Statistical analysis of the fractal gating motions of the enzyme acetylcholinesterase

T. Y. Shen,[1] Kaihsu Tai,[2] and J. Andrew McCammon[2,3]
[1]*Department of Physics, University of California, San Diego, La Jolla, California 92093-0365*
[2]*Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California 92093-0365*
[3]*Howard Hughes Medical Institute and Department of Pharmacology, University of California, San Diego, La Jolla, California 92093-0365*

(Received 15 September 2000; published 20 March 2001)

The enzyme acetylcholinesterase has an active site that is accessible only by a ''gorge'' or main channel from the surface, and perhaps by secondary channels such as the ''back door.'' Molecular-dynamics simulations show that these channels are too narrow most of the time to admit substrate or other small molecules. Binding of substrates is therefore ''gated'' by structural fluctuations of the enzyme. Here, we analyze the fluctuations of these possible channels, as observed in the 10.8-ns trajectory of the simulation. The probability density function of the gorge proper radius (defined in the text) was calculated. A double-peak feature of the function was discovered and therefore two states with a threshold were identified. The relaxation (transition probability) functions of these two states were also calculated. The results revealed a power-law decay trend and an oscillation around it, which show properties of fractal dynamics with a ''complex exponent.'' The cross correlation of potential energy versus proper radius was also investigated. We discuss possible physical models behind the fractal protein dynamics; the dynamic hierarchical model for glassy systems is evaluated in detail.

DOI: 10.1103/PhysRevE.63.041902 PACS number(s): 87.15.He, 05.40.−a, 05.45.Df, 87.18.Bb

I. INTRODUCTION

The ideas of fractal geometry [1] have been applied to many branches of science, especially those involving complex systems. It is widely believed that peculiar power-law behaviors emerge from a complex system when there is a lack of characteristic scale (scale-invariance), i.e., the physical observable is of the form $f(bx)=b^{\alpha}f(x)$. The solution is assumed to be a power law with a real exponent α, but some situations can result in a power law with complex exponents [2–4], giving $f(x)\sim \mathrm{Re}(x^{\alpha+i\phi})\sim x^{\alpha}\cos(\phi \ln x)$. Or the general solution is expressed as $f(x)=\mathrm{Re}(\Sigma_m A_m x^{\alpha+im\phi})$. In other words, the signature of a complex exponent is the logarithmic oscillation around the real exponent power-law trend. The situation in which a system shows such behavior can be described by the renormalization-group (RG) analogs, in which case the exponents are eigenvalues of a RG transformation matrix around a fixed point. For most cases, the matrix is assumed to be diagonalizable with real eigenvalues, but in general, we could have complex eigenvalues. It turns out that certain RG calculations in disordered (glassy) systems have found complex eigenvalues [3].

Due to the complex nature of biological systems, there are a variety of phenomena in molecular biophysics [2] that can be characterized by fractal methods, from biopolymer structures to diffusion and chemical kinetics. In this paper, we use these ideas to analyze the dynamics of the functionally important motions (FIM's) [5] of the enzyme acetylcholinesterase (AChE) [6]. AChE is an important serine esterase that catalyzes hydrolysis of the cationic neurotransmitter acetylcholine (ACh^+), thus terminating the synaptic transmission cycle. In order to be hydrolyzed, ACh^+ has to reach the active site of AChE, which lies at the bottom of a narrow and deep gorge. Since the gorge has a very narrow bottleneck region for a ligand of the size of ACh^+ to pass through, the gorge width fluctuation is a crucial FIM for this catalysis. Some earlier work has been conducted to investigate the gating motion [7]. Based on a large data set from a new molecular-dynamics (MD) simulation of mouse acetylcholinesterase (mAChE), we analyze this important gating motion of protein with the language of fractals.

II. DATA COLLECTION

We collected more than 10 000 snapshots from the MD simulation of mAChE at 1-ps intervals. This 10.8-ns trajectory is unusually long for a full atomistic, explicitly cubic-box solvated MD simulation with a protein of this size (mAChE has 543 residues). The simulation was performed on 32 processors of a Cray T3E parallel supercomputer at the San Diego Supercomputer Center over a period of three years, consuming a total of over seven months of supercomputer time. Since this trajectory is an extension of that reported in [8], the MD setup was the same. The time step of the MD simulation was 2 fs.

In order to characterize the degree of the gorge opening with a single variable, we defined the proper radius for the conformation of each snapshot as the maximum radius of a spherical ligand that can go through the gorge from outside the protein to reach the bottom. Equivalently, it is the maximum probe radius with which we can still generate a solvent accessible surface with a continuous topology; that is, if the probe radius is too large, the surface of the bottleneck region will break, the surface near the bottom part of gorge will no longer connect to that on the exterior of the protein, and we will have a discontinuous topology. Figure 1 compares the open and closed conformations thus defined. In our algorithm, we first generate the Shrake and Rupley surface with a testing probe radius [8]; then we try to determine whether the surface generated by the atom O_γ in residue Ser203 (the bottom of the gorge) is topologically continuous with the surface of the bottleneck region (represented by the atoms

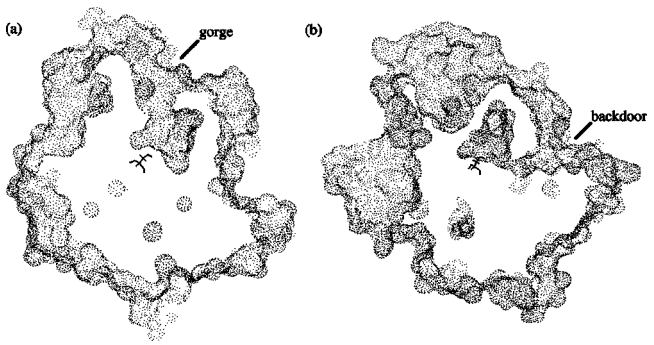

FIG. 1. Comparison of surface for continuous and discontinuous topology. (a) shows a conformation with an open gorge, while (b) shows a closed gorge, but with the ''back door'' open. (Ser203 is shown in the center of each figure.)

Leu76 $C_{\delta 1}$, Trp286C_{β}, and Tyr72 O_H) for that probe radius. Using a binary search algorithm to decide what will be the next probe radius, we can determine the proper radius with desired precision. We started with a test value of 1.6 Å, and assumed the value of the proper radius was bounded between 0.8 and 2.4 Å. With six iterations in the binary search algorithm, we achieved a final discretization of 0.05 Å; in other words, we narrowed down the proper radius to some value within ± 0.025 Å. We calculated a proper radius for each snapshot. The results of these calculations are shown as a time series of the gorge proper radius in Fig. 2.

In addition to the gorge, results from this and previous [8] MD simulations of mAChE also showed an alternative opening occasionally large enough for at least water molecules to pass. This opening, named the ''back door'' [9], is conjectured to assist in releasing the catalyzed products. The opening motion of the ''back door'' is formed by a shutterlike motion of Trp86, Gly448, Tyr449, and Ile451. Instead of using multiple probe radii to calculate the proper radius, as in the case of the gorge, we used a single probe radius of 1.4 Å to test whether the ''back door'' is open or closed. We always blocked the gorge entrance, while probing for ''back door'' opening events. Similarly, we blocked the ''back door'' region in the gorge calculation described above. Among the 10 000 snapshots, only 78 had ''back door'' opening events; for the rest of the time it was closed.

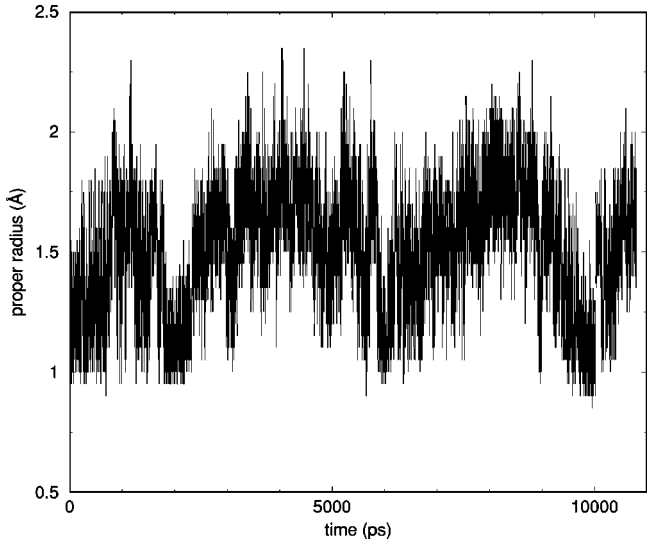

FIG. 2. Time series of the gorge proper radius.

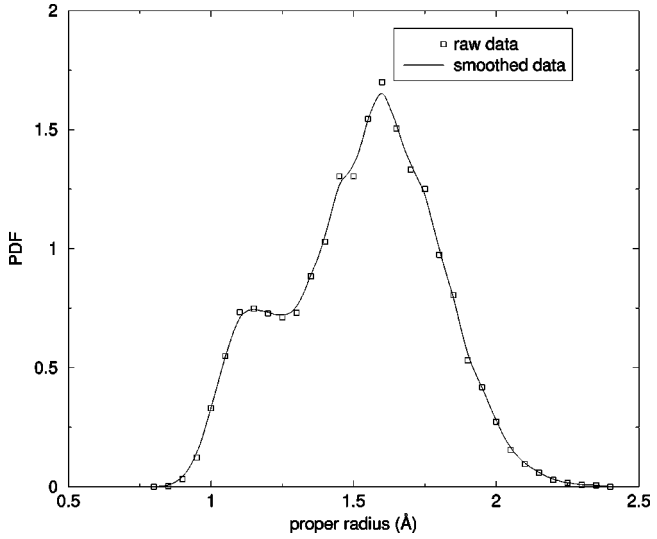

FIG. 3. Probability density function of the gorge proper radius (0.7 ns to 10.8 ns).

III. ANALYSES AND RESULTS

First we should choose a starting point for our equilibrium statistics instead of using all the data points from the MD trajectory. The MD simulation starts from the crystal structure of AChE, a conformation with the gorge closed and high potential energy due to crystal packing effects. During the MD simulation, the solvated protein gradually relaxes; this equilibration process normally takes less than 1 ns. Here we choose the cutoff relaxation time to be 700 ps, which will be justified later.

The probability density function (PDF) of the proper radius is shown in Fig. 3, with the average of the radii being 1.52 Å and standard deviation 0.261 Å. We found that the distribution has two peaks, one centered around 1.15 Å and the other around 1.60 Å. Also shown in the same figure is the smoothed result of the raw discrete data.

By setting a threshold radius $r^* = 1.27$ Å at the local minimum between the two peaks in Fig. 3, we can define two states, which we call the closed and the open states. The physical meaning of the two states is whether, for that conformation, the gorge of AChE allows a spherical ligand of radius 1.27 Å to enter. We then define a dichotomous signal using the following coarse grain transformation function:

$$I(r) = \begin{cases} 0 & \text{(closed state)}, \quad r \leqslant r^*; \\ 1 & \text{(open state)}, \quad r > r^*. \end{cases} \tag{1}$$

We can then calculate the dwell times of each state from the dichotomous signal.

The average radii of the closed and open states are $\langle r_C \rangle = 1.133$ Å and $\langle r_O \rangle = 1.621$ Å. The mean values of the closed- and open-state dwell times are $\langle T_C \rangle = 2.86$ ps and

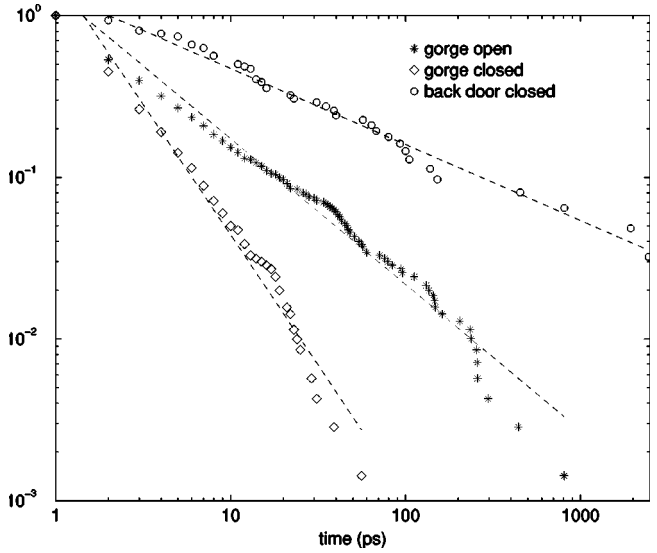

FIG. 4. Log-log plot of the tail of the cumulative probability distribution versus dwell time, showing the power-log behavior over three orders of magnitude. Note for ''back door,'' since we did not have enough data for statistics of the open dwell time, only the closed ''back door'' data are shown.

$\langle T_O \rangle = 11.55$ ps. The ratio of the sum of all the dwell times for the closed state over that for the open state is 1:4. Maximum open and closed dwell times for the gorge are 806 and 56 ps, respectively.

The dwell times are significant: they can be seen as how long the protein stays in a state before passing on to other states. The statistics of these dwell times give the relaxation (transition probability) function, which is an important function for dynamics.

The definition of the tail of the cumulative probability distribution for a PDF $f(t)$ is

$$P(t) = \int_t^\infty f(t')dt' = 1 - F(t), \tag{2}$$

where $F(t)$ is the cumulative probability distribution function as usually defined. Figure 4 is the log-log plot of the tail of the cumulative probability distribution versus dwell time. The data fall very nicely around a straight line over several orders of magnitude in the time axis; this suggests power-law functions $P(t) \sim t^{-\mu}$ and $f(t) \sim t^{-(\mu+1)}$.

In addition to the relaxation function, other common methods for the investigation of nonexponential dynamics include the time correlation function, the power spectrum, and the dynamic scaling exponent for fractional random walk. Figure 5 is the time autocorrelation function of the proper radius time series, showing an interesting nonexponential decay behavior, which initiated our quest for the dynamic nature of the fluctuation. Figure 6 is the log-log plot of the power spectrum of the proper radius time series; equivalently, this is the norm of the Fourier transform of the time series $r(t)$, $S(f) = |\hat{s}(f)|^2 \sim f^{-\beta}$. Again, it shows a general power-law decay trend. It appears to be a scaling noise, also

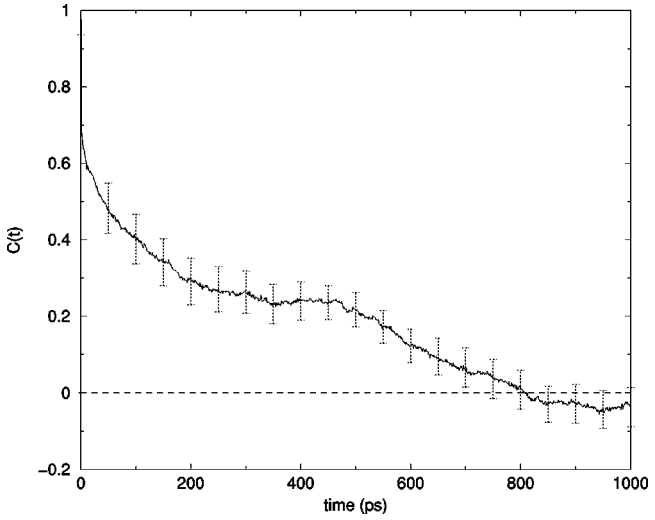

FIG. 5. Time autocorrelation function of the proper radius time series. The error bar is a rough estimation based on the assumption that each of the 500 ps blocks of data is independent. The presence of nonexponential decay makes the estimation of the error bar difficult.

called ''$1/f$'' noise [10]. It starts to level off as white noise at about 10^{11} Hz, due to the limit of our sampling frequency.

Figure 7 shows the calculation of the dynamic scaling exponents z, defined by

$$D(t) \equiv \sqrt{\langle [y(t+t_0) - y(t_0)]^2 \rangle - \langle y(t+t_0) - y(t_0) \rangle^2} \sim t^z, \tag{3}$$

where $y(t) = \Sigma^{i<t} r(i)$. If $z = \frac{1}{2}$, it could be a trivial case of a Brownian motion. Otherwise, $z > \frac{1}{2}$ gives a fractional persistent random walk, and $z < \frac{1}{2}$ antipersistent [11]. In this calculation, z is about 0.89, which is much more persistent than white noise. Dynamic exponents for the random walk are also used to track other systems such as DNA-encoded walk;

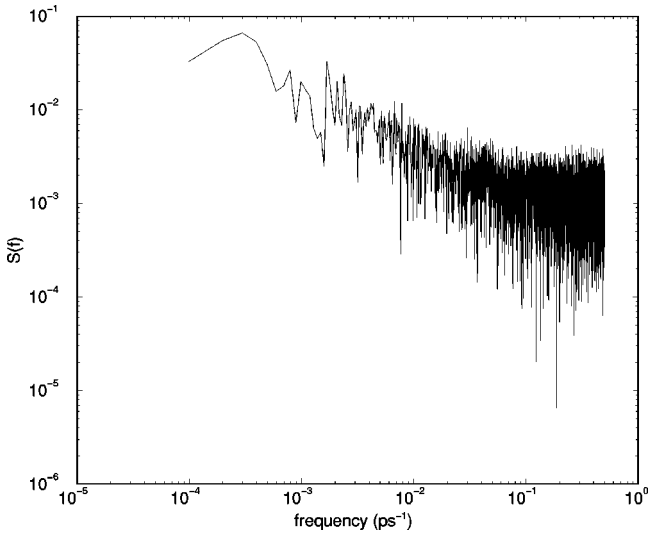

FIG. 6. Power spectrum of the proper radius time series.

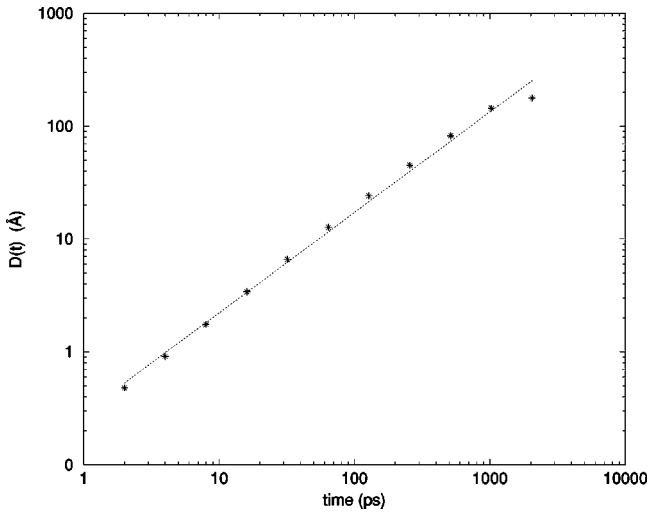

FIG. 7. Dynamic scaling exponent shown as the slope of the log-log plot of Eq. (3). The slope from the linear regression is 0.923 ± 0.007; the correlation coefficient is 0.99977.

this method was reported [12] to offer better results than autocorrelation or power spectrum methods, as is shown in our case.

Apart from the statistics of the proper radius alone, we also studied the cross correlation between the gorge proper radius and the potential energy of AChE (Fig. 8). One can easily see that the states in the first 0.7 ns are distinctly different from those afterwards. The first and second half of the trajectory after 0.7 ns are statistically similar. For the first 700 ps, the system starts from a high-energy state with a narrow gorge, and subsequently the system evolves with MD to a relaxed equilibrium, namely a lower-energy state with more time in the open gate state. This justifies our choice of the beginning of the equilibrium statistics. In order to see whether there is an energy barrier at the threshold between the two peaks, we averaged the potential energy for each value of the proper radius and plotted it for all values of the proper radius (data not shown). This resulted in a very rough energy landscape, making it difficult to describe conclusively.

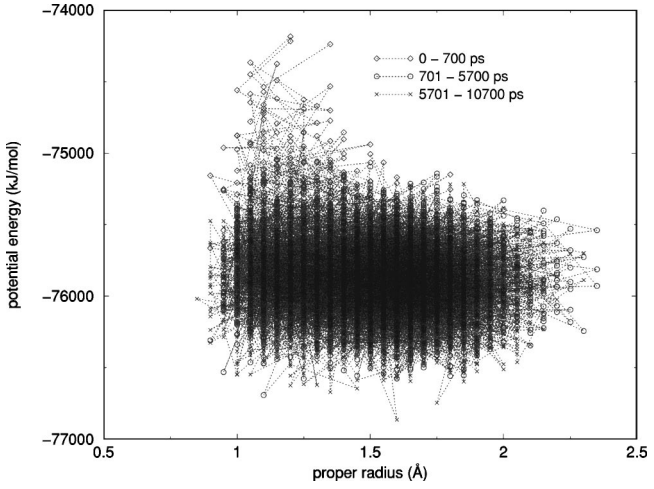

FIG. 8. Proper radius vs potential energy plot. Each dot represents a snapshot. The dashed line shows the transition from one snapshot to the next.

IV. DISCUSSION

Since fractals appear in many different complex systems, many theoretical models have been developed to explain them. Some are used to explain the dynamic fractals in biophysics, for example the amorphous two-level system (TLS), the time-dependent kinetics model, and the defect-diffusion model. For a brief review of models describing the fractal dynamics of protein, see Chaps. 6 and 7 of [2]. Though many of them give the final power-law function, some are easier to understand physically.

The hierarchical model [13], as a possible description of protein dynamics [14], is especially interesting in this case. It was originally proposed in the spin-glass literature to model different kinds of nonexponential dynamics problems, for example stretched exponential dynamics and power-law dynamics. In this model, the complex relaxation event is thought of as a hierarchy of simple Markov relaxations. The hierarchy of relaxations is divided into tiers. In each tier, the relaxations are considered to be parallel and uncorrelated. However, relaxations in the slower tiers can only happen when the faster tiers have moved into required conformations.

In the case of a polymer, the fast degrees of freedom may involve simpler motions with only a few atoms participating, while the slower ones involve a group of atoms, maybe a subdomain. Indeed, most of the bond rotations in a polymer are constrained by simple energy barriers (comparable to a single spin in a strongly coupled glassy material). These simple rotations have to reach a specific conformation in order to lower the barrier for the next tier of events to happen, and so on.

We can express the relaxation function as

$$f(t) = \sum_{i=1}^{N} (c_i/\tau_i)\exp(-t/\tau_i), \tag{4}$$

with $\Sigma_{i=1}^{N} c_i = 1$ as the requirement of normalization. Thus for a simple (single-component) event, the decay is exponential. Since the PDF of the dwell times is a Lévy-type distribution, which has a power-law decay tail, our objective is to find a set of c_i and t_i which can lead to $f(t) \sim A(t)/t^{1+\mu}$, where $A(t)$ can be a correction from the pure power-law behavior. Much effort has been spent in modeling this, e.g., in [15–17]. From both maximum-likelihood fitting of experimental data [18] and theoretical work, a likely solution for the equation above is the Weierstrass spectrum in geometric progression [13,16,19], which defines two ratios:

$$c_{i+1}/c_i = \alpha \quad \text{and} \quad \tau_{i+1}/\tau_i = \lambda, \tag{5}$$

where $\alpha \in (0,1)$ and $\lambda > 1$.

With the help of the Poisson summation formula [20], the solution of Eqs. (4) and (5) can be rewritten (see the RG approach in [19]) as $f(t) = B_N(t)/t^{1+\mu}$, with

$$B_N(t)=\frac{c\tau^{\mu}}{\ln\lambda}\left\{ I_N(\mu+1,\xi)+2\sum_{m=1}^{\infty}\|I_N(\eta_m,\xi)\| \times\cos[\arg I_N(\eta_m,\xi)+\beta_m\ln\xi]\right\}, \quad (6)$$

where $c=(1-\alpha)/(1-\alpha^N)$, $\tau=\tau_1$, $\mu=-\ln\alpha/\ln\lambda$, $\xi=t/\tau$, $\beta_m=(2\pi m)/\ln\lambda$, and $\eta_m=1+\mu-i\beta_m$. Here the integral is defined as

$$I_N(\eta_m,\xi)\equiv\int_{\xi\lambda-N+1}^{\xi} y^{n_m-1}e^{-y}\,dy. \quad (7)$$

In the case of $N\to\infty$, this approaches the complex-valued incomplete Γ function, $\gamma(\eta_m,\xi)$, and for longer time $t\gg\tau$, it asymptotically approaches the complete Γ function $\Gamma(\eta_m)$.

One of the merits of this model compared with others is that it gives not only the dominant power-law trend, but also higher-order oscillations around the trend. As mentioned before, this oscillation corresponds to the complex exponent, or the complex fractal dimension. It has been observed [19,21] in systems with abundant experimental data as well as in our simulation results. A few examples in biophysics include the human bronchial tree [22], ion channel kinetics [18,23], and the tritium exchange in rhodopsin and lysozyme [21]. Earlier phenomenological models successfully fit those data to the function $\varphi(t)=[A_2+A_3\cos(A_4\ln t)]/t^{A_1}$ [Eq. (4) of [21]]. From Eq. (6), we can see that besides the zeroth order with power-law trend, we have a series of cosine functions of decreasing magnitude with increasing order. The oscillation can help us to estimate the parameters of the model. Before we estimate the parameters, we should point out that in the above section, we plot the tail integration of that PDF instead. Fortunately, with simple substitution, one can prove [19] that the integration of the above equation nicely preserves the type of function only with different parameters $(\mu+1\to\mu,\beta_i\to\beta_i)$. Here we use data from the gorge open case as an example to evaluate parameters α and λ for this model. Based on the discussion above and Fig. 4, we can see that for an order of magnitude of change, about five-halves of oscillations occur, therefore the parameter $\beta_1=2\pi/\ln\lambda\approx\pi/(\frac{1}{5})$, so $\ln\lambda\approx\frac{2}{5}$ and $\lambda\approx1.5$. Since the slope $\mu=-\ln\alpha/\ln\lambda\approx0.85$, we estimate $\alpha\approx0.7$ for this case.

V. CONCLUDING REMARKS

We reported striking fractal dynamics of a protein molecule from a MD simulation. Several statistical analysis tools were applied to demonstrate this behavior, and a mathematical model was highlighted in an attempt to describe the system. We want to point out that some of the statistical methods and theoretical models are inspired by and borrowed from the well-studied fractal ion-channel kinetics. This is an example of fractal application, which often connects two systems seemingly without similarity. Here, a similar kind of molecular level gating motion might be a candidate for the underlying phenomena in the ion-channel case, which was also suggested by recent experimental results [24].

ACKNOWLEDGMENTS

We are grateful to L. S. Canino, U. Börjesson, M. Philippopoulos, and T. Hwa for very helpful discussions. K. T. wishes to acknowledge the La Jolla Interfaces in Science interdisciplinary training program and the Burroughs Wellcome Fund for support. This work was supported in part by NSF, NPACI/SDSC, and the Howard Hughes Medical Institute.

[1] B. B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, New York, 1983).
[2] T. G. Dewey, *Fractals in Molecular Biophysics* (Oxford University Press, New York, 1997).
[3] A. Weinrib and B. I. Halperin, Phys. Rev. B **27**, 413 (1983).
[4] B. Doucot *et al.*, Phys. Rev. Lett. **57**, 1235 (1986); D. Sornette, Phys. Rep. **297**, 239 (1998).
[5] H. Frauenfelder, F. Parak, and R. D. Young, Annu. Rev. Biophys. Biophys. Chem. **17**, 451 (1988); J. A. McCammon, in *Simplicity and Complexity in Proteins and Nucleic Acids*, edited by H. Frauenfelder, J. Deisenhofer, and P. G. Wolynes (Dahlem University Press, Berlin, 2000).
[6] J. L. Sussman, M. Harel, F. Frolow, C. Oefner, A. Goldman, L. Toker, and I. Silman, Science **253**, 872 (1991).
[7] H. X. Zhou, S. T. Wlodek, and J. A. McCammon, Proc. Natl. Acad. Sci. U.S.A. **95**, 9280 (1998).
[8] S. Tara, T. P. Straatsma, and J. A. McCammon, Biopolymers **50**, 35 (1999).
[9] M. K. Gilson, T. P. Straatsma, J. A. McCammon, D. R. Ripoll, C. H. Faerman, P. H. Axelsen, I. Silman, and J. L. Sussman, Science **263**, 1276 (1994).
[10] M. B. Weissman, Rev. Mod. Phys. **60**, 537 (1988).
[11] L. S. Liebovitch and W. Yang, Phys. Rev. E **56**, 4557 (1997).
[12] C.-K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Havlin, M. Simons, and H. E. Stanley, Phys. Rev. E **47**, 3730 (1993).
[13] R. G. Palmer, D. L. Stein, E. Abrahams, and P. W. Anderson, Phys. Rev. Lett. **53**, 958 (1984); R. Zwanzig, *ibid.* **54**, 364 (1985); R. G. Palmer, D. L. Stein, E. Abrahams, and P. W. Anderson, *ibid.* **54**, 365 (1985).
[14] D. L. Stein, Proc. Natl. Acad. Sci. U.S.A. **82**, 3670 (1985).
[15] S. Machlup, in *Sixth International Conference on Noise in Physical Systems* (NBS, Washington, D.C., 1981).
[16] T. F. Nonnenmacher and D. J. F. Nonnenmacher, Phys. Lett. A **140**, 323 (1989).
[17] L. S. Liebovitch, Math. Biosci. **93**, 97 (1989); L. S. Liebovitch and F. P. Czegledy, Ann. Biomed. Eng. **20**, 517 (1992).
[18] O. B. McManus and K. L. Magleby, J. Physiol. (London) **402**, 79 (1988).
[19] T. F. Nonnenmacher and D. J. F. Nonnenmacher, in *Stochastic Processes: Physics & Geometry* (*International Conference*), edited by S. A. Albeverio, G. Casati, D. Merlini, G. Cattaneo, and R. Moresi (World Scientific, Singapore, 1991).

[20] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), p. 467.
[21] T. G. Dewey and J. G. Bann, Biophys. J. **63**, 594 (1992).
[22] E. R. Weibel, *Morphometry of the Human Lung* (Academic, New York, 1963); M. F. Shlesinger and B. J. West, Phys. Rev. Lett. **67**, 2106 (1991).
[23] S. Mercik, K. Weron, and Z. Siwy, Phys. Rev. E **60**, 7343 (1999).
[24] S. M. Bezrukov and M. Winterhalter, Phys. Rev. Lett. **85**, 202 (2000).