Detection of defective DNA in carbon nanotubes by combined molecular dynamics/tight-binding technique

Yang Xu, Xiaobing Mi, and N. R. Aluru[a)]
University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

(Received 14 July 2009; accepted 25 August 2009; published online 18 September 2009)

A tight-binding method combined with molecular dynamics (MD) is used to investigate the electrostatic signals generated by DNA segments inside short semiconducting single-wall carbon nanotubes (CNTs). The trajectories of DNA, ions, and waters, obtained from MD, are used in the tight-binding method to compute the electrostatic potential. The electrostatic signals indicate that when the DNA translocates through the CNT, it is possible to identify the total number of base pairs and the relative positions of the defective base pairs in DNA chains. Our calculations suggest that it is possible to differentiate Dickerson and hairpin DNA structures by comparing the signals. © *2009 American Institute of Physics*. [doi:10.1063/1.3231922]

Rapid and cheaper techniques to sequence and detect DNA abnormality could revolutionize the use of genetic information and biomedical research.[1–3] Defective DNA segments can commonly be found in cancer patients, where one or multiple DNA bases as well as the corresponding backbones are inserted, missing or unpaired.[4,5] Defective DNA pairs can lead to cell death or mutations. A series of recent experiments studied the translocation of DNA molecules through narrow pores, which selectively allow single- or double-stranded DNAs to pass through.[3,6] Recent experiments[7,8] and molecular dynamics (MD) simulations[9,10] have shown that carbon nanotubes (CNTs) can function as molecular channels for water and DNA transport. CNTs have many exquisite properties that can be exploited to develop next-generation devices with high sensitivity, fast response, low cost, and large volume production.[11–13] The possibility to detect and identify different charged ions translocating through CNTs has recently been reported.[14] Since DNA bases with backbones are negatively charged, the confinement of DNA inside CNT makes the detection of electrostatic signals generated by DNA bases possible. Furthermore, abnormal and normal DNA segments have different charge distribution, so it is also possible to differentiate them by measuring the electrical signals. Electrical biosensing approaches are promising for fast DNA detection because of their potential for miniaturization and automation without chemically changing the molecular structures.[13] In this paper, we present combined MD and tight-binding simulations of DNA translocating through a CNT. We show that it is possible to count the number of base pairs and estimate the relative positions of defective base pairs in the DNA segments.

A schematic of the system is shown in Fig. 1. A DNA chain is translocated from reservoir B to reservoir A through the CNT. A semiconducting CNT with solvent reservoirs attached at both ends is considered. DNA electrostatic signal patterns are investigated when DNA bases translocate through the CNT together with water and ions. The configurational structure of water, ions, and DNA inside the CNT are obtained by performing MD simulations. The screening effects are computed by self-consistent tight-binding (TB) simulations.[15] Since semiconducting CNTs have much weaker screening ability than metallic CNTs,[14] we use semiconducting CNTs to obtain stronger electrical signals when DNA bases are present inside the CNT. The TIP3P model is used for the geometry and atomic charges of water molecules.[16] Two types of DNA sequences, namely, Dickerson and hairpin structures,[17] are considered. The topologies of a 12-segment double stranded Dickerson (*B*-type) DNA[17,18] and an eight-segment hairpin DNA[13,19] are obtained from the DNA database. Defective DNA segments are generated by removing one or multiple DNA bases as well as the corresponding backbone structure. To keep the system neutral, 40 Na^+ and 18 Cl^- ions are added for the Dickerson case and 40 Na^+ and 24 Cl^- ions are added for the hairpin case. For simplicity, we assume that the presence of CNT does not affect the partial charges of the DNA and water molecules. As DNA, water molecules and ions move from reservoir B to reservoir A, the electrostatic potential detected at the measuring point is a combination of the screening of CNT and the field generated by the DNA, ions, and water molecules inside the CNT. The measuring point is located above the center of the CNT at a distance of 2 nm from the nanotube axis. Water dipoles can also partially screen the electric field generated by the DNA. However, when water and DNA enter the hydrophobic interior of small-diameter

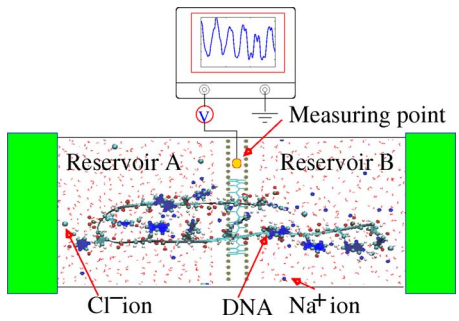

FIG. 1. (Color online) Schematic of the system. DNA is transported from reservoir B to reservoir A through the (22,0) CNT with a diameter of 1.724 nm and length of 0.426 nm. The potential variation at the measuring point is calculated as the DNA molecules pass through the CNT. The measuring point is located above the center of the CNT at a distance of 2 nm from the nanotube axis. The size of the simulation box is $4.0\times4.0\times19.8$ nm^3.

a)Author to whom correspondence should be addressed. Electronic mail: aluru@illinois.edu.

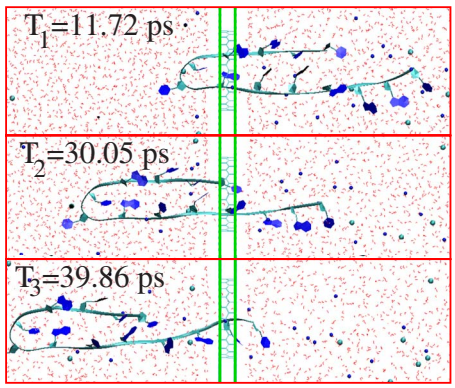

FIG. 2. (Color online) Three MD snapshots showing the translocation of an eight-segment hairpin DNA through the (22, 0) semiconducting CNT.

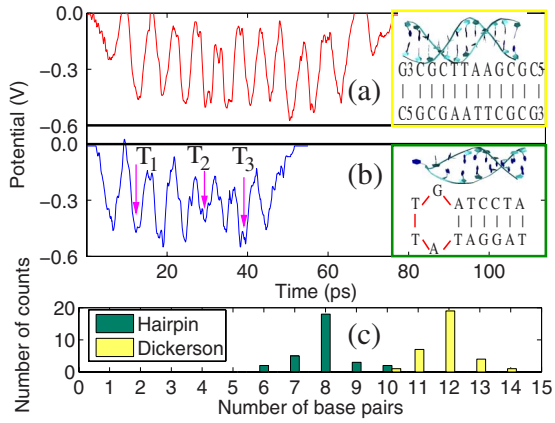

FIG. 3. (Color online) Variation of electrostatic potential as a function of time calculated at the measurement point (see Fig. 1) for (a) the 12-segment Dickerson DNA and (b) the eight-segment hairpin DNA as they translocate through the CNT. T_1, T_2, and T_3 correspond to the time of the MD snapshots in Fig. 2. The insets in [(a) and (b)] show the Dickerson and the hairpin structures used in the MD simulations. (c) For the Dickerson case, 19 of 30 MD simulations indicate that the Dickerson DNA has 12 segments. The margin of error is within two base pairs. For the hairpin case, 18 of 30 MD simulations indicate that the DNA has eight segments. The margin of error is within two base pairs.

CNTs, the overall screening ability of water molecules is reduced because of the limited number of water molecules inside the CNT.[20,21] Thus, the effective electrostatic potential signals at the measuring point are dominated by DNA charges. Because the DNA, ions, and water molecules are charged, the presence of fewer ions and water molecules surrounding the DNA inside the CNT will improve the measured signal quality. Our calculations indicate that a (22,0) CNT can provide good results. If multiple DNA base pairs appear in the CNT at the same time, then the signals can be convoluted. To minimize the number of DNA base pairs inside the CNT, we shorten the length of the CNT to one unit-cell (4.26 Å), which is a little longer than one nucleotide unit (3.4 Å). This length is comparable to the length of some of the biological nanopores found in cell membranes. A self-consistent sp^3 TB method is employed to calculate the electrostatic potential distribution of DNA and other charged molecules inside the CNTs. More details on the TB method can be found in Ref. 15.

MD simulations are performed with GROMACS 3.2.1 using an AMBER99 force field.[22,23] Nonequilibrium MD simulations are performed for DNA translocation by mimicking atomic force microscope tips[22,24] to pull the DNA. The coordinates of water, ions, and DNA at an interval of 0.01 ps are used to compute the electrostatic Coulomb potential and the TB Hamiltonian. The TB equations are then solved self-consistently to obtain the potential distribution.[15] In Fig. 2, three MD snapshots illustrate the translocation of DNA through the semiconducting CNT from reservoir B to reservoir A. To obtain the electrostatic potential shown in Fig. 3, the average potential at the measuring point, referred to as the reference potential, is computed before the DNA enters the CNT. Next, once the first base of the DNA molecule enters the CNT, the potential at the measuring point, denoted as the perturbed potential, is recorded which varies as a function of the DNA position and time. Since the perturbed signals are too noisy because of the fluctuations in the MD trajectory, statistical moving-average technique is used to distill electrostatic signals, i.e., $\bar{V}(t)=1/(2N+1)\Sigma_{j=-N}^{j=N}V(t+j\times dt)$ where t is the time, dt is the time interval between sampling MD snapshots, which is chosen as 0.01 ps, N is the number of MD snapshots to do the moving average, V is the electrostatic signal obtained by using the self-consistent TB method, and $\bar{V}$ is the signal after using the moving average. $N=40$ is found to be reasonable to smooth out the position fluctuations of the MD simulations as well as to keep the average displacement small after every N steps compared to one nucleotide unit. The potential difference between the perturbed and the reference potential is denoted as the actual potential. The actual potential for the 12 nondefective base pair DNA case is shown in Fig. 3(a).

When the DNA bases are outside the CNT, the electrostatic potential generated by them is mostly screened by the electrolyte reservoirs. When the bases are inside the CNT, they contribute most to the electrostatic potential at the measuring point. Therefore, wavelike potential signals are obtained, as shown in Fig. 3(a), when the DNA base pairs enter the short CNT one by one. Since one cycle of peak and valley of the electrostatic signal represents a base pair passing through the CNT, by counting the number of cycles, we can estimate the total number of base pairs in the DNA segment. Figure 3(a) shows 12 cycles revealing that the DNA segment has 12 base pairs. Because the first and the last base pairs of Dickerson DNA have a total charge of $-1e$ each, while the interior base pairs have a total charge of $-2e$ each, the electrostatic potential at the measuring point for the first and the last cycles is only about half of the interior cycles. This is a potential approach to detect whether the DNA sequence is of Crick–Watson type (e.g., Dickerson).[17] To demonstrate the accuracy of the method, data from 30 MD simulations are collected and each MD simulation is performed with a different initial configuration. The estimates for the number of base pairs within the DNA are shown in Fig. 3(c). These estimates are obtained by counting the number of cycles with data from each MD simulation. The data from some MD simulations indicates a different number of cycles than what is expected and this is because of the translocation of Na^+ ions along with the DNA bases through the CNT. We did not observe translocation of any Cl^- ions along with the DNA bases through the CNT. The reason could be that negatively charged DNAs can easily attract positive ions and repel negative ions. When the DNA bases translocate through the CNT together with positive ions, they generate mixed electrostatic signals, introducing errors in the total number of cycles and the DNA bases. However, by using data

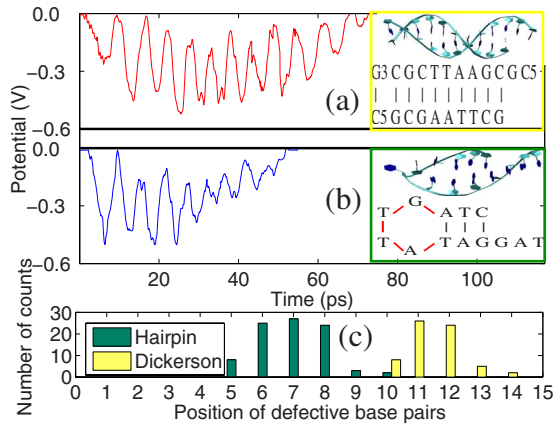

FIG. 4. (Color online) Variation of electrostatic potential as a function of time calculated at the measurement point (see Fig. 1) for the Dickerson DNA with two bases missing and the hairpin DNA with three bases missing. The insets in [(a) and (b)] show the defective Dickerson and the defective hairpin structures used in the MD simulation. (c) For the hairpin case, 25 of 30 MD simulations indicate that the sixth base pair is defective; 27 of 30 MD simulations indicate that the seventh base pair is defective; 24 of 30 MD simulations indicate that the eighth base pair is defective. The margin of error is within ±2 base pairs. For the Dickerson case, 26 of 30 MD simulations indicate that the 11th base pair is defective; 24 of 30 MD simulations indicate that the 12th base pair is defective. The margin of error is within ±2 base pairs.

from multiple MD simulations, the statistical results show that the number of base pairs can be estimated with reasonable accuracy.

Defective DNA base pairs have fewer charges than the normal DNA base pairs. Hence, they will generate smaller electrical signals compared to the normal DNA segments. For example, when a 12-segment Dickerson DNA with the last two bases missing translocates from reservoir B to reservoir A, the potential at the measuring point as a function of time is shown in Fig. 4(a). We observe ten normal cycles and two defective cycles. We again used data from 30 MD simulations each with a different initial configuration. The probability for the location of the defective base pair within the DNA is shown in Fig. 4(c). This is determined by counting the number of times where the defective cycles are found for the data from 30 MD simulations. The results show that the location of defective base pairs can be determined with reasonable accuracy.

Hairpin DNA structures are commonly used in experiments.[19] Single-strand DNA can form a hairpin structure, which can be very different from the standard helix structure.[10,17] An eight-segment hairpin DNA is also simulated in this work. The first few bases of the hairpin structure are usually unpaired, and every base has a charge of $-1e$, which is different from the Dickerson DNA charge pattern. The electrostatic potential for the first and last cycles are similar to the interior cycles for the hairpin structure as shown in Fig. 3(b). Note the clear difference from the Dickerson DNA results, which are shown in Fig. 3(a). Because of the charge and potential difference, it is possible to differentiate Dickerson and hairpin-type DNA structures by simply comparing the electrostatic potential signals from their first and last cycles. From the MD snapshots of Fig. 2, the second, fifth, and seventh base pairs of the hairpin DNA pass through the CNT at the times T_1, T_2, and T_3, respectively, and one can see that the times T_1, T_2, and T_3 in Fig. 3(b) approximately coincide with the second, fifth, and seventh valleys of the electrostatic potential signal. It is consistent with the physical observation that when one base-pair with net negative charge is completely located inside the CNT, the base-pair will generate the strongest potential signal at the measuring point.

When a hairpin structure with its last three base pairs defective translocates through the CNT, the relative position of the defective base pairs can be clearly read from the electrical signal patterns as shown in Fig. 4(b). With data from 30 MD simulations, the probabilities of the relative position of the three defective bases are as shown in Fig. 4(c). The results again show that the technique is an effective way to estimate the location of the defective base pairs. Finally, the theoretical results presented in this paper could pave the way toward developing next-generation nanosensors that could measure the lengths of various DNA strands and distinguish between different types of DNAs.

This research was supported by NSF under Grant Nos. 0120978, 0328162, 0810294, and 0852657.

[1] J. Kasianowicz, E. Brandin, D. Branton, and D. W. Deamer, Proc. Natl. Acad. Sci. U.S.A. **93**, 13770 (1996).
[2] J. Li, D. Stein, C. McMullan, D. Branton, M. J. Aziz, and J. A. Golovchenko, Nature (London) **412**, 166 (2001).
[3] J. B. Heng, A. Aksimentiev, C. Ho, P. Marks, Y. V. Grinkova, S. Sligar, K. Schulten, and G. Timp, Nano Lett. **5**, 1883 (2005).
[4] P. J. Goodfellow, B. Buttin, T. J. Herzog, J. S. Rader, R. K. Gibb, E. Swisher, K. Look, K. C. Walls, M. Y. Fan, and D. G. Mutch, Proc. Natl. Acad. Sci. U.S.A. **100**, 5908 (2003).
[5] H. Davies, G. R. Bignell, C. Cox, P. Stephens, S. Edkins, S. Clegg, J. Teague, H. Woffendin, M. J. Garnett, and W. Bottomley, Nature (London) **417**, 949 (2002).
[6] J. W. Lee and A. Meller, in *Perspective in Bioanalysis*, edited by K. Mitchelson (Elsevier Science, Amsterdam, 2007).
[7] S. Ghosh, A. K. Sood, and N. Kumar, Science **299**, 1042 (2003).
[8] M. Majumder, N. Chopra, R. Andrews, and B. J. Hinds, Nature (London) **438**, 44 (2005); J. K. Holt, H. G. Park, Y. Wang, M. Stadermann, A. B. Artyukhin, C. P. Grigoropoulos, A. Noy, and O. Bakajin, Science **312**, 1034 (2006).
[9] G. Hummer, J. C. Rasaiah, and J. P. Noworyta, Nature (London) **414**, 188 (2001); S. Joseph and N. R. Aluru, Nano Lett. **8**, 452 (2008).
[10] H. Gao and Y. Kong, Annu. Rev. Mater. Res. **34**, 123 (2004); Y. Xie, Y. Kong, A. K. Soh, and H. Gao, J. Chem. Phys. **127**, 225101 (2007).
[11] R. H. Baughman, A. A. Zakhidov, and W. A. de Heer, Science **297**, 787 (2002).
[12] L. Dai, *Carbon Nanotechnology: Recent Developments in Chemistry, Physics, Materials Science and Device Applications* (Elsevier Science, Amsterdam, 2006).
[13] M. Zwolak and M. D. Ventra, Rev. Mod. Phys. **80**, 141 (2008).
[14] Y. Xu and N. R. Aluru, Appl. Phys. Lett. **93**, 043122 (2008).
[15] Y. Xu and N. R. Aluru, Phys. Rev. B **76**, 075304 (2007); **77**, 075313 (2008).
[16] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[17] H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson, Proc. Natl. Acad. Sci. U.S.A. **78**, 2179 (1981).
[18] X. Zhao and J. K. Johnson, J. Am. Chem. Soc. **129**, 10438 (2007).
[19] B. McNally, M. Wanunu, and A. Meller, Nano Lett. **8**, 3418 (2008).
[20] R. J. Mashl, S. Joseph, N. R. Aluru, and E. Jakobsson, Nano Lett. **3**, 589 (2003).
[21] D. Lu, Y. Li, S. V. Rotkin, U. Ravaioli, and K. Schulten, Nano Lett. **4**, 2383 (2004).
[22] W. F. Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi, *Biomolecular Simulation: The GROMOS96 Manual and User Guide* (Hochschulverlag, ETH Zurich, Switzerland, 1996).
[23] E. J. Sorin and V. S. Pande, Biophys. J. **88**, 2472 (2005).
[24] D. Branton, D. W. Deamer, A. Marziali, H. Bayley, S. A. Benner, T. Butler, M. D. Ventra, S. Garaj, and A. Hibbs, Nat. Biotechnol. **26**, 1146 (2008).