

# Correctly validating results from single molecule data: The case of stretched exponential decay in the catalytic activity of single lipase B molecules

Ophir Flomenbom<sup>a,\*</sup>, Johan Hofkens<sup>b,\*</sup>, Kelly Velonia<sup>c</sup>, Frans C. de Schryver<sup>b</sup>, Alan E. Rowan<sup>d,\*</sup>, Roeland J.M. Nolte<sup>d</sup>, Joseph Klafter<sup>e</sup>, Robert J. Silbey<sup>a</sup>

<sup>a</sup> Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Av., Cambridge, MA 02139, United States

<sup>b</sup> Department of Chemistry, Katholieke Universiteit Leuven, Celestijnenlaan 200 F, 3001 Heverlee, Belgium

<sup>c</sup> Department of Organic Chemistry, University of Geneva, 30 quai Ernest Ansermet, Geneva 4, CH-1211, Switzerland

<sup>d</sup> Department of Organic Chemistry, Radboud University Nijmegen, 6525 ED Nijmegen, The Netherlands

<sup>e</sup> School of Chemistry, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel

Received 11 October 2006

Available online 20 October 2006

## Abstract

The question of how to validate and interpret correctly the waiting time probability density functions (WT-PDFs) from single molecule data is addressed. It is shown by simulation that when a stretched exponential WT-PDF,  $\phi_{\text{off}}(t) = \phi_0 e^{-(t/\tau)^\alpha}$ , generates the off periods of a two-state trajectory, a reliable recovery of the input  $\phi_{\text{off}}(t)$  from the trajectory is obtained even when the bin size used to define the trajectory,  $dt$ , is much larger than the parameter  $\tau$ . This holds true as long as the first moment of the WT-PDF is much larger than  $dt$ . Our results validate the results in an earlier study of the activity of single lipase B molecules and disprove recent related critique.

© 2006 Elsevier B.V. All rights reserved.

## 1. Introduction

Advances in single molecule experimental techniques allow studying at room temperature the dynamics of a large number of processes in real time. Examples include the flux of ions through individual channels [1], conformational fluctuations of biopolymers [2–5], and kinetics of single enzymes [6–9]. The aim of such sophisticated measurements is to learn about the underlying mechanism of the process to an extent that is unattainable from bulk measurements due to averaging.

In the simple scenario, the first step in the analysis of the data turns it into a trajectory of *on* and *off* periods (waiting times), Fig. 1a. A frequently used assumption describes the mechanism of the observed process by a multi-substate *on*–*off* Markovian kinetic scheme (KS) [10–17], Fig. 1b; see also [18–21] for other model types. Using statistical tools of data analysis [22–26], one wishes to learn as much as possible about the underlying KS.

It was recently shown in [10], that the way to utilize the information content in the two-state trajectory is to construct from it a canonical form of reduced dimensions. The mathematical mapping from the space of KSs into the space of reduced dimension forms enables finding, in the most efficient way, the relationship between the data and all possible KSs that can generate it. Technically, the analysis uses, along with other methods, the two dimensional histograms of successive events.

\* Corresponding authors. Fax: +1 617 253 7030 (O. Flomenbom), fax: +32 16 327989 (J. Hofkens), fax: +31 24 365 2929 (A.E. Rowan).

E-mail addresses: [flomenbo@mit.edu](mailto:flomenbo@mit.edu) (O. Flomenbom), [johan.hofkens@chem.kuleuven.be](mailto:johan.hofkens@chem.kuleuven.be) (J. Hofkens), [A.Rowan@science.ru.nl](mailto:A.Rowan@science.ru.nl) (A.E. Rowan).

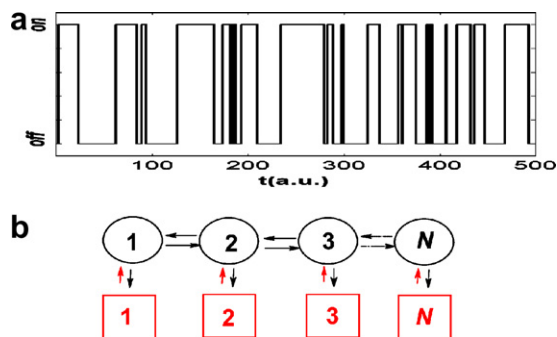


Fig. 1. (a) A trajectory of an observable that fluctuates between two values, *on* and *off*, as a function of time. Such a two-state trajectory can be generated from kinetic Monte-Carlo simulations, which model random walks on kinetic schemes, or on reduced dimension forms [10]. (b) An example for a KS with only reversible transitions. The *off*-state consists of a spectrum of coupled *off* substates (black circles), whereas the *on* substates (red squares) are uncoupled. Each *on* substate is connected to a single *off* substate, and vice versa. This KS was used in [8,9] as a model for the enzymatic activity of lipase B from *Candida Antartica*.

However, there are cases in which the experimental two-dimensional histograms are too noisy due to the limited number of events in the data, and other methods of analysis are used to detect correlations between successive periods. This was the case in the study of the activity of single lipase B molecules from *Candida antartica* [8,9]. The two-state trajectories contained approximately  $10^4$ -cycles. The bin size of the trajectories was 1 ms. The first stage in the analysis of the two-state trajectories constructed the WT-PDFs of the *on* and *off* periods. (This way of analysis is the most basic one when analyzing two-state trajectories, and must be accompanied by methods that look for correlations between successive events [8,9].) The *on* WT-PDF was approximated by an exponential decay with a fast decaying rate of the order of the bin size, which reflects short *on* periods. In contrast, the *off* periods were, on average, two orders of magnitude larger than the bin size. The experimental histograms for the *off* periods, at saturating substrate concentrations, were fitted by a stretched exponential,  $\phi_{\text{off}}(t) = \phi_0 e^{-(t/\tau)^\alpha}$ , with the parameters  $\alpha = 0.15$  and  $\tau = 0.00115$  ms, independent of the substrate concentration. The limited number of events in the trajectory dictates analyzing the *off* ordered waiting times trajectory, which is the trajectory of the *off* waiting times plotted vertically as a function of the occurrence index. This trajectory revealed correlations between successive *off* periods. Combining these two features, along with other experimental findings, led to a mechanism that involves simultaneously conformational changes and enzymatic activity (Fig. 1b). Importantly, the mechanism strongly deviates from the Michaelis–Menten mechanism that has been used to model enzymatic activity for almost a century, see for example [27].

A recent letter [28], incorrectly argues against the interpretation of the experimental *off* histogram found in Refs. [8,9] as a stretched exponential *off* WT-PDF. The critique in Ref. [28] wrongly concludes that the values for  $\alpha$  and  $\tau$

found for the experimentally stretched exponential *off* histogram cannot be extracted from a two-state trajectory with a bin size of 1 ms whose *off* periods were generated from such a WT-PDF. By means of a simulation, it is shown that a stretched exponential WT-PDF with the parameters  $\alpha = 0.15$  and  $\tau = 0.00115$  ms is reliably recovered from a  $10^4$ -cycle two-state trajectory with a bin size of 1 ms, whose *off* periods are generated by this WT-PDF. This result holds true as long as the first moment of the WT-PDF is still much larger than the bin size. We also point out on the technical error that led to the wrong conclusion in [28]; the incorrect choice of the prefactor in the plot of the experimental WT-PDFs led to the wrong conclusion in [28].

## 2. How to interpret experimental results correctly?

### 2.1. Moments analysis

Consider the normalized WT-PDF,

$$\begin{aligned} \phi_{\text{off}}(t) &= \phi_0 e^{-(t/\tau)^\alpha} \equiv \phi_0 \psi_{\text{off}}(t); \\ \phi_0 &= 1 / \int_0^\infty \psi_{\text{off}}(t) dt = \frac{\alpha/\tau}{\Gamma(1/\alpha)}. \end{aligned} \quad (1)$$

The  $n$ th moment of the WT-PDF in Eq. (1) is given by

$$\langle t^n \rangle = \int_0^\infty t^n \phi_{\text{off}}(t) dt = \tau^n \frac{\Gamma[(n+1)/\alpha]}{\Gamma(1/\alpha)},$$

where  $\Gamma(\cdot)$  is the gamma function.

In particular, the first moment,  $n = 1$ , with,

$$\alpha = 0.15, \quad \tau = 0.00115 \text{ ms}, \quad (2)$$

has the value of,

$$\langle t \rangle = 3300 \text{ ms}.$$

Although the functional form for the *off* WT-PDF in [8,9] was found to follow Eq. (1) with the parameters given by Eq. (2), the value of the average *off* waiting times found directly from the data was only 220 ms. The reason for this discrepancy is very simple; the largest random *off* period observed in the experiment was about 1000 ms. Accordingly, when calculating the moments (and the normalization) of the WT-PDF found from the experiment using Eq. (1), one should take into consideration the time window relevant to the experiment. With a cut-off of 1000 ms, the numerically obtained first moment is consistent with the experimental value:

$$\langle t_{<1000 \text{ ms}} \rangle = 210 \text{ ms}.$$

The above discussion emphasizes the subtleness of testing for self-consistency; one must correctly set the time window of the integration used in calculating the moment and the normalization of the WT-PDF, whose temporal decay pattern is found from the experiment, when comparing it with the average of the *off* periods found directly from the trajectory.

## 2.2. Plotting analytical WT-PDFs and experimentally obtained WT-PDFs

Suppose that the analytical WT-PDF in Eq. (1) with the parameter values in Eq. (2) is plotted with  $dt = 1$  ms. In a plot of the ratio  $\phi_{\text{off}}(t)/\phi_{\text{off}}(0)$ , the normalization constant vanishes, and we end up plotting  $\psi_{\text{off}}(t)$  (Fig. 2a). Thus, a simple manipulation of the axes immediately leads to a straight line for the above ratio. Namely, when plotting  $f(\phi_{\text{off}}(t)/\phi_{\text{off}}(0))$  as a function of  $\log t$ , where  $f(\cdot)$  stands for a  $-\log[-\log(\cdot)]$  and  $\log(\cdot)$  is a natural  $\log(\cdot)$ , we get a straight line for a normalized stretched exponential  $\phi_{\text{off}}(t)$  (Fig. 2b). From this plot the extraction of the parameters  $\alpha$  and  $\tau$  is straightforward. Now, consider an experimental WT-PDFs constructed as a function of discrete time,  $t = jdt, j = 1, 2, \dots, N$ , where  $Ndt$  is the largest relevant period. Note that the first value of  $t$  equals to the bin size  $dt$ , which is larger than zero. So, if one looks at the ratio  $\phi_{\text{off}}(t)/\phi_{\text{off}}(dt)$ , which is equal to  $\psi_{\text{off}}(t)/\psi_{\text{off}}(dt)$ , and plots  $f(\phi_{\text{off}}(t)/\phi_{\text{off}}(dt))$  versus  $\log t$ , one clearly does not get a straight line (Fig. 2c–d). The deviation of this ratio from a straight line does not mean that  $\psi_{\text{off}}(t)$  obeys a decay pattern different than a stretched exponential. It simply means that  $\psi_{\text{off}}(dt) \neq 1$ . Eq. (1) implies,  $\psi_{\text{off}}(dt) \leq 1$ , where the equality holds only in the limit  $dt \rightarrow 0$ , but in experimental conditions  $dt$  has a finite value. If one builds a histogram from an experiment as

$$\tilde{g}_{\text{exp}}(t) = \# \text{ of periods between } t - dt \text{ and } t / \# \text{ of periods,} \\ j = 1, 2, \dots, N,$$

(where as in [8,9], only the right edge of the  $j$  bin, among its two edges, contributes to  $\tilde{g}_{\text{exp}}(t)$  and the histogram bin size is taken to be equal to the trajectory bin size) and further looks at the ratio,

$$g_{\text{exp}}(t) = \tilde{g}_{\text{exp}}(t) / \tilde{g}_{\text{exp}}(dt),$$

one ends up looking at (or estimating),

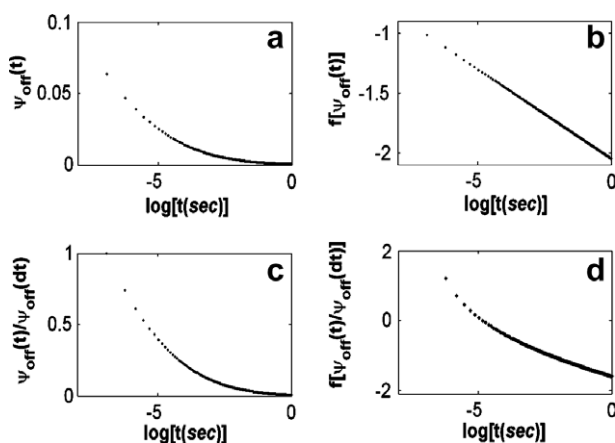


Fig. 2.  $\psi_{\text{off}}(t)$  (a),  $f[\psi_{\text{off}}(t)]$  (b),  $\psi_{\text{off}}(t)/\psi_{\text{off}}(dt)$  (c) and  $f[\psi_{\text{off}}(t)/\psi_{\text{off}}(dt)]$  (d) as a function of  $\log(t)$ . Although  $f[\psi_{\text{off}}(t)]$  follows a straight line when plotted against  $\log(t)$  (b), the ratio  $f[\psi_{\text{off}}(t)/\psi_{\text{off}}(dt)]$  does not (d). This can explain the results in [28], where Fig. 1 in [28] plots curves with similar prefactors as in (d). Here, and in the next figure,  $f[g(t)]$  stands for,  $-\log(-\log[g(t)])$ .

$$g_{\text{exp}}(t) = \psi_{\text{off}}(t) / \psi_{\text{off}}(dt).$$

To proceed correctly with the analysis, one should search for a constant  $c$ , such that plotting the product  $c g_{\text{exp}}(t) [\equiv -G_{\text{exp}}(t)]$  in the manipulated axes leads to a straight line. If such a constant exists, it is unique, and  $G_{\text{exp}}(t)$  approximates  $\psi_{\text{off}}(t)$ . This demand on  $G_{\text{exp}}(t)$  is equivalent to the condition,  $(G_{\text{exp}}(t))_{t \rightarrow 0} = 1$ , which was required in [9]. (In fact,  $c = \psi_{\text{off}}(dt)$ .  $c$  can be easily found from the first couple of data points.)

Note that the above means that for any value of  $t$ :

$$\frac{\psi_{\text{off}}(t) / \psi_{\text{off}}(dt)}{\phi_{\text{off}}(t) / \phi_{\text{off}}(0)} = e^{+(dt/\tau)^\alpha} \geq 1,$$

where the equality holds in the limit  $dt \rightarrow 0$ . This explains Figs. 1 and 2 in [28]; Fig. 1 in [28] plots curves with improper prefactors, and Fig. 2 in [28] plots  $e^{+(dt/\tau)^\alpha}$  as a function  $dt$  for different values of  $\alpha$  and  $\tau$ . Thus, a general conclusion regarding the disability of reconstructing the WT-PDF in Eqs. (1) and (2) from a 1 ms trajectory cannot be drawn from the analysis given by Molski [28]. Moreover, below it is shown by simulation that reconstructing the WT-PDF of Eqs. (1) and (2) with 1 ms bin size is easily performed.

## 2.3. Self consistent test: drawing random times from a stretched exponential WT-PDF and constructing it back

The simplest self-consistent test for reconstructing the stretched exponential WT-PDF in Eqs. (1) and (2) from a trajectory with  $dt = 1$  ms is performed. The numerical test draws random times from  $\psi_{\text{off}}(t)$  using the rejection method [29]. Fig. 3a plots  $f[G_{\text{exp}}(t)]$  obtained from a  $10^4$ -cycle trajectory, with  $c = 0.063$ . Also shown is  $f[\psi_{\text{off}}(t)]$ . These WT-PDFs coincide. Importantly, the decay of  $G_{\text{exp}}(t)$  spans two to three orders of magnitude both in time and in amplitude. For comparison, Fig. 3b plots the same curves, but now  $G_{\text{exp}}(t)$  (with the same value of  $c$ ) is obtained from a  $10^6$ -cycle trajectory. As expected,  $G_{\text{exp}}(t)$  built from the longer trajectory approximates better the input WT-PDF. However, a trajectory with  $10^4$ -cycles is long enough for  $G_{\text{exp}}(t)$  to reliably approximate the temporal decay of the true WT-PDF, and allows an accurate parameter extraction.

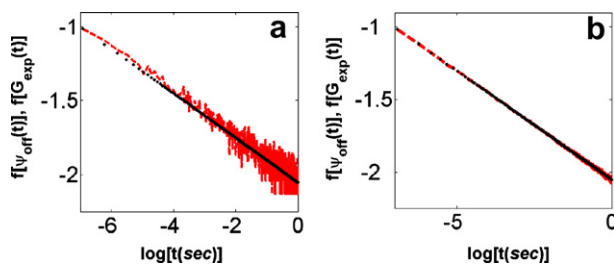


Fig. 3.  $G_{\text{exp}}(t)$  (dotted curve) with  $c = 0.063$  and  $\psi_{\text{off}}(t)$  (circled symbol) in the manipulated axes. In (a),  $10^4$  random times were used to construct  $G_{\text{exp}}(t)$  where in (b)  $10^6$  random times were used to construct  $G_{\text{exp}}(t)$ . Clearly, a  $10^4$ -cycle trajectory allows a reliable construction of  $G_{\text{exp}}(t)$  and an accurate parameters extraction.

### 3. Concluding remarks

In this Letter, it was shown by simulation that a reliable stretched exponential decay pattern with the parameters  $\alpha = 0.15$  and  $\tau = 0.00115$  ms can be recovered from a two-state trajectory with  $dt = 1$  ms whose *off* periods are generated by this WT-PDF. This validates the interpretation of the experimental findings made in [8,9] that for saturating substrate concentrations the *off* periods in the catalytic activity of single lipase B molecules from *Candida antarctica* construct a stretched exponential WT-PDF, and disproves the critique made in [28] about the correctness of such an interpretation. The result of this numerical test should not be a surprise because the moments of the WT-PDF in Eq. (1) with the parameters in Eq. (2) are all much larger than the bin size of 1 ms. The basic physical relevant timescale of a WT-PDF is its first moment (if it is finite). The misinterpretation of  $\tau$  as the timescale in the problem may arise from inappropriate comparison between a stretched exponential WT-PDF and an exponential WT-PDF. A stretched exponential WT-PDF is characterized by 2 parameters,  $\alpha$  and  $\tau$ . Strictly speaking, only for  $\alpha = 1$  is  $\tau$  the relevant timescale in the problem. Although one can consider  $\tau$  as the relevant timescale in the problem even in the vicinity  $\alpha \approx 1$ , in the case of Refs. [8,9], the small value of  $\alpha (=0.15)$  means a highly non-exponential decay pattern. This, in turn, assigns an important role for  $\alpha$  in the determination of the timescale in the problem, which is evident in the moment expressions.

An explanation for the error that led to the wrong critique in [28] is also given. The use of an incorrect prefactor when plotting the experimental curves in the manipulated axes led to the wrong conclusion. It should be noted that for a stretched exponential WT-PDF with a first moment comparable or smaller than the bin size, difficulties in a reliable recovery of the WT-PDF from the trajectory are expected to arise. Such a situation can happen with any form of a WT-PDF; it is a general outcome of the loss of information about the measured process when extracted from badly resolved data.

The occurrence of a stretched exponential decay pattern in measurements of protein activity has a long history, e.g., [30]. The origin for this behavior is usually explained by an ensemble of states, or conformations, that contribute to the activity in a particular way. This argument was used in [8,9] to explain the observed stretched exponential decay pattern. Moreover, the findings in [8,9] showed not just that the periods of the catalytic activity in between successive

dissociation events of product molecules from the enzyme construct a highly stretched exponential WT-PDF, but also that successive such periods are correlated. The implementation of the stretched exponential decay and correlations in the observed catalytic activity into a mechanistic description led to the KS in Fig. 1b, and can be viewed as constrains on, or, properties of, the underlying energy landscape of the process. These properties of the energy landscape can be further related to properties deduced from (e.g.) a theoretical study of the process. Establishing such relationships will most certainly constitute an important step forward in the understanding of biopolymer activity.

### References

- [1] E. Neher, B. Sakmann, Nature Lond. 260 (1976) 799.
- [2] T. Ha et al., Proc. Natl. Acad. Sci. USA 96 (1999) 893.
- [3] H. Yang et al., Science 302 (2003) 262.
- [4] E. Rhoades, E. Gussakovsky, G. Haran, Proc. Natl. Acad. Sci. USA 100 (2003) 3197.
- [5] G. Bokinsky et al., Proc. Natl. Acad. Sci. USA 100 (2003) 9302.
- [6] H. Lu, L. Xun, X.S. Xie, Science 282 (1998) 1877.
- [7] L. Edman, Z. Földes-Papp, S. Wennmalm, R. Rigler, Chem. Phys. 247 (1999) 11.
- [8] K. Velonia et al., Angew. Chem. Int. Ed. 44 (2005) 560.
- [9] O. Flomenbom et al., Proc. Natl. Acad. Sci. USA 102 (2005) 2368.
- [10] O. Flomenbom, R.J. Silbey, Proc. Natl. Acad. Sci. USA 103 (2006) 10907.
- [11] W.J. Bruno, J. Yang, J. Pearson, Proc. Natl. Acad. Sci. USA 102 (2005) 6326.
- [12] O. Flomenbom, J. Klafter, A. Szabo, Biophys. J. 88 (2005) 3780.
- [13] O. Flomenbom, J. Klafter, Acta Phys. Pol. B 36 (2005) 1527.
- [14] R.J. Bauer, B.F. Bowman, J.L. Kenyon, Biophys. J. 5 (1987) 961.
- [15] P. Kienker, Proc. Royal Soc. London B 236 (1989) 269.
- [16] D.R. Fredkin, J.A. Rice, J. Appl. Prob. 23 (1986) 208.
- [17] J. Cao, Chem. Phys. Lett. 327 (2000) 38.
- [18] N. Agmon, J. Phys. Chem. B 104 (2000) 7830.
- [19] H. Qian, E.L. Elson, Biophys. Chem. 101 (2002) 565.
- [20] R. Granek, J. Klafter, Phys. Rev. Lett. 95 (2005) 098106.
- [21] O. Flomenbom, J. Klafter, Phys. Rev. Lett. 95 (2005) 098105.
- [22] J.B. Witkoskie, J. Cao, J. Chem. Phys. 121 (2004) 6373.
- [23] O. Flomenbom, J. Klafter, J. Chem. Phys. 123 (2005) 064903.
- [24] F. Qin, A. Auerbach, F. Sachs, Biophys. J. 79 (2000) 1915.
- [25] D. Colquhoun, A.G. Hawkes, Philos. Trans. Royal Soc. London B Biol. Sci. 300 (1982) 1.
- [26] E. Yeramian, P. Claverie, Nature 326 (1987) 169.
- [27] N. Richter-Dyn, N.S. Goel, Stochastic Models in Biology, Academic Press, 1974.
- [28] A. Molski, Chem. Phys. Lett. 428 (2006) 196.
- [29] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes, Cambridge University press, New York, 1986.
- [30] H. Frauenfelder, P.G. Wolynes, R.H. Austin, Rev. Mod. Phys. 71 (1999) S419.