *two* orders of magnitude in polymer length. At $t = 0$ the polymer resides entirely in the right chamber at $x > (N_x/2)\Delta x$. A separating wall is located in the mid-section of the x direction, at $x = (N_x/2)\Delta x$. At the center of this wall, a cylindrical pore of length $l_p = 3\Delta x$ and diameter d_p is opened up. We have used two different pore diameters: small ($d_p = 5$ in units of Δx) and large ($d_p = 9$ in units of Δx). Translocation is induced by a constant electric field, localized around the pore, like in the experimental settings [2], acting along the x direction and confined to a cylindrical channel of the same size as the pore and length $3\Delta x$ along the streamwise (x) direction. All parameters are measured in units of the LB time step and spacing, Δt and Δx, respectively, which are both set equal to 1. The MD time step is five times smaller than Δt. With the pulling force associated with the electric field in the experiments set at 0.02 and the temperature at 10^{-4}, the process falls in the *fast* translocation regime.

The monomers interact through a Lennard-Jones 6–12 potential with parameters $\sigma = 1.8$, $\epsilon = 10^{-4}$ and a cut-off at 2.02. The interaction of the monomers with the wall is modeled by a Lennard-Jones 6–12 potential with parameters $\sigma_w = 1.5$, $\epsilon_w = 10^{-3}$ and a cut-off at 1.68. Accordingly, the effective width and radius of the surrounding pore should take into account the repulsive monomer–wall interactions, so that a monomer is counted as being inside the pore if contained in a cylinder of effective width $\simeq 6\Delta x$ and radius $\simeq 3.5$ and 7.5 for $d_p = 5$ and 9, respectively. The bonds between adjacent beads are modeled through springs, with a spring constant $k = 0.5$ and an effective equilibrium bond length $b = 1.2$. The solvent has density $\rho = 1$, kinematic viscosity $\nu = 0.1$ and

J. Stat. Mech. (2009) P06009

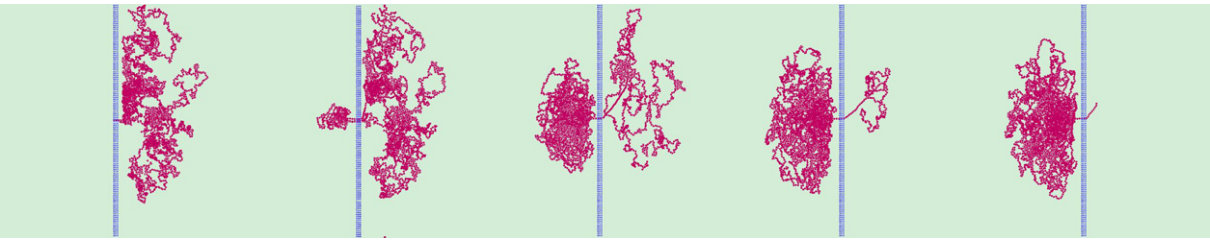

Figure 1. Snapshots of a biopolymer ($N_0 = 4000$) translocating through a wide pore ($d_p = 9$). Different folding conformations are visible. A force (not shown) is applied in the pore region and initiates translocation from right to left. In order to reveal the polymer conformation as it passes though the pore, the wall is shown thinner than it actually is in the simulations.

damping coefficient with the embedded particles $\gamma = 0.1$. Further details can be found in [14]. One lattice spacing $\Delta x = 42$ nm, so the bond length $b = 1.2\Delta x$ corresponds to the persistence length of double-stranded DNA (50 nm). In measuring the residence number of beads in the pore region we have defined a cylinder of length $10\Delta x$ and radius d_p centered at the pore mid-point and with its axis aligned with the pore axis.

3. Configurational analysis

The ensemble of simulations is generated by different realizations of the initial polymer configuration, to account for the statistical nature of the process. The combined statistics over initial conditions and time evolution of the simulations delivers an aggregate ensemble ranging from 500 000 events for the shortest history ($N_0 = 100$, $d_p = 9$) up to nearly 10 million events for the longest one ($N_0 = 8000$, $d_p = 5$). Figure 1 shows different snapshots of such a translocation event for a $N_0 = 4000$ long biopolymer. In the initial stage of the translocation, the nanopore gets populated, with the biopolymer undertaking a conformation with a high number of folds as it passes through the pore. The range of the number of folds explored by the translocation trajectories grows approximately with the cross-section of the pore and the polymer length.

The observed translocation time results from a weighted average over a whole set of multi-folded configurations attained by the polymer during translocation; the weights in this average depend on the number of folds that a polymer of a given length undertakes within the pore, and correspond to the probability distribution function (number of counts normalized to unity) shown in figure 4. Within this picture, a single scaling exponent characterizing the translocation time as a function of the polymer length appears to be insufficient because, above a given length, states with many folds are simultaneously excited, as their number is an increasing function of the pore diameter. The translocation time of these multi-folded configurations is shown to be dominated by the states with a low number of folds, which explains the relatively minor deviations of the translocating time from the single-file power-law expression $\tau \sim N_0^{1.27}$ [2], where τ and N_0 are the translocation time and the length of the biopolymer, respectively. In order to illustrate this behavior, in figure 2, we report the translocation time as a function of the polymer length, N_0, for both pores and all events for each length. The figure shows that, up to

J. Stat. Mech. (2009) P06009

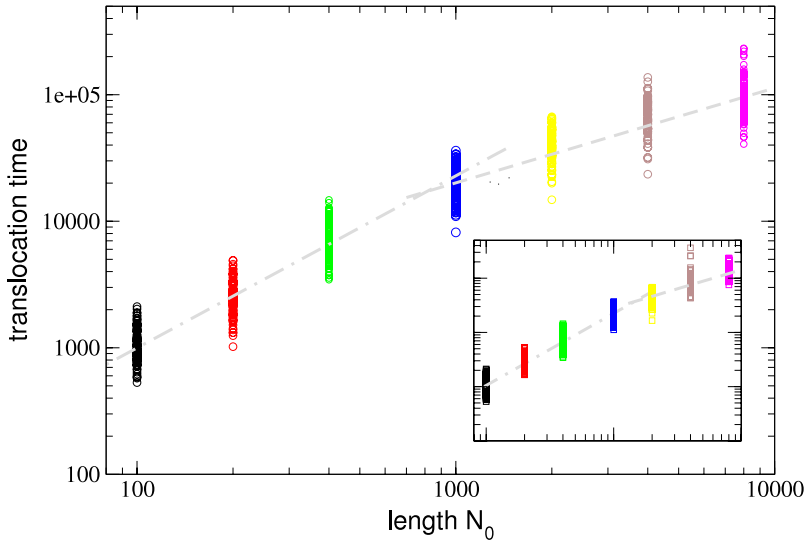

Figure 2. Scatter plots for all events for biopolymers translocating through the wide pore ($d_p = 9$). The two different exponents, 1.36 and 0.75 (see the text), are denoted by the two lines. The inset shows similar events for a narrower pore ($d_p = 5$). Again, the two lines indicate the two exponents 1.31 and 0.70.

a length of $N_0 = 1000$ for the wide ($d_p = 9$) and $N_0 = 2000$ for the narrower ($d_p = 5$) pores, the most probable translocation time τ obeys a scaling law of the form $\tau \sim N_0^{\alpha}$, with $\alpha \sim 1.31$ and $\alpha \sim 1.36$, respectively. These values are slightly larger than the corresponding values for narrow nanopores [2, 14].

This feature does not occur only due to statistical uncertainties, but may also depend on the pore width. Polymers translocating across a pore with $q > 1$, where q denotes a q-times-folded configuration, possibly enhance the resistance to the electric drive, due to the additional energy needed to keep them folded against internal flexional-relaxation forces. These forces tend to unfold the polymer, increasing its interaction with the walls, which may slow down the process as compared with the single-strand translocation [10, 11, 17]. By restricting the analysis to the longest chains, $N_0 \geq 1000$ ($N_0 \geq 2000$) for the wide (narrow) pore, the bending of the curve could be interpreted as indicating the emergence of a new scaling exponent, $\alpha_2 \sim 0.75$ (0.70). Evidently, there are not enough data points to support the exact values for the scaling exponents, but the significant decrease of these exponents for long biopolymers is qualitatively apparent. In figure 2, the two different exponents are denoted by the gray lines (with a different prefactor in each case in order to match the scaling law to the simulation data). We have found that this bending is due to the multi-fold conformation of the translocating biopolymer, which does not necessarily follow a power-law dependence on the polymer size. As a result, in contrast to single-file translocation (the case for narrow pores), multi-fold translocation does not need to obey a standard power-law scaling.

An insightful way to monitor the translocation process is through the evaluation of the fraction of translocating beads with time. Averages over all events for all lengths are plotted in figure 3, where both the number of the translocated beads and the time are presented in reduced units, i.e. scaled with respect to the total number of beads

J. Stat. Mech. (2009) P06009

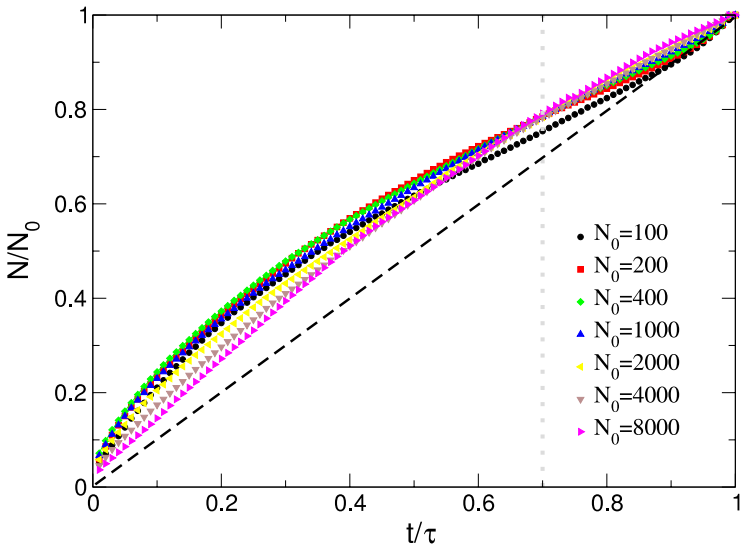

Figure 3. Scaled number of translocated beads with time for all lengths ($N_0 = 100$–8000). The pore diameter is $d_p = 9$ and the two axes are scaled with respect to the total number of beads and the total translocation time, respectively. The black dashed line corresponds to a constant translocation speed, $dN(t)/dt = N_0/\tau$. The vertical dotted line shows the crossover discussed in the text.

and the total translocation time for each event and length. In such units, universality would result in a collapse of the translocation data at all lengths onto a single master curve.

Figure 3 shows a clear trend with size, which only very short chains ($N_0 = 100$) do not follow. Initially, the shorter biopolymers exhibit a larger translocation speed and also a larger acceleration, but at the final stages this trend is reversed, as the longer chains accelerate and eventually translocate faster through the pore. The crossover where the translocation speed of the long biopolymers becomes larger than the corresponding speed of the shorter ones occurs at $0.7N_0$ (see the dotted line in figure 3), a point at which 70% of the chain has already translocated. It is interesting to observe that this is a universal value for all lengths studied here, although at this point, this is a purely observational fact. Again, only polymers with $N_0 = 100$ do not follow this trend, as these are too short. For all times, the beads follow a super-linear trajectory compared to the constant speed translocation (dashed line in the figure), but at the final stages the end parts of the polymers translocate with constant speed. In the case of the narrower pore ($d_p = 5$), the trend is similar, although no universal crossover close to the constant speed limit is observed.

3.1. Quantization of the folding number

In order to investigate the quantization of the resident beads in the pore, we monitored the distribution of the number of pore-resident beads N_r with the number q for a q-times-folded translocation. The resident monomers block the current across the channel, so N_r provides a direct measure of the current drop associated with the biopolymer passage

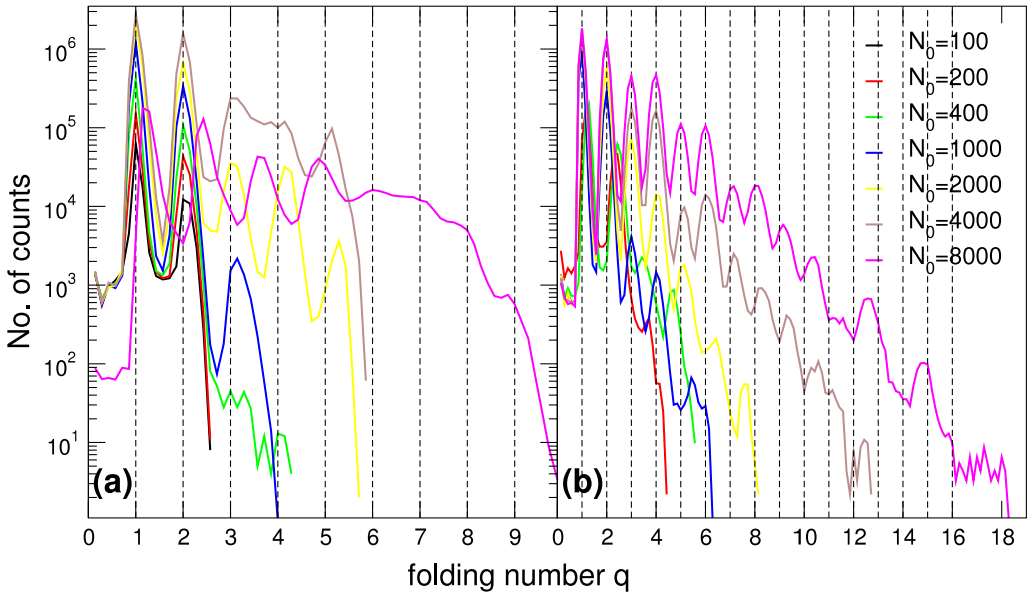

Figure 4. Probability distribution of the folding number q for the entire set of polymer lengths, $N_0 = 100$–8000, and both pore diameters (a) $d_p = 5$ and (b) $d_p = 9$.

through the nanopore. All distributions are peaked around quantized values of q (as defined below). A large fraction of the events are around $q = 1$, and hence have single-file translocation, which is also the conformation at the late stages of the process for all lengths. In figure 4, the cumulative statistics of the folding number, $q = N_r/N_1$, are shown as collected at each time step of every single trajectory for a series of 100 realizations for each polymer length. Here, $N_1 \sim 7$ is the single-file value of the resident number, as also confirmed by visual inspection of the configurations.

The wide pore ($d_p = 9$) reveals a sharp quantization of the distribution of the folding number, with a very well-defined peak in the distribution. Simulation data show that the range of numbers q occupied in this case grows up to $q \sim 10$. The average q over all realizations for each length remains approximately constant, $\sim$1.3, up to $N_0 < 1000$, and grows up to about 2.6 for $N_0 = 8000$. Accordingly, the departure of the translocation time from a power law at large N_0 is mainly due to the increase of the average q with the polymer length (for $N_0 > 1000$). This results from the shift of the probability distribution of the translocation time towards higher numbers q of folds as N_0 increases. For all lengths inspected, the time average of the folding number q remains below 3, because the states with $q = 1$ and 2 continue to be the most populated ones.

The narrower pore ($d_p = 5$) shows the same trends, although for a smaller range of numbers q, up to $q \sim 6$. Above this value, the quantized peaked structure in the folding probability is lost. For this pore, the average q over all realizations for each length remains close to 1.3, up to $N_0 < 2000$, and grows up to about 2.1 for $N_0 = 8000$. These data indicate that quantization of the folding number is better manifested by long chains crossing wide pores. The fact that the average q remains the same ($\sim$1.3) for the two pore sizes and relatively short biopolymers shows that, for these lengths, there is no essential

J. Stat. Mech. (2009) P06009

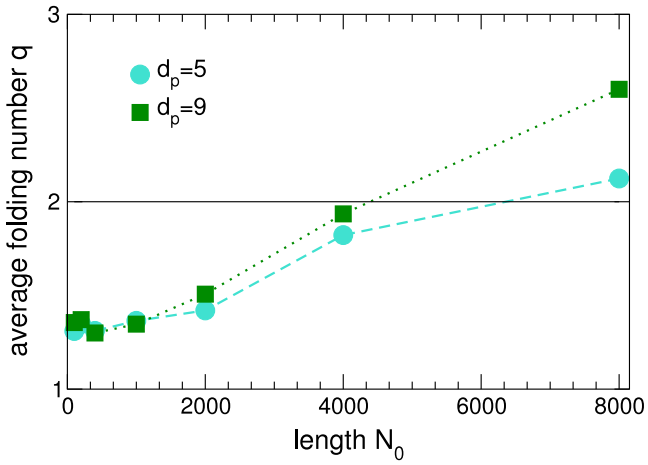

Figure 5. The average folding number q as a function of the polymer length for the wide ($d_\mathrm{p} = 9$) and narrower ($d_\mathrm{p} = 5$) pores.

effect of varying the pore size. In contrast, for a wide pore, the average q increases faster in the range of long biopolymers than for a narrower pore. It is quite natural to expect that long polymers transiting through large pores will prove capable of supporting a greatly richer spectrum of folded conformations as compared to the case for short polymers transiting through narrow pores. Graphical evidence for these points is given in figure 5, where the average folding number q for all cases studied is plotted as a function of the length N_0. Evidently, the average q is almost constant up to $N_0 = 1000$–2000 and increases thereafter, exposing the difference between the two pore sizes, which remains nonetheless rather mild. The highest folding number supported by the polymer–pore system should increase quadratically with the pore diameter, as the number of monomers that a given pore can accommodate is clearly proportional to the cross-section of the pore.

4. Forces influencing the translocation process

The translocation dynamics depends on the strength of the frictional forces exerted by the wall. In single-file translocation, it is well known that strong friction changes the power-law exponent from $\simeq 1.2$ to a linear dependence $\tau \propto N_0$ [18]. In the case of multi-file translocation, it is unclear whether the configurations with high numbers of folds induce high friction conditions, that would decrease the rate of translocation of beads ($\dot{N} = \mathrm{d}N/\mathrm{d}t$) with the resident monomers (N_r) for high values of the folding number q. We have observed that $\dot{N}$ is linearly correlated with N_r with basically the same average slope for all folds. Therefore, frictional forces affect a small layer close to the wall, and have little effect on the group of monomers translocating in the inner region of the pore. This seems to rule out the possibility that the change of exponent is caused by the pore frictional forces. Supporting data from about 100 events for the longest biopolymer ($N_0 = 8000$) translocating through the wide pore ($d_\mathrm{p} = 9$) reveal that all events are centered around integer values of the folding number q. As the characteristic number q increases, the rate of the translocation of monomers with time increases linearly. Inspection of the results for all lengths ($N_0 = 100$–8000) and both pore lengths ($d_\mathrm{p} = 9$ and 5) shows that (a) for

J. Stat. Mech. (2009) P06009

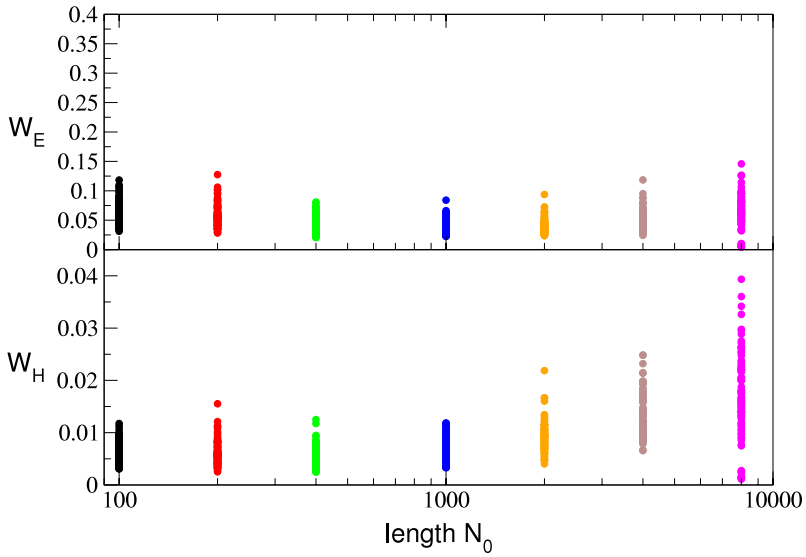

Figure 6. Scatter plot of the translocation work: (top panel) the work for the driving field and (lower panel) the work for the hydrodynamic field as functions of length N_0 for translocation through the wide pore ($d_p = 9$).

the same pore, the slope of the q–$\dot{N}$ curve increases with increasing chain length (long polymers slightly slow down), and (b) for the same polymer length, the linear correlation between q and $\dot{N}$ is better manifested for wider pores.

Another set of forces which greatly affect the translocation are the fluid–biopolymer interactions and the driving force at the pore region. The simulations reveal that solvent *correlated* motion makes a substantial contribution to the translocation energetics. The role of hydrodynamic correlations is best highlighted by computing the work done by the moving fluid on the polymer, W_H, over the entire translocation process as compared to the case for a passive fluid at rest. Inspection of all events for all polymers translocating through wide pores, shown in figure 6, reveals that the cooperation of the surrounding solvent and the solute monomers becomes greater as the length N_0 of the polymer increases. The work of the driving force, W_E (always positive), varies only slightly with the length N_0, as compared to the corresponding behavior of the hydrodynamic work. A rough estimate shows that, as the length increases from $N_0 = 100$ to 8000, there is an increase of about 150% in W_H, while the corresponding increase in W_E is less than 10%. (For these estimates, we used the average values of W_H and W_E for each length N_0.) For a narrower pore ($d_p = 5$), these effects are still visible, though not as strong, revealing a higher cooperativity of the solvent during translocation through wide pores.

5. Summary

In summary, multiscale simulations of the translocation of long polymers, consisting of up to 8000 beads, across wide nanopores, capable of hosting multi-file configurations, provide clear evidence of a sharp quantization of the translocation process. Throughout their translocation, the polymers undertake multi-folded configurations, associated with well-defined integer folding numbers. The observed translocation time reveals a bending

J. Stat. Mech. (2009) P06009

of the scaling law with the biopolymer length, which gives rise to two different exponents, one of them describing the short biopolymers and the other the longer ones. The longer the polymer and the wider the pore, the more evident the quantization of the conformational folds, and this is accompanied by an enhancement of the synergistic role of the hydrodynamic field.

Acknowledgment

SM and MB acknowledge support by the Initiative in Innovative Computing at Harvard University.

References

[1] Li J, Gershow M, Stein D, Brandin E and Golovshenko J A, 2003 *Nat. Mater.* **2** 611
[2] Storm A J, Storm C, Chen J, Zandbergen H, Joanny J-F and Dekker C, 2005 *Nano Lett.* **5** 1734
[3] Lodish H, Baltimore D, Berk A, Zipursky S, Matsudaira P and Darnell J, 1996 *Molecular Cell Biology* (New York: Freeman)
[4] Deamer D W and Akeson M, 2000 *Trends Biotechnol.* **18** 180
Dekker C, 2007 *Nat. Nanotechnol.* **2** 209
[5] Kasianowicz J J, Brandin E, Branton D and Deamer D W, 1996 *Proc. Nat. Acad. Sci.* **93** 13770
[6] Meller A, Nivon L, Brandin E, Golovchenko J and Branton D, 2000 *Proc. Nat. Acad. Sci.* **97** 1079
[7] Kantor Y and Kardar M, 2004 *Phys. Rev.* E **69** 021806
[8] Sung W and Park P J, 1996 *Phys. Rev. Lett.* **77** 783
[9] Matysiak S, Montesi A, Pasquali M, Kolomeisky A B and Clementi C, 2006 *Phys. Rev. Lett.* **96** 18103
[10] Forrey C and Muthukumar M, 2007 *J. Chem. Phys.* **127** 015102
[11] Lubensky D K and Nelson D R, 1999 *Biophys. J.* **77** 1824
[12] Reboux S, Capuani F, Gonzalez-Segredo N and Frenkel D, 2006 *J. Chem. Theory Comput.* **2** 495
[13] Bernaschi M, Melchionna S, Succi S, Fyta M and Kaxiras E, 2008 *Nano Lett.* **8** 1115
[14] Fyta M G, Melchionna S, Kaxiras E and Succi S, 2006 *Multiscale Model. Simul.* **5** 1156
[15] Benzi R, Succi S and Vergassola M, 1992 *Phys. Rep.* **222** 145
[16] Fyta M, Melchionna S, Succi S and Kaxiras E, 2008 *Phys. Rev.* E **78** 036704
[17] Luo K, Ala-Nissila T, Ying S-C and Bhattacharya A, 2007 *Phys. Rev. Lett.* **99** 148102
[18] Zwolak M and Di Ventra M, 2008 *Rev. Mod. Phys.* **80** 141