## Nanopore sequencing of polynucleotides assisted by a rotating electric field

C.-M. Chen<sup>a)</sup> and E.-H. Peng

Physics Department, National Taiwan Normal University, Taipei, Taiwan, Republic of China

(Received 14 June 2002; accepted 2 January 2003)

The translocation kinetics of a polynucleotide chain through a nanopore is studied using Monte-Carlo simulations for both lattice and off-lattice models, by which we demonstrate a method in sequencing polynucleotides assisted by a rotating electric field. At low frequencies of the rotating field, the translocation time of the chain is inversely proportional to the rotating field frequency. More specifically, in these cases, the translocation time of each nucleotide is nearly quantized, and thus it greatly improves the resolution of blockade-current time series obtained from nanopore sequencing experiments. The polynucleotide sequences can be accurately determined from analyzing several time series of current variation through the nanopore. © 2003 American Institute of Physics. [DOI: 10.1063/1.1554480]

Developing methods and instruments that permit fast polynucleotide sequencing has attracted considerable attention recently.<sup>1-4</sup> A device for rapid DNA decryption will allow quick identification of pathogens to save many lives during an epidemic or a bioterrorist attack. Moreover, doctors would be able to diagnose a disease, judge possible risks, and design a special treatment plan based on knowledge about which disease-related genes a patient carries. Although DNA sequencing has important medical applications, present methods in sequencing polynucleotides are slow, costly, and inaccurate.<sup>5-7</sup> A comparison of the Celera and Ensembl predicted gene sets has revealed little (20%) overlap in predicted genes.<sup>8</sup> Therefore, it is highly desired to develop accurate sequencing methods that are fast and inexpensive. An interesting idea in sequencing DNA is to monitor the variation of ionic current due to an applied electric field which drives single-stranded polynucleotides through a nanopore in a thin film.<sup>1,2,9</sup> Preliminary results of this method have shown its capability to distinguish long stretches of the same nucleotides, such as 30 adenines followed by 70 cytosines. Nevertheless, to make this method work for sequencing real DNA at single-nucleotide resolution, the translocation time of each nucleotide should be increased by a factor of 1000 and thermal fluctuations should be suppressed to allow sequence analysis. Based on computer simulations, we report a method to resolve the above difficulties in nanopore sequencing of polynucleotides by adding a rotating electric field above the thin film. With an adequate rotating field, the translocation time of each nucleotide is found to be about  $mT_c/4$ , where m is an integer and  $T_c$  is the period of the rotating field. The resolution of nanopore sequencing can be greatly improved by controlling the frequency of the rotating field. High accuracy in predicting the sequence of polynucleotides can be achieved by analyzing several time series of current variation.

In our simulations,<sup>10</sup> a single strand polynucleotide of length N is represented by the bond-fluctuation model, and

its motion is simulated by the Metropolis Monte-Carlo (MC) algorithm in a cubic lattice at a constant temperature T. This model has been shown to be a realistic and efficient method for studying polymer dynamics in various systems.<sup>11-15</sup> Simulations using an off-lattice bead-spring model have also been carried out to eliminate possible lattice effects. During our simulations, the polynucleotide is driven through a nanopore of size D in a thin film by an applied electric field with a constant amplitude E in the z-direction. Above the thin film, a rotating electric field  $\mathbf{E}_c = E_c \sin(\omega t)\hat{i} + E_c \cos(\omega t)\hat{j}$  is added to control the translocation of the polynucleotide, where  $\hat{i}$  and  $\hat{j}$  are unit vectors in the x- and y-directions. At each instant, a nucleotide is picked up at random and attempts to move in any of the six directions by one lattice spacing. If any attempted move of nucleotides satisfies the excluded volume constraint and the new bond vectors are still in the allowed set, the move is accepted with probability  $p = \min[1, \exp(-\Delta U/kT)]$ , where  $\Delta U$  is the energy change of the chain and kT is thermal energy. The energy of polynucleotides is expressed as  $U = U_{bend} + U_{electric} + U_{H-bond}$ , where  $U_{\text{bend}} = \sum_{i=1}^{N-2} e(1 - \cos \theta_i)$  is the bending energy of the chain with rigidity e and bending angles  $\{\theta_i\}, {}^{11-13} U_{\text{electric}}$  is the electric potential energy due to the above two electric fields, and  $U_{\text{H-bond}}$  is the hydrogen bonding energy of (A, T) and (G, C) pairs. Here, we consider the special case of negligible hydrogen bonding between bases, which can be realized by adjusting the pH value, raising the temperature, or adding ureas. The effect of hydrogen bonding on the translocation of polynucleotides will be discussed elsewhere.

To study the kinetics of a polynucleotide chain passing a nanopore, we have simulated its translocation process 50 times for each set of parameters. For a chain of 50 nucleotides, we choose the pore size D=3, temperature T=1, the constant electric field amplitude E=1.5, and the bending rigidity e=0.2. Here, thermal energy and electric charge of each nucleotide have been set to unity, and the corresponding electric field is of order  $10^7$  V/m. For the rotating field, its frequency  $\omega$  varies from  $10^{-1}$  to  $10^{-8}$  (MC step<sup>-1</sup>) and its amplitude  $E_c$  varies from 0.1 to 1.2. Figure 1 shows the number of nucleotides that have passed the pore in some

1308

<sup>&</sup>lt;sup>a)</sup>Electronic mail: cchen@phy.ntnu.edu.tw

<sup>© 2003</sup> American Institute of Physics



FIG. 1. Translocation processes of polynucleotides in two different regimes. (a) At high frequencies ( $\omega \ge 10^{-3}$ ),  $t_c (\approx 2 \times 10^4 \text{ MC steps})$  is independent of  $\omega$ . (b) At low frequencies ( $\omega \le 10^{-4}$ ),  $t_c$  is well controlled by varying  $\omega$ . Here,  $E_{\rm c}/E = 0.47$ .

typical time series of the translocation process. At high frequencies ( $\omega \ge 10^{-3}$ ), the polynucleotide passes through the pore smoothly and the translocation time of the whole chain,  $(t_c)$  is about a constant  $(t_c \approx 2 \times 10^4 \text{ MC steps})$ . In these cases, the translocation time of each nucleotide,  $(t_n)$  does not vary drastically. At low frequencies ( $\omega \le 10^{-4}$ ), two kinds of translocation kinetics are observed. For nucleotides locating at the middle,  $t_n$  is much longer than that of nucleotides near both ends. At  $\omega = 10^{-6}$ ,  $t_c$  is about  $10^8$  MC steps. It has been estimated previously that  $t_n$  is about 1  $\mu$ s at the strength of our driving field for a smooth translocation, and thus 1 MC step in our simulations is on the order of  $10^{-8}$  s.<sup>16</sup> We conclude that the frequency of the rotating field should be less than 10<sup>4</sup> Hz in order to slow down the translocation process. A detailed study reveals that  $t_n$  $\simeq mT_c/4$  for  $\omega \le 10^{-4}$  and  $E_c/E > 0.4$ , where *m* is an integer and  $T_c = 2 \pi / \omega$ . This effect of quantized  $t_n$  can be understood from the simulation in Ref. 10, which is because the aligned chain cannot move with the rotating field in our lattice model. Although this might result partly from lattice effects, the vanishing stretching force can be easily realized in experiments by switching off the rotating field briefly. As shown in the inset of Fig. 2, the quantization of  $T_n$  is also seen in off-lattice simulations using a bead-spring model, in which the rotating field is switched off for less than 0.02  $T_c$ every  $T_c/4$ . Figure 2 shows that  $t_c$  is inversely proportional to  $\omega$  for  $\omega \le 10^{-4}$ , and is almost a constant for  $\omega \ge 10^{-3}$ . If the response of the chain is faster than the rotating field, translocation of nucleotides cannot occur when the chain is stretched. If the response of the chain is too slow, it penetrates the pore smoothly. The boundary between these two regimes depends on the response of polynucleotides and can be varied by changing the viscosity of the solution or the friction of the thin film surface. Similar results have also been obtained for N = 30, 80, and 100.

То sequence polynucleotides by time series analysis of current variation, we consider a chain consisting of 26 randomly generated sequence Downloaded 20 Feb 2003 to 140.122.141.159. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/aplo/aplcr.jsp



FIG. 2. The dependence of translocation time of the chain  $(t_c)$  on the frequency of the rotating field.  $E_c/E$  is 0.47 (circle) and 0.8 (square), respectively. The slope is -1 for both sets of data. The inset shows a typical time series of  $t_n$  in off-lattice simulations.

(GTACTTCGCGTGTAGTCATTTAATCC) located at the middle, and two extra fragments AAAAAAAAAAA and ACCCCCCCCC, attached at the 3' and 5' ends, respectively. These two fragments are added because nucleotides near the ends pass the nanopore quickly and cannot be sequenced, as indicated from Fig. 1(b). In addition, they can be used to locate the main sequence; an initial increase of the current signals the beginning of the main sequence from the 3' end, while an initial drop of the current signals the beginning of the sequence from the 5' end. The change in current of the time series shows the change of nucleotides in the sequence. Many sets of time series of current variation are recorded for the same chain. Since  $T_n$  is quantized, the shortest duration of each segment of blockade current in these sets of data reveals the number of repeat nucleotides. As shown in Fig. 3, the predicting error of the random sequence from a single time series is about 30%. By analyzing more than 16 sets of time series, the sequencing error is reduced to zero. It is evident that the error decreases rapidly as the number of time series increases.

In this letter, we have demonstrated a simple and powerful method to control the translocation processes of polynucleotides through a nanopore assisted by a rotating electric field. Our method can be easily implemented in a nanopore sequencing experiment by adding two pairs of parallel electrodes above the thin film. Moreover, a recent development



FIG. 3. Prediction error of a polynucleotide sequence from analyzing several time series. Here,  $\omega = 10^{-4}$  and  $E_c/E = 0.47$ .

in fabricating nanopores in a  $Si_3N_4$  thin film by ion-beam sculpting<sup>1</sup> allows us to design an array of sequencing cells that can be used to boost sequencing speed without sacrificing sequencing resolution. If this method is implemented, our analysis shows that high accuracy in predicting polynucleotide sequences can be achieved when many sets of time series are analyzed.

We thank T. K. Lee and K. Dill for stimulating discussions. This work is supported, in part, by the National Science Council of Taiwan under Grant No. NSC 91-2112-M-003-023 and by National Taiwan Normal University under Grant No. ORD92-3.

<sup>3</sup>B. Essevaz-Roulet, U. Bockelmann, and F. Heslot, Proc. Natl. Acad. Sci. U.S.A. **94**, 11935 (1997).

- <sup>4</sup>M. Sauer, J. Ardenjacob, K. H. Drexhage, F. Gobel, U. Lieberwirth, K. Muhlegger, R. Muller, J. Wolfrum, and C. Zander, Bioimaging 6, 14 (1998).
- <sup>5</sup>L. M. Smith, J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. H. Kent, and L. E. Hood, Nature (London) **321**, 674 (1986).
- <sup>6</sup>F. Sanger, S. Nicklen, and A. R. Coulson, Proc. Natl. Acad. Sci. U.S.A. **74**, 5463 (1977).
- <sup>7</sup>For a review, see A. Marziali and M. Akeson, Annu. Rev. Biomed. Eng. **3**, 195 (2001).
- <sup>8</sup>J. B. Hogenesch, K. A. Ching, S. Batalov, A. I. Su, J. R. Walker, Y. Y.
- Zhou, S. A. Kay, P. G. Schultz, and M. P. Cooke, Cell **106**, 413 (2001). <sup>9</sup>M. Akeson, D. Branton, J. J. Kasianowicz, E. Brandin, and D. W. Deamer, Biophys. J. **77**, 3227 (2001).
- <sup>10</sup> For more details, see [http://www.phy.ntnu.edu.tw/~cchen/dna/ index.htm].
- <sup>11</sup>C.-M. Chen, Phys. Rev. E 63, 010901 (2001).
- <sup>12</sup>C.-M. Chen and Y.-A. Fwu, Phys. Rev. E 63, 011506 (2001).
- <sup>13</sup>C.-M. Chen and P. G. Higgs, J. Chem. Phys. 108, 4305 (1998).
- <sup>14</sup>I. Gerroff, A. Milchev, K. Binder, and W. Paul, J. Chem. Phys. **98**, 6526 (1993).
- <sup>15</sup>H. P. Deutsch and K. Binder, J. Chem. Phys. **94**, 2294 (1991).
- <sup>16</sup>P. G. de Gennes, Proc. Natl. Acad. Sci. U.S.A. 96, 7262 (1999).

<sup>&</sup>lt;sup>1</sup>J. Li, D. Stein, C. Mcmullan, D. Branton, M. J. Aziz, and J. A. Golovchenko, Nature (London) **412**, 166 (2001).

<sup>&</sup>lt;sup>2</sup>A. Meller, L. Nivon, E. Brandin, J. Golovchenko, and D. Branton, Proc. Natl. Acad. Sci. U.S.A. **97**, 1079 (2000).