Single-base DNA discrimination via transverse ionic transport

James Wilson and M Di Ventra

Department of Physics, University of California, San Diego, La Jolla, CA 92093, USA

E-mail: jawilson@physics.ucsd.edu and divernta@physics.ucsd.edu

Received 18 July 2013, in final form 16 August 2013
Published 23 September 2013
Online at stacks.iop.org/Nano/24/415101

Abstract
We suggest the discrimination of single DNA bases via transverse ionic transport, namely by detecting the ionic current that flows in a channel while a single-stranded DNA is driven through an intersecting nanochannel. Our all-atom molecular dynamics simulations indeed show that the ionic currents of the four bases are statistically distinct, thus offering another possible approach to sequencing DNA.

1. Introduction
Measuring transverse electronic currents within a nanochannel to differentiate between DNA/RNA bases [1–3] is a promising new approach for sequencing DNA fast and at low cost. It consists of feeding a single-stranded DNA molecule through a channel equipped with nanometer-scale electrodes able to differentiate (via transverse electrical current) the electronic structure of the various bases as they pass by [1–3]. This approach has been recently realized in various experiments [4–8].

On the other hand, ionic current through a nanopore has been known for over a decade to be useful in detecting when a DNA translocation event has occurred [9–14], and has even been able to give some information about the sequence [15, 16]. However, within a nanochannel, the ionic current blockaded by an arbitrary strand of DNA is a non-trivial convolution of a large number of blockade events from different bases [17], and as such it is difficult to sequence at the single-base level with this physical mechanism, unless the bases are fed one at a time through the opening.

Recent advances in the fabrication of ionic nanochannels [18] have shown that it is possible to use a pair of intersecting nanochannels to detect the transport of DNA. In that particular experiment a double-stranded DNA has been translocated through one channel, while the ionic current flowing through the second, transverse, channel is modulated based on the presence or absence of DNA at the intersection. So far, fabrication techniques have only realized nanochannel widths of about 20–30 nm, still too large to be able to achieve single-base resolution, if at all possible. In fact, the linear dimension of an individual base is on the order of 1 nm, so in order to distinguish a single base, channels of widths comparable to or less than 2 nm are required. Even if those were possible to make, it is not at all obvious whether single-base discrimination would be achievable with transverse ionic transport.

Inspired by these experimental advances, and the possibility to realize devices with intersecting channels as those mentioned above, in this work we seek to determine whether changes in transverse ionic conductance are sufficient to discriminate the different DNA bases, were intersecting nanochannels of width comparable to or less than 2 nm fabricated. To be more specific, the set up we have in mind is illustrated in figure 1. We place a strand of poly(dX)7 where X is one of the bases A, C, G, T in the pore. To optimize the simulation time required, the single-stranded DNA (ss-DNA) is already placed inside of a Si3N4 nanopore of diameter of 1.8 nm that runs along the y-axis. The simulation box has a regular hexagonal shape with an in-radius of 3.3 nm in the x–y plane to correspond with hexagonal periodic boundary conditions. Intersecting this, a transverse pore of diameter 1.4 nm extends in the z-axis direction. The membrane is 2.8 nm thick in the z direction and the simulation box, which contains water stacked in the z direction, has rectangular periodicity in the z direction. In addition, an atom in the backbone of each of the outer two bases in the ss-DNA is fixed in space during the simulation. This allows the bases
that are near the intersection of the pores to move without any constraint, but does not allow the DNA to move out of the pore or to fold up. Moreover, after each simulation run we vary the position of the two outermost bases along the y axis by 0.1 Å, while also keeping the distance between the above two points fixed. This way the ionic currents we detect correspond to different configurations of the single bases, facing the transverse channel. In fact, bases were found to shift several angstroms during the simulation while the end points remained fixed.

2. Simulation

An electric field is then imposed in the z-direction causing ions to flow through the transverse channel in the Si$_3$N$_4$. As has been discussed in [1] the electric field in the longitudinal and transverse directions can be independently manipulated to allow the DNA to translocate through the pore slowly, yet still have sufficient time to make measurements. Many measurements of current for each base can then be made, allowing a current distribution to be built for each base as it passes the channel intersection.

The system is simulated using the molecular dynamics (MD) package NAMD [19]. It is solvated with 5 nm of water on each side of the channel in the z-axis and K and Cl ions are added such that the molarity of the system is 2 M.\(^1\) We use periodic boundary conditions all around. The static energy is minimized and then the system is brought up to room temperature. Next, a Langevin piston is used to equilibrate at room temperature and 1 atmosphere of pressure in the NPT ensemble. Finally the system is run in the NVT ensemble with a Langevin damping term to keep the temperature steady. An electric field is applied transverse to the DNA such that the voltage drop across the cell in the z direction is 0.5 V. The first 2 ns are used as a further equilibration with the same conditions as the production run. During this first 2 ns the system approaches a steady-state current (see also below).

In the previous work covering transverse electronic transport [1, 3], the current was calculated using a single-particle scattering approach [20]. The molecular dynamics simulated the structural fluctuations inherent in the water–pore system, and these fluctuations caused the variations in the current. That study was possible because the timescales in which the electronic transport occurred (due to tunneling) are much shorter than the timescales of the structural fluctuations. Unfortunately, the ionic transport timescales are much longer, so we must simulate many nanoseconds to correctly capture the fluctuations of the ionic current. For each base, 31 simulations were run. In all, we have simulated 1736 ns of MD, with each simulation run of 14 ns. The computational resources required to characterize the current distributions were significantly greater than in the electronic transport study. To accomplish this, we made use of a cluster of 30 dual-core nodes and 10 eight-core nodes. We also used computing time from the Open Science Grid project [21], scavenging time whenever resources were available.

3. Results

We calculate the current directly by looking at the motion of the ions in the system

$$I = \sum_i \frac{q_i \Delta z_i}{l_z \Delta t},$$ (1)

where $l_z$ is the cell length of the system, and the sum is over all ions in the system. $\Delta z_i$ is the distance the $i$th ion has traveled in an interval of time $\Delta t$. Other methods of calculating the current were tested and they yielded similar results to equation (1). For instance, the slope of the curve $Q(t) = \int_0^t dq$ at steady state, where $S$ is the surface of the transverse channel (at steady state the choice of this surface is irrelevant [20]) yields similar currents.

In about 1–2 ns, the ions move into a steady-state current-carrying state, so we do not begin to calculate the current until after this time period. The next 12 ns are then used to calculate the current for that run. This short transient is clear from figure 2 where we plot $Q(t)$ for a typical run. Other, shorter, lengths of time were checked, but the best, most stable results were obtained with the present timescale. A rough estimate of the standard deviation of the current measurement is that it would scale as the inverse square root of the measurement window [22]. Experimental measurements for a reasonable sampling frequency would average over a longer timescale than we have simulated, so we expect the current distributions to be even less noisy than we have calculated. Our rate of 12 ns per sample would correspond to a sampling frequency of approximately 80 MHz. Ionic current measurements have been recently conducted at 1 MHz [23], which would be a factor of 80 slower than our simulations.

The current distributions from the all-atom MD simulations are shown in figure 3. First we note, as was also

\(^1\) The choice of a larger molarity than the typical 1 M is again to reduce the already demanding computational requirements.
However, as also indicated by the reverse order of C and T, the exclusion volume alone cannot fully explain the order of the average currents we have obtained, and other microscopic effects must also intervene such as the strength of the base dipoles, roughness of the surfaces at the channel intersection, and possibly other ones. Again, this confirms that the actual distributions will likely depend on the structural properties of each device.

4. Analysis

In order to sequence with this particular approach, we then parallel the protocol suggested in [1] by one of us (MD). First, run a strand of DNA with known composition through the longitudinal channel, e.g., a homogeneous strand of the four different bases. Then, while the DNA is translocating through the channel, measure the current as many times as possible and build up a distribution of currents for each DNA base. Once this distribution for each base is obtained, the DNA strand that is to be read is sent through. The current is read N times for each base X as it passes through the intersection, and each of those N readings is analyzed for the probability that it could be an A, C, T, or G. The probability of X being correctly identified after one measurement is given by [1]

\[
\langle P_X(I) \rangle = \frac{n_X(I)}{n_A(I) + n_C(I) + n_T(I) + n_G(I)}
\]

where \( n_X(I) \) is the height of the normalized distribution for current \( I \), and \( X \) is one of the bases A, C, G, T.

After \( N \) measurements, the probability that the base at the intersection of the two channels is correctly identified is [1]

\[
\langle P_X^N \rangle = \frac{\prod_{i=1}^{N} n_X(I_i)}{\sum_{J=A,C,G,T} \left( \prod_{i=1}^{N} n_J(I_i) \right)}.
\]

A Monte Carlo method was employed with the distribution data obtained from our simulations to calculate the probability of correctly identifying a base after \( N \) measurements. For each base, \( I_1 \) through \( I_N \) were randomly generated from the current distribution for that base. The probability \( P_X \) is then calculated from equation (3). Assuming an equal proportion of each of the bases within the strand being sequenced, the probability of correctly identifying a base at random after \( N \) independent measurements is just the average

\[
P_N = \frac{1}{4} (P_N^A + P_N^C + P_N^G + P_N^T).
\]

We then average over realizations to arrive at the probability, \( P = \langle P_N \rangle \), that we correctly identify a random base after \( N \) measurements.

In the inset of figure 3, we plot a graph of uncertainty, \( E = 1 - P \), in identifying bases versus number of independent measurements. When at least 20 measurements are taken, the probability of error is less than 0.1%. This seems to suggest that, with these particular distributions, we need fewer measurements with this method than with transverse electronic transport for the same error rate [1]. However, as already noted ionic transport measurements are typically slower than electronic ones. Nevertheless, if we measure the ionic current at a reasonable rate of 100 kHz, then DNA could
be identified at 5000 bases per second using this scheme. Therefore, an entire genome could be sequenced in less than 7 days without parallelization. Clearly, this estimate may be somewhat off according to the actual distributions obtained experimentally. Furthermore, the error rate assumes that the pore is made small enough that neighboring bases do not affect the current, and that it is easy to tell when a base passes over the intersection. If that is not the case, it may take several more measurements per base simply to know that a change in base at the intersection has occurred. These effects would all increase the number of measurements required.

5. Conclusions

We conclude that the approach we have described in this paper can indeed discriminate between the four DNA bases and has thus potential as an alternative sequencing method. Furthermore, this method would easily lend itself to parallelism. For example, a device dedicated to each chromosome would decrease the time required to sequence an entire genome by a factor of 10. Finally, this same idea could be implemented as a protocol for protein sequencing. We are now in the process of assessing the feasibility of this last possibility.

Acknowledgments

This research has been partially supported by the NIH-National Human Genome Research Institute. We also gratefully acknowledge discussions with M Ramsey. Part of the calculations in this work were performed by using resources provided by the Open Science Grid, which is supported by the National Science Foundation and the US Department of Energy.

References

[1] Lagerqvist J, Zwolak M and Di Ventra M 2006 Fast DNA sequencing via transverse electronic transport Nano Lett. 6 779–82
[2] Zwolak M and Di Ventra M 2005 Electronic signature of DNA nucleotides via transverse transport Nano Lett. 5 421–4
[3] Krems M, Zwolak M, Pershin Y V and Di Ventra M 2009 Effect of noise on DNA sequencing via transverse electronic transport Biophys. J. 97 1990–6
[4] Ohshiro T, Matsuura K, Tsutsui M, Furuhashi M, Taniguchi M and Kawai T 2012 Single-molecule electrical random resequencing of DNA/RNA Sci. Rep. 2 501–7
[5] Ivanov A P, Instulti E, McGilvery C M, Baldwin G, McComb D W, Albrecht T and Edel J B 2010 DNA tunneling detector embedded in a nanopore Nano Lett. 11 279–85
[6] Liang X and Chou S Y 2008 Nanogap detector inside nanofluidic channel for fast real-time label-free DNA analysis Nano Lett. 8 1472–6
[7] Chang S, Huang S, He J, Liang F, Zhang P, Li S, Chen X, Sankey O and Lindsay S 2010 Electronic signatures of all four DNA nucleosides in a tunneling gap Nano Lett. 10 1070–5
[8] Tsutsui M, Matsuura K, Ohshiro T, Furuhashi M, Taniguchi M and Kawai T 2011 Electrical detection of single methylcytosines in a DNA oligomer J. Am. Chem. Soc. 133 9124–8
[9] Kasianowicz J J, Brandin E, Branton D and Deamer D W 1996 Characterization of individual nucleotide molecules using a membrane channel Proc. Natl Acad. Sci. 93 13770–3
[10] Deamer D W and Branton D 2002 Characterization of nucleic acids by nanopore analysis Acc. Chem. Res. 35 817–25
[11] Merchant C A et al 2010 DNA translocation through graphene nanopores Nano Lett. 10 2915–21
[12] Garaj S, Hubbard W, Reina A, Kong J, Branton D and Golovchenko J A 2010 Graphene as a subnanometre trans-electrode membrane Nature 467 190–3
[13] Schneider G F, Kowalczyk S W, Calado E V, Pandrau G, Zandbergen H W, Vandersypen L M K and Dekker C 2010 DNA translocation through graphene nanopores Nano Lett. 10 3163–7
[14] Howorka S and Siwy Z 2009 Nanopore analytics: sensing of single molecules Chem. Soc. Rev. 38 2360–84
[15] Derrington I M, Butler T Z, Collins M D, Manrao E, Pavlenok M, Niederweis M and Gundlach H 2010 Nanopore DNA sequencing with MspA Proc. Natl Acad. Sci. 107 16060–5
[16] Manrao E A, Derrington I M, Laszlo A H, Langford K W, Hopper M K, Gillgren N, Pavlenok M, Niederweis M and Gundlach H J 2012 Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase Nature Biotechnol. 30 349–53
[17] Branton D et al 2008 The potential and challenges of nanopore sequencing Nature Biotechnol. 26 1146–53
[18] Menard L D, Mail C E, Woodson M E, Alaire J P and Ramsey J M 2012 A device for performing lateral conductance measurements on individual double-stranded DNA molecules ACS Nano 6 9087–94
[19] Phillips J C, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel R D, Kale L and Schulten K 2005 Scalable molecular dynamics with NAMD J. Comput. Chem. 26 1781–802
[20] Di Ventra M 2008 Electrical Transport in Nanoscale Systems (Cambridge: Cambridge University Press)
[21] Pordes R et al 2007 The open science grid J. Phys.: Conf. Ser. 78 012057
[22] Aksamitieiev A 2010 Deciphering ionic current signatures of DNA transport through a nanopore Nano Lett. 10 468–83
[23] Rosenberg J K, Watanu M, Merchant C A, Dnmic M and Shepard K L 2012 Integrated nanopore sensing platform with sub-microsecond temporal resolution Nature Methods 9 487–92
[24] Zwolak M and Di Ventra M 2008 Colloquium: physical approaches to DNA sequencing and detection Rev. Mod. Phys. 80 141