DNA sequencing based on physical properties at single nucleotide resolution

F. Ebadi Jalal and H. Nejat Pishkenari*

Lab. of Computational Nanomechanics, Department of Mechanical Engineering, Sharif University of Technology, Tehran, P.O. Box 11155-9567, Iran

Received 9 January 2018; received in revised form 17 March 2019; accepted 25 June 2019

KEYWORDS
DNA sequencing; Nanochannel; Nanotube; Physical properties; Molecular dynamics simulation.

Abstract. To realize such objectives as genomic medicine, it is required to develop economical and fast methods for DNA sequencing. In this paper, a novel approach is developed to significantly improve the efficiency of DNA sequencing based on physical differences between nucleotides. Here, it is claimed that the reason for the rather low resolution of sequencing based on physical differences is the extremely nonlinear and complex dynamics of DNA, making DNA translocation with respect to detectors highly dependent on initial conditions and environmental disturbances. In various sequencing methods, the position and orientation of nucleotides are different at the detection time. At a low signal-to-noise ratio, different dynamics of nucleotides make it difficult to detect slight differences in the physical properties of DNA bases. This claim is confirmed by designing a sequencing nanodevice in which the motion of a stretched single-stranded DNA (ssDNA) is constrained in such a way that the axis of ssDNA backbone is fixed and each nucleotide lies on a fixed plane at the detection time. Further, nonlinear effects of ssDNA and the interactions of detectors are reduced as low as possible. Results of molecular dynamics indicate that specific and distinct signals of each type of nucleotide will be generated under constrained conditions.

© 2020 Sharif University of Technology. All rights reserved.

1. Introduction

One of the most common techniques widely used in DNA sequencing based on physical properties is the application of nanopores [1–12]. Two different types of nanopores exist: biological nanopores [13–15] and solid-state nanopores [16–20]. Biological nanopores, such as α-hemolysin [13,14] or MspA [15], inserted into the lipid bilayer are stable only in limited conditions and are characterized by high thickness and constant diameter [21,22]. Solid-state nanopores are fabricated by different methods such as ion beam sculpting [17], transmission electron microscope [23], electron beam lithography [24], and ion beam lithography [25] from silicon nitride [16], silicon oxide [17], aluminum oxide [18], or graphene [19,20] substrates. They are much more stable than biological nanopores and are of different sizes, and DNA translocates through them fast [22]. In addition to the nanopores, fluidic nanochannel is another nanostructure proposed for DNA sequencing based on physical differences [26,27]. In the following, some physical approaches to DNA sequencing are presented briefly.

Most of the physical approaches to DNA sequencing use the differences between electronic characteristics in nucleotides. For example, to measure ionic current blockade in an applied electric field, DNA electrophoretically translocates through the nanopore and one of its nucleotides blocks the nanopore space

* Corresponding author. Tel.: +98 21 66165543
E-mail address: nejet@sharif.edu (H. Nejat Pishkenari)

DOI: 10.24200/sci.2019.50231.1591
mechanically at any moment. Both types of the aforementioned nanopores have been employed in this approach [14.28–31]. Usually, an ionic current signal is too poor to allow DNA sequencing at a single-nucleotide resolution [22]. In addition to the mentioned example approach, the differences in nucleotides conductance are considered by measuring transverse electronic transport between two electrodes spaced closely to each other and embedded in the nanopore when DNA translocates through it [32–35]. As it is clear, under the same above-mentioned conditions, distinct nucleotides, fixed between two electrodes, transfer completely different electronic currents [32]; however, in practice, fluctuations in the orientation of nucleotides significantly reduce the signal-to-noise ratio [36]. By applying a small transverse electric field, the orientation of nucleotides towards electrodes can be relatively controlled upon detection [37]. Another idea entails good electrical conductivity of a single atomic layer of graphene nanogap, according to which sequencing can be performed by measuring the transverse conductance of graphene nanogap when ssDNA translocates through it [38]. The third instance considers the differences in nucleotide capacitance by measuring voltage fluctuations in a capacitor embedded in a nanopore as DNA translocates through it [39,40]. Contrary to the transverse electronic current method, this approach considers long-range coulomb interactions between the nucleotides and detectors [36]. Hence, effects of adjacent nucleotides on the signal of a single-nucleotide increase upon detection.

Moreover, researchers attempt to improve the resolution of sequencing by modifying the previous approaches. By functionalizing, some methods could achieve better interactions between nucleotides and detectors or control DNA translocation in nanodevices [41–43]. Some others combine physical approaches and chemical reactions [15,44,45]. One of them is the use of optical methods in which an amplified DNA is ripped apart in a nanopore and sequencing is performed by analyzing the optical signal emitted from it [46]. In addition, the force detection method uses optical tweezers to decelerate DNA translocation through nanopores in ionic-current detection [47]. Another proposed approach is based on the differences of the molecular friction of nucleotides when an ssDNA is pulled through a pore in a substrate [48], or when a tightly fitting molecular ring is pulled over an ssDNA [49]. The major difficulty of these methods is to thread ssDNA through the pore [48]. However, many ideas presented so far have not been implemented experimentally. They use computer simulations to model how the DNA sequencing process is performed in these methods.

In computer simulations, molecular dynamics, a method that can reveal all events happening in interactions among DNA, detectors, and surroundings, provides a route for designing the sequencing nanodevices and understanding how DNA transports and nucleotides are oriented in such devices [50–52]. Despite the many theoretical and experimental researches about DNA sequencing, the developed methods based on physical properties have failed to play their role perfectly so far. However, the methods in which dynamics of DNA translocation is somehow controlled achieve better results than others [26,37]. Therefore, to attain a single-nucleotide resolution, ssDNA motion should be controlled during DNA sequencing [51]. Complex and nonlinear dynamics of ssDNA during translocation [53] created by extremely nonlinear Lennard-Jones and electrostatic forces, numerous degrees of freedom, and many bonds that connect various atoms, in addition to fluctuations in DNA, caused a significant reduction in the signal-to-noise ratio. Furthermore, it has been shown that when ssDNA nucleotides have the same position and orientation with respect to detectors, they can obtain specific signals for each type of a detector [32]. Hence, to resolve this problem, it is necessary that the dynamics of DNA be constrained and the nonlinear effects in ssDNA-detectors interactions be minimized. Constrained dynamics of ssDNA should be able to enforce nucleotides with the same position and orientation upon detection. This paper applies molecular dynamics simulations and considers the differences in the Lennard-Jones interactions between distinct nucleotides and two nanotubes when ssDNA passes through them; in this respect, DNA sequencing can be performed fast at a high rate of accuracy.

2. Simulation setup

Hence, the proposed approach functions by constraining the dynamics of ssDNA such that the backbone axis of stretched ssDNA moves on a straight line and each nucleotide nearly lies on a given fixed plane at the detection time. To this end, a stretched ssDNA is translocated through a gold nanochannel whose width is linearly reduced from ~30 Å to ~8 Å through ~65 Å vertical displacement (see Figure 1). A gradual decrease in width reduces the molecular friction and, therefore, causes a smooth and uniform motion for ssDNA. After ssDNA translocates through this v-shaped part, it lies between two flat faces, which is the location of detected nucleotides (Figure 1). Thus, ssDNA repulsive Lennard-Jones interactions with these faces enforce ssDNA nucleotides to lie in the mid-plane of these nucleotides. In addition, the ssDNA backbone, whose atoms are located in a more compact and regular arrangement, with attractive Lennard-Jones interactions adheres to the end of the channel, provided that the backbone is close to it. Therefore, the dynamics of ssDNA motion is constrained so that the
1. Bonded interactions among atoms of one nanotube ($K_b$);
2. Lennard-Jones interactions between atoms of one nanotube and those of the other nanotube ($K_{LL}$).

Assuming that these two kinds of stiffness act in parallel, the equal stiffness, $K_{eq}$, resulting from all bonded and non-bonded interactions is calculated through $K_{eq} = K_b + K_{LL}$, where the term $K_b$ is constant at small values of $\delta$: hence, a linear relationship holds between $\delta$ and the forces applied by nucleotides. In contrast, $K_{LL}$ is extremely nonlinear and varies by changing $\delta$ in terms of $1/r$ and $1/r^6$, where $r$ is the distance between two nanotubes. The term stiffness is the reason why $K_{eq}$ becomes large when two nanotubes approach each other; then, it is reduced to a considerable extent as the nanotubes retract. Initially, nanotubes are rotated to ensure the proximity of their two ends to each other and, as a result, the value of $K_{LL}$ is reduced to become equal to the value of stiffness $K_{eq}$; consequently, the relationship between the forces from nucleotides and $\delta$ is approximately linear. Therefore, nonlinear effects of the interaction between nucleotides and nanotubes are reduced and the resolution of resulting signals is improved. It should be noted that as the length of nanotubes increases, $K_b$ reduces and $K_{LL}$ increases.

Second, after detecting the nucleotides, two parts of them are found in the space between two nanotubes simultaneously (Figure 1). The two aforementioned parts include: (a) the sugar part which is of the same amount in all nucleotides and, thus, does not help detect the nucleotides and (b) the base part which is not common in distinct nucleotides. The configuration of nanotubes is the reason why each part of the nucleotides has a different role in changing $\delta$ since a common part of nucleotides passes through a stiffer site where the distance between two nanotubes is relatively large. In contrast, the base parts have passed through a more flexible site, where the nanotubes are closer to each other. Thus, the base parts affect $\delta$ more than the sugar part.

3. Method

All simulations are performed in an all-atom implicit solvent using NAMD2 [54] and are visualized in VMD [55]. Each ssDNA is generated by X3DNA software [56] and parameterized in the CHARMM27 force field [57]. Lennard-Jones parameters of gold are assumed to be equal to $\varepsilon = 0.039 \text{ kcal mol}^{-1}$ and $\sigma = 2.934 \text{ Å}$ [58]. For simulating constant tensile force applied to ssDNA backbone by AFM tip, Steered Molecular Dynamics (SMD) simulations are used in which two forces equal in magnitudes and opposite in directions are applied to the phosphorous atoms of the
first and last nucleotides of ssDNA. The magnitude of the forces is 50 kcal mol⁻¹, which is large enough to keep
the ssDNA backbone straight. At first, the backbone
of the stretched ssDNA is placed close to the end
face of the channel so that attractive Lennard-Jones
interactions could fix the axis of backbone of ssDNA
during simulations.

All simulations should undergo an equilibration step. Equilibration consists of energy minimization in 1000 steps and relaxation in 500 ps under the NVT ensemble. During relaxation, temperature is set to 300 K using the Langevin thermostat with a damping coefficient of 5 ps⁻¹ [59]. After equilibration, the production stage consisting of SMD simulation is conducted in which ssDNA is pulled at a constant velocity of (0.25 Å ns⁻¹) through probes, while it remains
straight by applying two equal forces. It is done by a
spring with stiffness:

$$50 \kappa = 500 \text{ kcal mol}^{-1} \text{Å}^2$$

one end of which is connected to the phosphor of the
first nucleotide and its other end is pulled at a constant
velocity. The stiffness is large enough to prevent the
“stop and go” ratchet-like motion of ssDNA. Such high
stiffness of a spring can be achieved by multiple covalent
bonds operating in parallel. Periodic boundary
conditions are applied in three axis directions, and
Particle Mesh Ewald (PME) [60] is used for computing
long-ranged electrostatic interactions. velocity-Verlet
algorithm [59] was employed for time integration at a
time step of 2 fs. By using SHAKE algorithm [61],
hydrogen bonds were restrained. Lennard-Jones in-
teractions were reduced using a switching function
starting at 10Å and vanishing at 12 Å.

4. Results and discussion

Initially, sequencing of four homogeneous polynu-
cleotides including:

3'-AAAAAA-5',
3'-TTTTTTTT-5',
3'-CCCCCCCC-5',
3'-GGGGGGGG-5',

is simulated. The results of these simulations are shown
in Figure 2, in which δ parameter versus time and a
smooth curve fitting to it are given. The fitted curve
smooths the data using a moving average filter. Then,
sequencing of decussate arrangements including:

3'- ATATATAT-5',
3'- ACACACAC-5',
3'- AGAGAGAG-5',
3'- CTCTCTCT-5',
3'- GTGTGTGT-5',
3'- GCGCGCGC-5',

Figure 2. Diagram of parameter δ versus time during simulations and a smooth curve fitting to it (simulation results of
sequencing homogeneous polynucleotides): (a) 3'-AAAAAA-5', (b) 3'-TTTTTTTT-5', (c) 3'-CCCCCCCC-5', and (d)
3'-TTTTTTTT-5'.


is performed. In addition, the results of these simulations are shown in Figure 3, in which $\delta$ parameter versus time and the smoothed data are indicated.

According to Figure 2, the proposed nano-device can generate identical and distinct signals for each nucleotide in homogeneous arrangements. This means that all signals generated for a specific nucleotide are approximately similar, while the signals generated for different nucleotides are easily distinguishable. For nucleotides A, T, C, G, $\Delta\delta = \delta_{\text{max}} - \delta_{\text{min}}$ and they are 1.9 Å, 2.2 Å, 1.9 Å, and 1.7 Å, respectively, in steady-state conditions. As can be seen from Figure 2, it takes a longer amount of time to translocate the nucleotide G than other nucleotides.

This setup can generate a certain distinct signal between each pair of different nucleotides, except C and T (Figure 3). For sequences that include nucleotide T, the minimum value of the signal, $\delta_{\text{min}}$, is smaller than that in other sequences. Of note, at the beginning of the simulation, dynamics of nanotubes has not reached a steady state and, therefore, the resulting signals do not have enough resolution. However, after applying a transient unsteady response, the resolution of signals is enhanced. For instance, if the simulation of decussate arrangement $3'-\text{CTC}..\text{TCT}-' \delta$ is performed, the signal resolution will significantly improve (Figure 4) such that one can easily distinguish signals of the nucleotides C and T. According to Figure 4, it takes a longer amount of time to translocate the nucleotide T than the nucleotide C. The time required for the translocation of the nucleotides T is about 20 picoseconds, while it is approximately 15 picoseconds for nucleotides C. Moreover, in Figure 4, the maximum value of the signal, $\delta_{\text{max}}$, for the nucleotide C is larger than that for the nucleotide T.

Based on Figures 2 and 3, it can be concluded that signals of a certain nucleotide depend on the adjacent nucleotides. For instance, signals of the nucleotide A arranged as $3'-\text{AAAAA}..\text{A}-'$, $3'-\text{ATA}..\text{AT}-'$, $3'-\text{ACA}..\text{ACA}-'$, and $3'-\text{AGA}..\text{AGA}-'$ (according to Figures 2(a) and 3(a), (b), and (c), respectively) or signals of the nucleotide T arranged as $3'-\text{TTTTT}..\text{T}-'$, $3'-\text{ATA}..\text{AT}-'$, $3'-\text{CTC}..\text{TCT}-'$, and $3'-\text{GTG}..\text{GTG}-'$ (according to Figures 2(b) and 3(a), (d), and (e), respectively).
Besides, the force fields used in simulations which are linearly approximated for small deformations might lose validity for simulations, too. It is thus required to check for the deformation of ssDNA bonds during simulations. Moreover, of note, similar to the system dimensions, an implicit solvent model was used in these simulations to reduce the calculation costs. This approach might affect the results by omitting some nonlinear effects caused by explicit water molecules.

Finally, one of the most common challenges of designing devices on a nano scale is the possibility of its fabrication. Fabricating a nanodevice for a nanochannel and embedding nanotubes with \( \sim 3.13 \) Å diameter in a \( \sim 10 \) Å gap are important related challenges. Perhaps, such a nanochannel can be made using lithography techniques [62]. In addition, nanotube detectors may be replaced by other detectors such as those used in transverse electronic current and capacitance methods, where sequencing can be performed based on electronic differences between nucleotides. By removing disturbances from nanotubes’ dynamics, these detectors can be of higher resolution. However, complications in fabrication are out of the scope of this study, and the purpose of this paper is to demonstrate that the resolution and accuracy of DNA sequencing can be significantly improved by constraining the dynamics of ssDNA during sequencing.

5. Conclusion

In summary, in this paper, a novel approach in DNA sequencing was proposed based on physical differences of nucleotides and, then, a sequencing nano-device was designed accordingly. This study argued that the reason for the rather low resolution of sequencing methods based on physical differences of nucleotides was the nonlinear and complex nature of the DNA dynamics. This nonlinear dynamic caused the positions and orientations of nucleotides with respect to detectors to be dependent on initial conditions and environmental disturbances. For this reason, to design a sequencing nano-device, an attempt was made to reduce the nonlinear effects of nucleotides and detectors. Accordingly, the resolution of the resulting signals can be improved significantly. From the results of molecular dynamics simulations, it can be concluded that the resolution of signals has highly improved so much so that one can specify the type of bases from output signals with insignificant data processing. Theoretically, sequencing by this method is very fast such that a DNA strand can be sequenced in only a matter of a few seconds.

The occurrence of large deformation of single-standard DNA (ssDNA) structure while translocating through the channel is probably one of the limitations of this approach. Large deformation can cause plastic deformations to ssDNA structure and destroy it;

References


Biographies

Farhad Eladi Jalal received his BSc from Bu-Ali Sina University, School of Mechanical Engineering. He continued his career by attaining MSc degree from Sharif University of Technology. He performed his researches at the Computational Nano-Mechanics Laboratory.

Hossein Nejat Pishkenari received his BSc, MSc, and PhD degrees in Mechanical Engineering from the Sharif University of Technology in 2003, 2005, and 2010, respectively. Then, he joined the Department of Mechanical Engineering at the Sharif University of Technology in 2012. Currently, he is directing the Computational Nano-mechanics and Nano-robotics Laboratories. His research interests are molecular dynamics, robotics and control, nonlinear dynamics, and vibration.