Chinese Chemical Letters 25 (2014) 1029–1032

Contents lists available at ScienceDirect

Chinese Chemical Letters

journal homepage: www.elsevier.com/locate/cclet

Original article

Exponentially modified Gaussian relevance to the distributions of translocation events in nanopore-based single molecule detection

Zhen Gu [a], Yi-Lun Ying [b], Bing-Yong Yan [c,*], Hui-Feng Wang [c,*], Pin-Gang He [a,*], Yi-Tao Long [b]

[a] Department of Chemistry, East China Normal University, Shanghai 200241, China
[b] Key Laboratory for Advanced Materials & Department of Chemistry, East China University of Science and Technology, Shanghai 200237, China
[c] School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China

ARTICLE INFO

Article history:
Received 20 February 2014
Received in revised form 14 April 2014
Accepted 23 April 2014
Available online 13 May 2014

Keywords:
Nanopore
Single-molecule detection
Exponentially modified Gaussian

ABSTRACT

Nanopore technique plays an important role in single molecule detection, which illuminates the properties of an individual molecule by analyzing the blockage durations and currents. However, the traditional exponential function is lack of efficiency to describe the distributions of blockage durations in nanopore experiments. Herein, we introduced an exponentially modified Gaussian (EMG) function to fit the duration histograms of both simulated events and experimental events. In comparison with the traditional exponential function, our results demonstrated that the EMG provides a better fit while covers the entire range of the distributions. In particular, the fitted parameters of EMG could be directly used to discriminate the sequence length of the oligonucleotides at single molecule level.

1. Introduction

The nanopore has become a unique tool for single molecule detection with the advantages of label free and high-throughput sensing [1,2]. It has been successfully used to explore a large variety of biomolecules including oligonucleotides [3–7], peptides [11–13], proteins [8–10], and protein–DNA complexes [14,15]. Previous studies have showed that a biological nanopore is an ultra-sensitive sensor for monitoring the conformational changes of biomolecule which is induced by the biological weak interactions [16–20]. In particular, nanopore technique offers the prospect of sequencing a human genome at the expense of ~$1000 within 24 h [21–23]. Generally, in a nanopore experiment, an individual molecule is driven into a nanopore under a biased voltage in the electrolyte solution. When the molecule transverses through the nanopore, it will produce a characteristic ionic blockage. After statistical analysis of currents and durations of the blockages, the properties of an individual molecule, such as the composition, length and secondary structure, could be elucidated. During the translocation process of biomolecules at a certain electrophoretic force, the blockage duration not only depends on the size of the biomolecule but also the interactions between nanopore and the analyte. Therefore, the histogram for the blockage duration appears to be a half Gaussian and half exponential distribution [24]. The curve follows a relatively steep rise and fall, but the blockage durations larger than the value of Gaussian peak is approximated by an exponential decay. In a conventional data analysis process, the distributions of durations are fitted by exponential functions. However, the exponential fittings barely cover the whole distribution of durations, which exclude all of the events with the durations shorter than the value of Gaussian peak. The inappropriateness of exponential distributions only reveals the random walks of the biomolecules translocation rather than other deterministic phenomena. Therefore, the exponential function is insufficient to model the blockage durations of translocation events in nanopore analysis.

Exponentially modified Gaussian function could be used to describe process which begins with Gaussian distribution and ends with exponential distribution [25]. It has been widely applied to analyze peaks in chromatography [26], model the distributions of intermitotic time and extracts variabilities of protein expression in biology [27]. Herein, EMG was introduced into analyzing the blockage durations for nanopore-based single molecule detection. The EMG was compared to conventional methods and validated by analyzing the stimulated distributions as well as the histograms for

* Corresponding author.
E-mail addresses: byyan@ecust.edu.cn (B.-Y. Yan), whuifeng@ecust.edu.cn (H.-F. Wang), pghe@chem.ecnu.edu.cn (P.-G. He).

http://dx.doi.org/10.1016/j.cclet.2014.05.009

oligonucleotides translocations by nanopore experiments. Our results demonstrated that, compared with traditional exponential function, EMG provided a better fit and covered the entire range of the distributions. In particular, the fitted parameters of EMG could be directly used to discriminate the sequence length of oligonucleotides at single molecule level.

2. Experimental

α-Hemolysin (α-HL) was purchased from Sigma–Aldrich (St. Louis, MO, USA) and used without purification. Diphytanoyl-phosphatidyl-choline was purchased from Avanti Polar Lipids Inc. (Alabaster, AL, USA). Poly(dA)$_{60}$ (oligo **1**) and the sequence 5′-ACCTGGGGGAGTATTGTAAAAAAGAATGTCGCAAAAAAAAAAA-3′ (oligo **2**) were synthesized by Invitrogen Life Technologies (Shanghai, China). The final concentration of the analyte in 1 mL *cis* chamber was 0.15 μmol/L. Prior to use, DNA solutions were annealed at 95 °C for 10 min and then cooled to room temperature. Millipore water (18 MΩ cm^{-1}) was used throughout the experiments. Unless otherwise noted, all other chemicals were of analytical grade.

The nanopore electrical recording was conducted according to our previous studies [15–17,28,29]. The lipid bilayers were created by applying diphytanoyl-phosphatidyl-choline (30 mg/mL) in decane (≥99%, Sigma–Aldrich, St. Louis, MO, USA) to a 150 μm orifice in a 1 mL bilayer cup integrated into a lipid bilayer chamber (Warner Instruments, Hamden, CT, USA) filled with KCl (1.0 mol/L) and Tris–HCl (10 mmol/L, pH 7.8). Temperature was fixed at 25 ± 0.5 °C by amounting the lipid bilayer chamber on a thermal stage (Dagan Corporation, Minneapolis, MN, USA). The α-HL was injected adjacent to the aperture in the *cis* chamber, and pore insertion was determined by a well-defined jump in current value. Then, oligo **1** and oligo **2** were added in the *cis* solution under a voltage of +100 mV, respectively.

The ionic currents were acquired using a ChemClamp amplifier (Dagan Corporation, Minneapolis, MN, USA), filtered with a 3-pole low-pass Bessel filter at 3 kHz and sampled at 100 kHz by DigiData 1440A/D converter (Axon Instruments, Forest City, CA, USA). The ionic currents were recorded by a PC running PClamp 10.2 (Axon Instruments, Forest City, CA, USA). The recorded current data were analyzed by a home-made matlab program.

Theoretical simulation: A one-dimensional lattice Monte Carlo algorithm was used to simulate the translocation process of single-strand DNA through nanopore whose cross-section was close to the pore size of α-HL. The detailed simulation process was described in a previous study [30]. The blockage durations of each simulation were recorded and sorted into histograms.

3. Results and discussion

As a probability distribution, the EMG function is the combination of normal and exponential random variables. The equivalent form of the EMG could be written as follows:

$$P(t;t_c,\omega,t_s)=\frac{1}{t_s}\exp\left(\frac{1}{2}\left(\frac{\omega}{t_s}\right)^2-\frac{t-t_c}{t_s}\right)\left(\frac{1}{2}+\frac{1}{2}erf\left(\frac{z}{\sqrt{2}}\right)\right) \quad (1)$$

where t_c determines the position of peak, ω is the width of peak, t_s is the modification factor (skewness), and $erf(z/\sqrt{2})$ is the error function evaluated at $(z/\sqrt{2})$, in which $z=((t-t_c)/\omega)-(\omega/t_s)$.

As shown in Fig. 1, we generated examples of EMG function through a Matlab program, with appropriate domain ranging from 0 to 10. Obviously, the EMG showed a characteristic positive skew, which is consistent with the histogram of duration of translocation events in nanopore. In addition, position, width and skew of the peak are tunable by adjusting the parameters ω, t_c and t_s. Thus, EMG is able to cover the whole distribution of blockage durations for nanopore experiments.

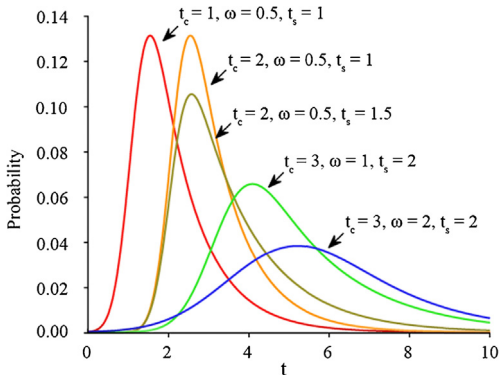

Fig. 1. The three parameters t_c, t_s, and ω could influence the shape of EMG in peak position, width and skewness.

Since the shape of the duration histogram originates from the translocation behavior of each oligonucleotide, it cannot be known ab initio. Here, we use the Monte Carlo simulation to generate the translocation events of oligonucleotides for the purpose of validating the introduced EMG. In the Monte Carlo simulation, the nanopore was simplified to a one-dimension channel as long as 40 bases. The oligonucleotides were considered as ideal, self-avoiding, and rod-like chains with lattice coordination number set to 3. The simulation was carried out at atmospheric temperature of 288.15 K and external electric field was 2. Chains with length from 20 to 120 bases were applied to simulate the translocation processes, respectively. For each chain, we generated 100,000 events for the substantial statistic analysis. Then, the histograms of duration time were fitted by both EMG and exponential function. The expression of the exponential function applied in this paper was $f(x)=\lambda_1\exp(-\lambda_2 x)$, where λ_1 and λ_2 were both parameters of the function. The bin-width of the duration histograms was set to 0.01 ms, while the histograms were normalized by dividing the count number of each bin by the maximum count number among all bins. Fig. 2 illustrates the EMG fitting examples of oligonucleotides with 20 and 120 bases, respectively. The parameters (t_c, t_s, ω) were obtained by fitting the EMG function to the histogram through trust region algorithm in Matlab program. The commonly used goodness of fit criteria (R-squared, R^2) was applied to check the fitting performances. Compared with the traditional exponential fittings, the values of R^2 of all fitted results carried out by EMG were much higher (Table 1). Therefore, the histograms of blockage durations were well fitted by the EMG. The fitted value t_c, which

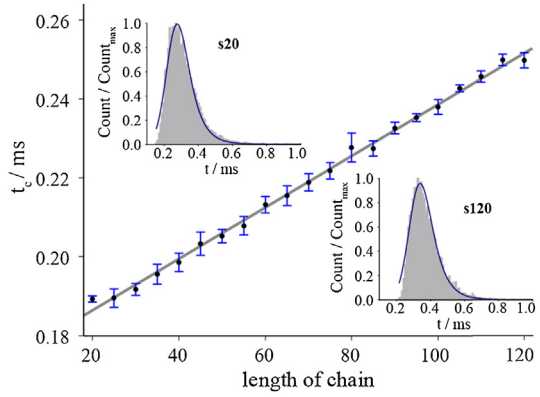

Fig. 2. The values of t_c as a function of chain length ranging from 20 to 120 bases. Insertion: Duration histogram of the simulated events of **s20** and **s120**, which were fitted by EMG distribution respectively.