JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Reaction Event Counting Statistics of Biopolymer Reaction Systems with Dynamic Heterogeneity

Yu Rim Lim,[†] Seong Jun Park,[†] Bo Jung Park,[†] Jianshu Cao,*[,‡] Robert J. Silbey,[‡,∥] and Jaeyoung Sung*[,†,§]

[†]Department of Chemistry, Chung-Ang University, Seoul 156-756, Korea

[‡]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

ABSTRACT: We investigate the reaction event counting statistics (RECS) of an elementary biopolymer reaction in which the rate coefficient is dependent on states of the biopolymer and the surrounding environment and discover a universal kinetic phase transition in the RECS of the reaction system with dynamic heterogeneity. From an exact analysis for a general model of elementary biopolymer reactions, we find that the variance in the number of reaction events is dependent on the square of the mean number of the reaction events when the size of measurement time is small on the relaxation time scale of rate coefficient fluctuations, which does not conform to renewal statistics. On the other hand, when the size of the measurement time interval is much greater than the relaxation time of rate coefficient fluctuations, the variance becomes linearly proportional to the mean reaction number in accordance with renewal statistics. Gillespie's stochastic simulation method is generalized for the reaction system with a rate coefficient fluctuation. The simulation results confirm the correctness of the analytic results for the time dependent mean and variance of the reaction event number distribution. On the basis of the obtained results, we propose a method of quantitative analysis for the reaction event counting statistics of reaction systems with rate coefficient fluctuations, which enables one to extract information about the magnitude and the relaxation times of the fluctuating reaction rate coefficient, without a bias that can be introduced by assuming a particular kinetic model of conformational dynamics and the conformation dependent reactivity. An exact relationship is established between a higher moment of the reaction event number distribution and the multitime correlation of the reaction rate for the reaction system with a nonequilibrium initial state distribution as well as for the system with the equilibrium initial state distribution.

INTRODUCTION

A chemical reaction is, in principle, a stochastic process and the number of chemical reaction events occurring in a time interval or the number of the product molecules generated in the time interval is a random variable with a time-dependent probability distribution. While being negligible in a macroscopic reaction system, the stochastic nature of chemical reactions in such a small reactor as a biological cell has important consequences on cell-to-cell variation in the level of important biomolecules including m-RNA and regulatory proteins that control a cell's biological function, its decision making, and its ultimate fate.[1−3] One of the goals in modern network biology is to understand how a lifeform can achieve a systematic development or a precise control over its function in spite of the stochastic nature of those chemical reactions comprising a cell's reaction network, and how a lifeform takes advantage of the noise or the randomness in chemical reactions occurring in itself.[4,5] To achieve this goal, it is desirable to have a quantitative description of the probabilistic outcome of biological chemical reactions and their networks in small and heterogeneous reaction environments posed by cells.

The master equation approach or Gillespie's stochastic simulation approach has been one of the most popular approaches in the investigation of the number fluctuation of reaction events or product molecules.[6,7] One of the crucial assumptions in the conventional master equation approach or Gillespie's stochastic simulation approach is that rate coefficients of elementary reactions are time-independent constants. However, it is not clear whether the conventional master equation approach or Gillespie's stochastic simulation approach is applicable to biological reaction systems in which reaction rate coefficients are dynamically heterogeneous. Modern single molecule experimental studies tell us that the reaction rate coefficient of a biopolymer keeps fluctuating in line with the conformational dynamics of the biopolymer even in a highly controlled homogeneous reaction environment.[8,9] For biopolymer reactions occurring in cells, reaction rates are different from cell to cell due to heterogeneous reaction environments posed by the cells, which adds additional complexity in quantitative description for the probabilistic outcome of reactions in cells.[10]

With advances in single molecule experimental techniques, the observations of individual reaction trajectories have been made possible for various biopolymer reactions including the stepping of a single molecular motor,[11] the catalytic turnover of a single enzyme,[8,9] the gene expression from a DNA,[10] and single molecule DNA sequencing.[12,13] Individual reaction trajectories recorded in these experiments constitute ideal data for the investigation of probabilistic dynamics of the biopolymer reaction systems. As far as the average behavior of those reaction trajectories is concerned, the conventional chemical kinetics founded on the law of mass action provides a satisfactory description. For example, the average velocity of kinesin motors and the mean enzymatic turnover time of β-galactosidase are

[§]On sabbatical leave at Massachusetts Institute of Technology.

Received: November 4, 2011
Published: January 18, 2012

   dx.doi.org/10.1021/ct200785q | *J. Chem. Theory Comput.* 2012, 8, 1415−1425

consistent with the Michaelis−Menten (MM) relation derived from the conventional chemical kinetics for the simple MM enzyme reaction scheme.[9,14] It is now known that the average enzymatic turnover time obeys the MM relation for a variety of different models of enzyme reactions with fluctuating rate coefficients.[15−17] Recently, a generalized MM relation is established for the mean turnover time of a general multistate enzyme reaction model, which reduces to the MM relation whenever the detailed balance condition between different conformational states is satisfied.[18] However, the conventional chemical kinetics is not so satisfactory in the description of statistical fluctuations in biopolymer reaction trajectories.[8,9] For example, the enzymatic turnover time distribution of a single β-galactosidase enzyme and the waiting time distribution of kinesin motors look inconsistent with the prediction of the conventional chemical kinetics for the simple MM reaction scheme.[9,11] A while ago, the variance in the time-dependent positions of a kinesin motor could be successfully explained by assuming multiple intermediate biochemical states in each step of the kinesin motor, which shows that the statistical distribution of waiting times between steps of a kinesin motor is a nonexponential function.[19] Recently, a general quantitative description of statistical fluctuations in single enzymatic turnover times was also achieved,[17,18,20] for which a generalization of chemical kinetics had to be made in the description of non-Poisson reaction processes of the enzyme−substrate (ES) complex.[16,21] Another interesting statistical property of the reaction trajectories of biomolecules, which cannot be explained within the conventional chemical kinetics, is the correlation between individual reaction times,[22,23] from which direct information about dynamics of reactivity fluctuation can be extracted.[24]

Among various statistical properties of biopolymer reaction trajectories, the probability distribution of the number of the reaction events or the number of product molecules has arguably the greatest biological relevance.[25−27] There are elegant theories for counting statistics of photons emitted from a chromophore with a few optically relevant discrete states;[28−33] however, most of these theories are not directly applicable to reaction event counting statistics (RECS) of a biopolymer reaction whose rate is dependent on a number of continuous conformational state variables of the biopolymer and the surrounding environment. The rate coefficient of the biopolymer reaction is a stochastic process of which the statistical property is most of the time beyond direct measurement and hence unknown. There have been a few general theories of RECS applicable to the biopolymer reaction with rate coefficient fluctuations. Brown introduced the generating function approach to counting statistics of transitions between a bright state and a dark state for a single molecule chromophore of which transition processes are generalized Poisson processes with rate coefficients being stationary Markov processes of any kind.[34] Sung and Silbey presented a formally exact form for RECS of a non-Poisson reaction process with an arbitrary conformation dependent reaction time distribution,[25] and applied it to an investigation of the RECS of the enzyme reaction model in which the rate determining step is the product release process through a fluctuating channel of the enzyme.[35] Cao presented a general theory of photon emission statistics for a single molecule chromophore with a conformation dependent emission rate. On the other hand, Gopich and Szabo investigated the RECS for kinetic networks composed of several Poisson reaction processes.[27]

These theories reduce to Cox's renewal theory when the fluctuation of the reaction rate occurs on a time scale far less than the time scale of individual reaction events.[36] However, recent single biopolymer experiments show that correlations between individual reaction events of a single biopolymer can persist on time scales much longer than the time scale of individual reaction events.[9] In this case, the reaction time distribution of a single biopolymer at one time interval can be significantly different from that of the single biopolymer at another time interval, and the RECS does not obey simple renewal theory.[25] Also, the reaction rate coefficient of a single biopolymer can be persistently different from that of another biopolymer throughout an observation time even though the two molecules are the same kind of chemical species, whenever the observation time scale is not longer than the time scale of the conformation relaxation of the biopolymer system. The RECS of such heterogeneous reaction system can be qualitatively different from the prediction of renewal theory.[25,26]

Although the rate coefficient fluctuation is ubiquitous in a small biopolymer system, still missing is the general relationship between the statistical properties of the rate coefficient and the number distribution of the reaction events even for the simplest elementary biopolymer reaction. In this work, we focus on the number fluctuation of the product molecules or RECS for the elementary biopolymer reaction, $S + E(\mathbf{r}) \xrightarrow{k(\mathbf{r})} P + E(\mathbf{r})$, where S, P, $E(\mathbf{r})$, and $k(\mathbf{r})$ respectively denote the substrate, the product, the biopolymer, and the reaction rate of the biopolymer when our system, including the biopolymer molecules and the surrounding environment, is in state $\mathbf{r}$.

In the present work, we generalize the counting statistics of the Poisson reaction process, presenting the exact expressions of the RECS for a general biopolymer reaction model in which the reaction rate coefficient $k(\mathbf{r})$ can be any function of system state variable $\mathbf{r}$ that undergoes arbitrary multidimensional Markov dynamics. Exact results obtained for the general model predict a universal kinetic phase transition in the fluctuation of reaction event statistics of the dynamically heterogeneous reaction system; the difference $\Delta(t)$ of the variance $\sigma_n^2(t)$ from the mean $\langle n(t)\rangle$ of the number of reaction events occurring in time t is proportional to t^2 or $\langle n(t)\rangle^2$ at short times, but it becomes proportional to t or $\langle n(t)\rangle$ on time scales much longer than the time scale of the rate coefficient fluctuation of the biopolymer. We also generalize Gillespie's stochastic simulation method for the numerical investigation of RECS of a reaction with slow reactivity fluctuation and confirm that the simulation results for RECS are in good agreement with the prediction of the analytic results obtained for the general biopolymer reaction model. On the basis of the results, we propose a method of quantitative analysis for the first two moments of the RECS of a biopolymer reaction. The new analysis method enables one to extract information about the magnitude of equilibrium fluctuation and the relaxation of the conformation dependent reaction rate of the biopolymer, without a bias that can be introduced by assuming a particular model of conformational dynamics and the conformation dependent reactivity. Our new stochastic simulation method developed for investigation of chemical fluctuations in reaction systems with rate coefficient fluctuations is particularly useful when the propagator of the system state is available; it gives the numerical results for the RECS of a reaction system with a state-dependent rate coefficient much more efficiently compared to the stochastic

simulation method involving an explicit simulation of the system dynamics in the state space.

To begin, we briefly review RECS for several important models of stochastic reaction processes, including Poisson reaction process, renewal reaction process, and the bottleneck enzyme reaction, an exactly solvable model of a nonrenewal reaction process. We demonstrate the emergence of the kinetic phase transition in the RECS of the bottleneck enzyme model, which results from dynamic fluctuation of the heterogeneous rate coefficient in the enzyme model. Next, we present the exact analytic expressions for the mean and the variance in RECS for a general model of the elementary biopolymer catalytic reaction and show that the kinetic phase transition in the RECS is universal for any reaction system with conformation dependent reactivity and Markov conformational dynamics. The correctness of the analytic results is confirmed by our stochastic simulation results. In the subsequent section, we discuss the RECS of the reaction system in which the rate coefficient fluctuations have multiple relaxation time scales and establish the general relationship between the higher order moment of the reaction event number distribution and the multitime autocorrelation function of the rate coefficient. In the beginning of this work, we assume that the initial state distribution of the reaction system composed of a number of molecules obeys the equilibrium distribution. Nevertheless, in the Discussion, we also discuss the RECS of a system with a nonequilibrium initial state distribution that is relevant to modern single molecule experiments. Particular emphasis is laid on the RECS of the single molecule system under the nonequilibrium initial state distribution attained by sampling the initial measurement time only at the moment at which each individual single molecule reaction event begins. In the Conclusion section, we conclude the present work. A detailed account of our stochastic simulation method is appended in the Appendix.

BRIEF REVIEW ON RECS

Poisson Reaction Process. Let us consider a hypothetical single enzyme E^0 with a single state, at which the enzyme consecutively performs the catalytic reaction, $S + E^0 \xrightarrow{k} P + E^0$, with steady-state reaction rate k. The steady-state reaction rate k of the single enzyme is constant in time but may be dependent on substrate concentration $[S]$, i.e., $k = k([S])$. For the well-known Michaelis−Menten (MM) enzyme reaction scheme, $k([S])$ is given by $k([S]) = k_{max}[S]/([S] + K_M)$, where k_{max} and K_M respectively denote the steady-state catalytic turnover frequency of the single enzyme E^0 in the high substrate concentration limit and the MM constant. We note that the enzyme reaction is, in reality, a non-Poisson process that involves an intermediate enzyme−substrate complex with a fluctuating reactivity.[17] However, the hypothetical enzyme considered here performs the catalytic reaction in a single elementary reaction step without any reactivity fluctuation. The probability distribution $p_n^{(0)}(t)$ of the number, n, of the reaction events catalyzed by the single hypothetical enzyme $E^{(0)}$ in time t satisfies the following master equation:[6]

$$\frac{\partial}{\partial t}p_n^{(0)}(t) = k[p_{n-1}^{(0)}(t) - p_n^{(0)}(t)]$$

with the initial condition, $\lim_{t\to 0} p_n^{(0)}(t) = \delta_{n0}$. Given that the substrate concentration $[S]$ is so high that its change during our observation time is negligible, the solution of the above partial differential equation can be obtained as

$$p_n^{(0)}(t) = \exp(-kt)\frac{(kt)^n}{n!} \quad (n \geq 0) \tag{1}$$

$p_n^{(0)}$ given in eq 1 is well-known as Poisson distribution. The first two moments of the Poisson distribution are given by $\langle n(t)\rangle = kt$ and $\langle n^2(t)\rangle = \langle n(t)\rangle + \langle n(t)\rangle^2$, so that the variance $\sigma_n^2(t)$ ($\equiv \langle n^2(t)\rangle - \langle n(t)\rangle^2$) of the Poisson distribution is the same as its mean, $\langle n(t)\rangle$.

The reaction-free probability, $P_0(t)$, designates the probability that the reaction has not occurred until time t given that the reaction begins at time 0. The reaction-free probability $p_0^{(0)}(t)$ of the Poisson reaction process decays as a single exponential function, $p_0^{(0)}(t) = \exp(-kt)$ according to eq 1 with n being equal to 0. An important related quantity is the reaction time distribution, $\psi(t)$, defined by $\psi(t) \equiv -dP_0(t)/dt$; $\psi(t)dt$ is the probability that the reaction event is completed in time interval $(t, t + dt)$ given that the reaction event begins at time 0. The reaction time distribution $\psi^{(0)}(t)$ of the Poisson reaction process is given by $\psi^{(0)}(t) = k\exp(-kt)$, whose first two moments are $\langle t\rangle = 1/k$ and $\langle t^2\rangle = 2/k^2$, satisfying $\langle t^2\rangle = 2\langle t\rangle^2$. The latter relation may not hold for stochastic reaction processes other than the Poisson process. One example of such stochastic reaction process more general than the Poisson process is the renewal reaction process discussed next.

Renewal Reaction Process. When our enzyme reaction involves multiple transition states or when it has multiple reaction channels, the reaction time distribution $\psi(t)$ associated with the single enzymatic turnover can be an arbitrary nonexponential function, and the RECS of the enzyme reaction deviates from the Poisson distribution. If individual enzymatic turnover reaction events are statistically identical and uncorrelated with each other, the RECS of the enzymatic reactions can be described by renewal theory.[36] The expression for the probability distribution $p_n^{ren}(t)$ of the number, n, of the renewal reaction events occurring in time t is simpler in the Laplace domain:

$$\hat{p}_n^{ren}(u) = \frac{1 - \hat{\psi}(u)}{u}\hat{\psi}^n(u) \tag{2}$$

From here on, $\hat{f}(u)$ denotes the Laplace transform of $f(t)$, i.e., $\hat{f}(u) \equiv \int_0^\infty e^{-ut} f(t)\, dt$. In the Laplace domain, the first two moments of p_n^{ren} are given by $\langle \hat{n}(u)\rangle = \hat{\psi}(u)/(u[1 - \hat{\psi}(u)])$ and $\langle \hat{n}^2(u)\rangle = \langle \hat{n}(u)\rangle + 2u\langle \hat{n}(u)\rangle^2$, where $\langle \hat{n}^k(u)\rangle$ and $\langle \hat{n}(u)\rangle^k$ denote $\int_0^\infty dt\, e^{-ut}\langle n^k(t)\rangle$ and $[\int_0^\infty dt\, e^{-ut}\langle n(t)\rangle]^k$, respectively. These results for the renewal reaction process reduce to those of the Poisson reaction process when the reaction time distribution $\psi(t)$ is chosen to be the single exponential function, i.e., when $\psi(t) = k\exp(-kt)$.

As long as the mean $\langle t\rangle[\equiv \int_0^\infty dt\, t\psi(t)]$ of reaction time distribution $\psi(t)$ exists, the mean reaction number $\langle n(t)\rangle$ of $p_n^{ren}(t)$ exhibits the following asymptotic behavior:

$$\langle n(t)\rangle \cong t/\langle t\rangle \quad (t/\langle t\rangle \gg 1) \tag{3}$$

irrespective of the functional form of $\psi(t)$, which is qualitatively the same as the mean reaction number of the Poisson reaction process with k being given by $\langle t\rangle^{-1}$. In comparison, the reaction time fluctuation, $\langle t^2\rangle - \langle t\rangle^2$, and the reaction number fluctuation, $\langle n^2\rangle - \langle n\rangle^2$, of a renewal reaction process can be qualitatively different from those of the Poisson reaction process. One of the quantities that measures the non-Poisson

character of a stochastic process is randomness parameter R, defined by

$$R \equiv \frac{\langle t^2\rangle - \langle t\rangle^2}{\langle t\rangle^2} - 1 \tag{4}$$

where $\langle t^k\rangle$ denotes the kth moment of the reaction time distribution $\psi(t)$ of the stochastic process. Another quantity that characterizes the deviation of a stochastic process from the Poisson process is Mandel's Q parameter, defined by

$$Q(t) \equiv \frac{\sigma_n^2(t) - \langle n(t)\rangle}{\langle n(t)\rangle} \tag{5}$$

where ${\sigma_n}^2(t)$ denotes the variance, ${\sigma_n}^2(t) = \langle n^2(t)\rangle - \langle n(t)\rangle^2$, in RECS. Both R and $Q(t)$ vanish for a Poisson reaction process; however, neither R nor $Q(t)$ vanishes for a non-Poisson reaction process. One can show that Mandel's Q parameter defined in eq 5 becomes approximately the same as randomness parameter R for a renewal process at a time t much longer than the mean reaction time, $\langle t\rangle$, which means that

$$\Delta(t)[\equiv \sigma_n^2(t) - \langle n(t)\rangle] \cong R\langle n(t)\rangle \quad (t/\langle t\rangle \gg 1) \tag{6}$$

for a renewal reaction process with randomness parameter R. The linear dependence of $\Delta(t)$ on $\langle n(t)\rangle$ or t given in eq 6 holds for any renewal reaction process as long as the first two moments of reaction time distribution $\psi(t)$ exist. When our reaction process is a nonrenewal process, however, eq 6 may not hold even if the first two moments of the reaction time distribution exist. An exactly solvable model of such a non-renewal reaction process is discussed next.

Bottleneck Enzyme Reaction. In ref 35, Zwanzig introduced the bottleneck enzyme model, whose catalytic reaction rate is controlled by the product escape process out of the enzyme's active site through a dynamically fluctuating channel,[37] and investigated the effects of rate coefficient fluctuation on the time dependence of the reaction free probability for the enzyme reaction. In the bottleneck enzyme model, the catalytic reaction rate of the enzyme is proportional to the square of the radius $|r(t)|$ of the product escape channel with the time series constructed by $r(t)$ being the Ornstein−Uhlenbeck (OU) process, which is statistically equivalent to the time series constructed by the position of a Brownian particle under harmonic potential in the high friction limit.[38]

In this section, we review the RECS of the bottleneck enzyme.[25] If $p_n(r,t)dr$ denotes the joint probability that the value of $r(t)$ is in interval $(r, r + dr)$ and the number of the reaction events catalyzed by the bottleneck enzyme in time internal $(0, t)$ is n, then $p_n(r,t)$ satisfies the following generalized master equation:

$$\frac{\partial}{\partial t}p_n(r,t) = \kappa r^2[p_{n-1}(r,t) - p_n(r,t)] + D\frac{\partial}{\partial r}\left[\frac{\partial}{\partial r}p_n(r,t) + \frac{r}{b^2}p_n(r,t)\right] \tag{7}$$

In eq 7, κr^2, D, and b^2 respectively denote the reaction rate coefficient proportional to the area of the product escape channel, the diffusion constant governing the relaxation rate of stochastic variable $r(t)$, and the variance of r at equilibrium. We assume that the initial value $r_i[\equiv r(t_i)]$ of $r(t)$ is distributed according to the equilibrium distribution. From eq 7, one can obtain the exact expression for the characteristic function, $F(q,t)$, also called the moment generating function, defined by $F(q,t) = \sum_{n=0}^{\infty} q^n p_n(t)$, with $p_n(t)$ being the probability that the number of catalytic turnovers occurred in time t is n, i.e., $p_n(t) = \int_{-\infty}^{\infty} dr\, p_n(r,t)$:[34]

$$F(q,t) = \exp\left[-\frac{D(S-1)t}{2b^2}\right] \times \left(\frac{(S+1)^2 - (S-1)^2\chi(t)}{4S}\right)^{-1/2} \tag{8}$$

Here, S and $\chi(t)$ are defined by $S \equiv (1 + 4\kappa(1-q)b^4/D)^{1/2}$, and $\chi(t) \equiv \exp(-2DSt/b^2)$, respectively. Making use of the well-known property of the characteristic function, $\langle n(t)\rangle = [\partial F(q,t)/\partial q]_{q=1}$, and $\langle n(n-1)(t)\rangle = [\partial^2 F(q,t)/\partial q^2]_{q=1}$, one can obtain the exact expressions for $\langle n(t)\rangle$ and $\Delta(t)$ ($\equiv \langle n^2(t)\rangle - \langle n(t)\rangle^2 - \langle n(t)\rangle$) as follows:

$$\langle n(t)\rangle = \kappa b^2 t \tag{9}$$

$$\Delta(t)(\equiv \sigma_n^2(t) - \langle n(t)\rangle) = 2(\kappa b^2 t)^2 g(\lambda t) \tag{10}$$

with $g(x) = 2[x - 1 + \exp(-x)]/x^2$ and $\lambda = 2D/b^2$. The mean reaction number $\langle n(t)\rangle$ increases linearly in measurement time t, which is qualitatively the same as that given in eq 3 for the renewal process with the mean reaction time being given by $\langle t\rangle = (\kappa b^2)^{-1}$. In comparison, the time dependence of $\Delta(t)$ given in eq 10 exhibits a kinetic phase transition:

$$\Delta(t) \cong \begin{cases} 2(\kappa b^2 t)^2 = 2\langle n(t)\rangle^2 & (t \ll \lambda^{-1}) \quad (11a) \\ 4\dfrac{(\kappa b^2)^2}{\lambda}t = 4\dfrac{\kappa b^2}{\lambda}\langle n(t)\rangle & (t \gg \lambda^{-1}) \quad (11b) \end{cases} \tag{11}$$

which is qualitatively different from the time dependence of $\Delta(t)$ given in eq 6 for the renewal process. Note that the kinetic phase transition in $\Delta(t)$ from the short time behavior proportional to t^2 or $\langle n(t)\rangle^2$ to the long time behavior linear in t or $\langle n(t)\rangle$ occurs near $t \cong \lambda^{-1}$, which is the characteristic relaxation time of the bottleneck enzyme's configuration variable, $r(t)$.

For the OU process, $r(t)$, the probability density $G_{OU}(r,t + t_i|r_i,t_i)$ that the value of $r(t + t_i)$ is r, given that the value of $r(t_i)$ is r_i, is given by $G_{OU}(r,t + t_i|r_i,t_i) = (2\pi b^2[1 - \phi^2(t)])^{-1/2} \exp(-([r - (\phi(t)\, r_i)]^2)/(2b^2[1 - \phi^2(t)]))$ where $\phi(t)$ denotes the normalized time correlation function of $r(t)$, i.e., $\phi(t) = \langle r(t + t_i)r(t_i)\rangle_{eq}/b^2 = \exp(-\lambda t/2)$ with $\lambda = D/\sigma^2$. At a time t far smaller than λ^{-1}, we have $\phi^2(t) \cong \phi(t) \cong 1$ and $G_{OU}(r,t + t_i|r_i,t_i) \cong \delta(r - r_i)$; in other words, conformational relaxation of the enzyme does not occur significantly and the conformation, $r(t + t_i)$, of the enzyme at time $t + t_i$ is nearly the same as the initial conformation, $r(t_i)$ ($\equiv r_i$). In the latter case, the dynamic fluctuation of rate coefficient $\kappa r^2(t)$ is negligible so that the RECS of the single bottleneck enzyme with initial configuration r_i can well be approximated by the Poisson distribution ${p_n}^{(0)}(t)$ given in eq 1 with rate coefficient k being equal to $\kappa {r_i}^2$, and the first two moments of the Poisson distribution are given by $\langle n(t|r_i)\rangle = \kappa {r_i}^2 t$ and $\langle n^2(t|r_i)\rangle = \langle n(t|r_i)\rangle^2 + \langle n(t|r_i)\rangle$. By taking the average of the latter results over the equilibrium distribution, $(2\pi b^2)^{-1} \exp(-2^{-1}{r_i}^2/b^2)$, for an initial configuration r_i of the bottleneck enzymes, we obtain $\langle n(t)\rangle = \kappa b^2 t$, $\langle n^2(t)\rangle = 3\kappa^2 b^4 t^2 + \kappa b^2 t$, and $\Delta(t) = 2(\kappa b^2 t)^2$, which is in agreement with the asymptotic short time behavior given in eq 11a.

The behavior of $\Delta(t)$ at a time t longer than λ^{-1} given in eq 11b for the bottleneck enzyme is linear in $\langle n(t)\rangle$, qualitatively

the same as that given in eq 6 for a renewal reaction process.[36] Making a comparison between eq 6 and eq 11b, one can see that, at a time t longer than λ^{-1}, the first two moments, $\langle n(t)\rangle$ and $\langle n^2(t)\rangle$, of the RECS of the bottleneck enzyme become the same as those of the renewal process with a reaction time distribution $\psi(t)$ of which mean reaction time $\langle t\rangle$ and randomness parameter R are given by $\langle\tau\rangle = (\kappa b^2)^{-1}$ and $R = 4\kappa b^2/\lambda$. In the fast fluctuation limit where $\lambda \gg \kappa b^2$, randomness parameter R for the reaction of the bottleneck enzyme vanishes, which indicates that the reaction of the enzyme with a fluctuating bottleneck reduces to a simple Poisson process in the fast fluctuation limit. In the slow fluctuation limit, or in the small λ limit, R diverges. These limiting behaviors of R for the bottleneck enzyme are consistent with Zwanzig's result for the reaction free probability $p_0(t)$, which is the same as $F(q = 0,t)$.[35] According to Zwanzig, $p_0(t) \cong \exp(-\kappa b^2 t)$ in the fast fluctuation limit, but $p_0(t) \cong (1 + 2\kappa b^2 t)^{-1/2}$ in the slow fluctuation limit. Noting that the reaction time distribution $\psi(t)$ is related to $p_0(t)$ by $\psi(t) = -\partial p_0(t)/\partial t$, one can show that $\lim_{\lambda\to\infty} R = 0$ and $\lim_{\lambda\to 0} R = \lim_{T\to\infty}(1+(2\kappa b^2 T))^{1/2} \to \infty$. However, one cannot obtain eq 11b, correct asymptotic RECS of the bottleneck enzyme model, in the framework of simple renewal theory with the use of the exact reaction time distribution given by $\psi_{exact}(t) = -\partial F(q = 0,t)/\partial t$. This is because the reaction process of the bottleneck enzyme is not really a simple renewal process.

RECS of Biopolymer Undergoing Non-Renewal Elementary Reactions. The kinetic phase transition of $\Delta(t)[\equiv \sigma_n^2(t) - \langle n(t)\rangle]$ found for the bottleneck enzyme model is not limited to the model; instead, it is universal for any reaction system with a state dependent rate coefficient and Markovian state dynamics. To show this, we investigate the RECS of a general reaction model in which catalytic rate coefficient $k(\mathbf{r})$ is an arbitrary function of the system state vector $\mathbf{r}$. The reaction event counting statistics of such enzyme can be described by the following generalized master equation:[25]

$$\frac{\partial}{\partial t}p_n(\mathbf{r},t) = k(\mathbf{r})[p_{n-1}(\mathbf{r},t) - p_n(\mathbf{r},t)] + L(\mathbf{r})\,p_n(\mathbf{r},t) \tag{12}$$

Here, $p_n(\mathbf{r},t)$ denotes the probability density that the dynamic state vector $\mathbf{r}(t)$ of the system state is given by $\mathbf{r}$ and the number of reaction events catalyzed by the biopolymer in a time interval $(0,t)$ is n, which satisfies the following normalization condition, $\sum_{n=0}^{\infty}\int d\mathbf{r}\, p_n(\mathbf{r},t) = 1$. $L(\mathbf{r})$ denotes an evolution operator describing the Markov dynamics of state vector $\mathbf{r}(t)$. The only restriction on the mathematical form of operator $L(\mathbf{r})$ is that it should not contain any explicit time dependence. Representative examples of state dynamics that can be described by the time independent operator $L(\mathbf{r})$ include diffusive dynamics, Langevin dynamics, Newtonian dynamics, Nose–Hoover dynamics, and Monte Carlo dynamics, under time-independent external potential. The initial condition associated with eq 12 is chosen as $\lim_{t\to 0} p_n(\mathbf{r},t) = \delta_{n0} f_{eq}(\mathbf{r})$, where $f_{eq}(\mathbf{r})$ denotes the equilibrium distribution of $\mathbf{r}$, satisfying $L(\mathbf{r})\, f_{eq}(\mathbf{r}) = 0$. From eq 12 with this initial condition, one can obtain exact expressions for $\langle n(t)\rangle$ and $\Delta(t)$:[25,26]

$$\langle n(t)\rangle = k_{eq}t \tag{13}$$

$$\Delta(t)(\equiv \sigma_n^2(t) - \langle n(t)\rangle) = 2\int_0^t d\tau(t-\tau)\langle \delta k(\tau)\,\delta k(0)\rangle_{eq} \tag{14}$$

where k_{eq} and $\langle \delta k(t)\,\delta k(0)\rangle_{eq}$ denote the equilibrium reaction rate coefficient and the time correlation function of the rate coefficient fluctuation $\delta k(t)[\equiv k(\mathbf{r}(t)) - k_{eq}]$, defined by $k_{eq} = \int d\mathbf{r}\, f_{eq}(\mathbf{r})\, k(\mathbf{r})$ and $\langle \delta k(t)\,\delta k(0)\rangle_{eq} = \int d\mathbf{r}\, d\mathbf{r}_0\, \delta k(\mathbf{r})\, G(\mathbf{r},t|\mathbf{r}_0)\, \delta k(\mathbf{r}_0)\, f_{eq}(\mathbf{r}_0)$, respectively. $G(\mathbf{r},t|\mathbf{r}_0)$ denotes the propagator defined by $\exp[tL(\mathbf{r})]\,\delta(\mathbf{r} - \mathbf{r}_0)$. It satisfies $\partial G(\mathbf{r},t|\mathbf{r}_0)/\partial t = L(\mathbf{r})\, G(\mathbf{r},t|\mathbf{r}_0)$ with the initial condition $G(\mathbf{r},0|\mathbf{r}_0) = \delta(\mathbf{r} - \mathbf{r}_0)$. Mandel's Q calculated from eq 14 is obtained as

$$Q(t) = 2\frac{\langle \delta k^2\rangle}{k_{eq}}\left[\int_0^t d\tau\phi_k(\tau) - \frac{1}{t}\int_0^t d\tau\tau\phi_k(\tau)\right]$$

with $\phi_k(t)$ being the normalized autocorrelation function $\langle \delta k(t)\,\delta k(0)\rangle_{eq}/\langle \delta k^2\rangle_{eq}$ of the rate coefficient fluctuation.

From eq 14, we obtain the following asymptotic behaviors for $\Delta(t)$:

$$\Delta(t) \cong \begin{cases} \langle \delta k^2\rangle t^2 = \dfrac{\langle \delta k^2\rangle}{k_{eq}^2}\langle n(t)\rangle^2 \quad (t \ll \xi) \quad (15a) \\ 2\langle \delta k^2\rangle \xi t = 2\dfrac{\langle \delta k^2\rangle \xi}{k_{eq}}\langle n(t)\rangle \quad (t \gg \xi) \quad (15b) \end{cases} \tag{15}$$

where ξ is the characteristic time of rate coefficient fluctuations defined by $\xi = \int_0^\infty dt\, \phi_k(t)$ with $\phi_k(t)$ being the normalized autocorrelation function of the rate coefficient fluctuation δk, i.e., $\phi_k(t) = \langle \delta k(t)\,\delta k(0)\rangle/\langle \delta k^2\rangle$. Equation 15 clearly shows the kinetic phase transition in $\Delta(t)$ for the general model of the elementary biomolecule reaction. Equations 13–15 correctly reduce to eqs 9–11 for the bottleneck enzyme model, for which $\langle \delta k^2\rangle_{eq} = \kappa^2(\langle r^4\rangle_{eq} - \langle r^2\rangle_{eq}^2) = 2(\kappa b^2)^2$, $\phi_k(t) = \exp(-\lambda t)$, and $\xi = \lambda^{-1}$.

Comparing eqs 13 and 15b to eqs 3 and 6 obtained for the renewal process, one can see that, when the size t of the measurement time bin is much greater than the characteristic relaxation time ξ of rate coefficient fluctuations, the first two moments of the nonrenewal RECS of the general enzyme model satisfying eq 12 can be approximated by the RECS of the renewal reaction process with a reaction time distribution $\psi(\tau)$ of which mean $\langle\tau\rangle$ and randomness parameter R are given by

$$\langle\tau\rangle = k_{eq}^{-1} \tag{16}$$

and

$$R = 2(\langle \delta k^2\rangle/k_{eq}^2)(k_{eq}\xi) \tag{17}$$

In contrast, when the size t of the measurement time bin is smaller than ξ, $\Delta(t)$ given in eq 15a for a system of biomolecules with reactivity fluctuation is proportional to t^2 or $\langle n(t)\rangle^2$, qualitatively different from the prediction, eq 6, of renewal theory. In the latter case, renewal theory is inapplicable to the RECS of the system. More details on this issue are presented in the Discussion section.

Note that the mean reaction number $\langle n(t)\rangle$ in eq 13 provides information about the equilibrium reaction rate coefficient, k_{eq}, only. By analyzing the behavior of fluctuations in RECS further, one can extract quantitative information about the magnitude and the dynamics of the rate coefficient fluctuation of the biopolymer reaction. From eqs 13 and 15, we get

$$\ln\Delta(t) \cong 2\ln\langle n(t)\rangle + \ln(\langle \delta k^2\rangle/k_{eq}^2) \tag{18a}$$

at the short time ($t \ll \xi$) and

$$\ln\Delta(t) \cong \ln\langle n(t)\rangle + \ln(\langle \delta k^2\rangle/k_{eq}^2) + \ln(2k_{eq}\xi) \tag{18b}$$

dx.doi.org/10.1021/ct200785q | J. Chem. TheoryComput. 2012, 8, 1415–1425

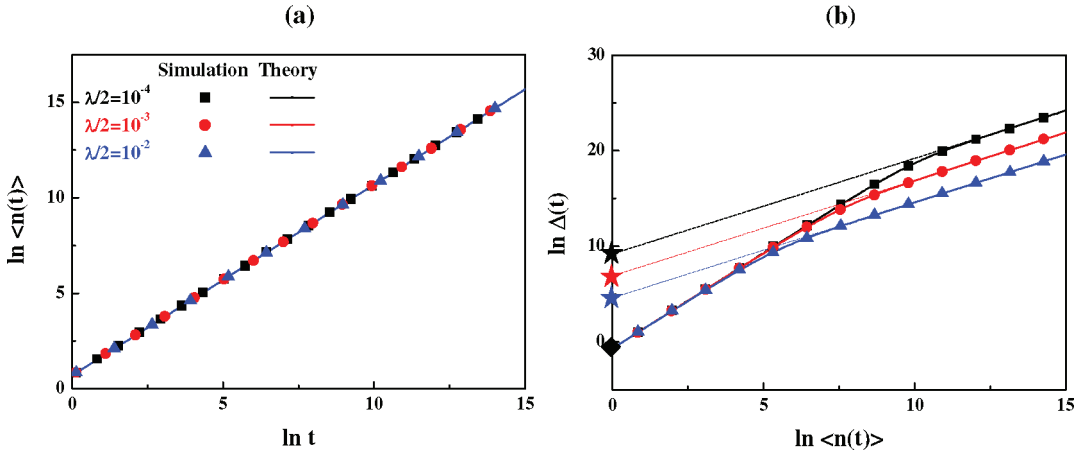

Figure 1. Comparison of predictions of eqs 13 and 14 to the stochastic simulation results for a model of rate coefficient fluctuation, $k(t) = k_0 + \kappa r^2(t)$ with $r(t)$ being a Gaussian random variable with time correlation function $\langle r(t + t_0)\, r(t_0)\rangle_{eq} = b^2 \exp(-\lambda t/2)$. $\Delta(t)$ denotes $\sigma_n^2(t) - \langle n(t)\rangle$ where $\sigma_n(t)$ and $\langle n(t)\rangle$ are the variance and the mean of the number of reaction events occurring in time t. Values of k_0, κ, and b^2 are set equal to unity. (a) The mean reaction number $\langle n(t)\rangle$ is given by $k_{eq}t$, independent of the time scale, $\xi = \lambda^{-1}$, of the rate coefficient fluctuation; k_{eq} is given by $k_0 + \kappa b^2$ for this model. (b) $\Delta(t)$ is proportional to $\langle n(t)\rangle^2$ at short times, but it becomes linear in $\langle n(t)\rangle$ at long times. The phase transition in the counting statistics occurs at time $t \cong \xi = \lambda^{-1}$. The intercepts (♦) of the curve at short times represents ln $\langle \delta k^2\rangle/k_{eq}^2$, whereas the intercepts (∗) of the lines extrapolated from the curves at long times represent $\ln(\langle \delta k^2\rangle/k_{eq}^2) + \ln(2k_{eq}\xi)$.

at the long time ($t \gg \xi$). Therefore, in the plot showing the dependence of $\ln\Delta(t)$ on $\ln\langle n(t)\rangle$, the slop of the data curve would change from 2 to 1 as the size t of the measurement time bin increases from a value much smaller than the characteristic relaxation time ξ of the rate coefficient fluctuation to a value much greater than ξ. From the values of the intercepts of the two lines, one in the short time regime and the other in the long time regime, one can separately measure the values of $\langle \delta k^2\rangle/k_{eq}^2$ and ξ, which respectively measure the magnitude and the relaxation time of the rate coefficient fluctuation.

Comparison to Stochastic Simulation Result. Gillespie's stochastic simulation method provides the numerical solution of the master equation, which provides numerical results for the RECS of a reaction with a constant reaction rate.[7] In this work, we generalize Gillespie's stochastic simulation method to investigate the RECS of a biopolymer reaction with rate coefficient fluctuation, which is described by the generalized master equation given in eq 12. Our simulation method is particularly useful when the propagator $G(\mathbf{r},t|\mathbf{r}_0)$ of the system state is available, and it gives the numerical results for the RECS of a reaction system with a state-dependent reaction rate coefficient much more efficiently compared to the stochastic simulation method involving an explicit simulation of the system dynamics in the state space. A detailed account of our stochastic simulation method is postponed to the Appendix.

In Figure 1, direct comparison is made between the analytic results, eqs 13 and 14, and our stochastic simulation results for a model of rate coefficient fluctuation, $k(t) = k_0 + \kappa r^2(t)$, where $r(t)$ is assumed to be the OU process with the time correlation function being given by $\langle r(t + t_0)\, r(t_0)\rangle_{eq} = b^2 \exp(-\lambda t/2)$. k_0 and κ are time-independent constants. As shown in Figure 1, the predictions of eqs 13 and 14 are in good agreement with the stochastic simulation results. The simulation results confirm that $\langle n(t)\rangle$ increases linearly in time, independent of the time scale, $\xi(\equiv \lambda^{-1})$, of the rate coefficient fluctuation but $\Delta(t)$ exhibits the kinetic phase transition predicted by eq 15. In Figure 1, the values of k_0 and κb^2 are chosen to be unity in a frequency unit. In accordance with eq 16, intercepts of curves at short times are given by ln $\langle \delta k^2\rangle/k_{eq}^2$, marked by diamonds in Figure 1b, whereas intercepts of the lines extrapolated from the curves at times much longer than ξ are given by $\ln(\langle \delta k^2\rangle/k_{eq}^2) + \ln(2k_{eq}\xi)$, marked by stars. This example demonstrates that one can estimate k_{eq} from the time dependence of the mean reaction number, i.e., $k_{eq} = \langle n(t)\rangle/t$; furthermore, one can separately estimate the variance $\langle \delta k^2\rangle$ and the relaxation time scale ξ of rate coefficient fluctuations from the dependence of ln $\Delta(t)$ on ln $\langle n(t)\rangle$.

We note that the time dependence of $\Delta(t)$ is independent of the value of equilibrium rate coefficient k_{eq} or the constant part k_0 of the fluctuating rate coefficient, $k(r) = k_0 + \kappa r^2$, in the reaction model. This is because $\delta(t)$ depends only on the fluctuation δk of the rate coefficient around its mean value, as given in eq 14. For this reason, the time dependence of $\Delta(t)$ for the model considered in Figure 1 is the same as that for the fluctuating bottleneck enzyme for which $k_0 = 0$. In contrast, the mean reaction number $\langle n(t)\rangle$ is dependent on k_0, as k_{eq} is given by $k_0 + \kappa b^2$ and eq 13 yields $\langle n(t)\rangle = (k_0 + \kappa b^2)t$.

DISCUSSION

In the previous section, we showed that the relationship, eq 15, between the mean and the variance of the number of biopolymer reaction events can be qualitatively different from eq 6 of a renewal process when the size, t, of measurement time is smaller than characteristic relaxation time ξ of rate coefficient fluctuations. Care must be taken in applying renewal statistics to an analysis of RECS for a biopolymer reaction with a fluctuating rate coefficient, as renewal theory is applicable only when the measurement time, t, is much greater than the characteristic relaxation time, ξ, of the rate coefficient fluctuation.[39] In Figure 2, we display simulated single molecule reaction trajectories and the time dependence of Mandel's Q parameter both at a short time regime and at a long time regime for the model considered in Figure 1. In the model, the reaction rate coefficient k is dependent on the conformational state variable, $r(t)$, of which relaxation to the equilibrium distribution occurs on time scale ξ. In the short time regime, where measurement time, t, is far smaller than the characteristic relaxation time scale, ξ, the distribution of conformational state

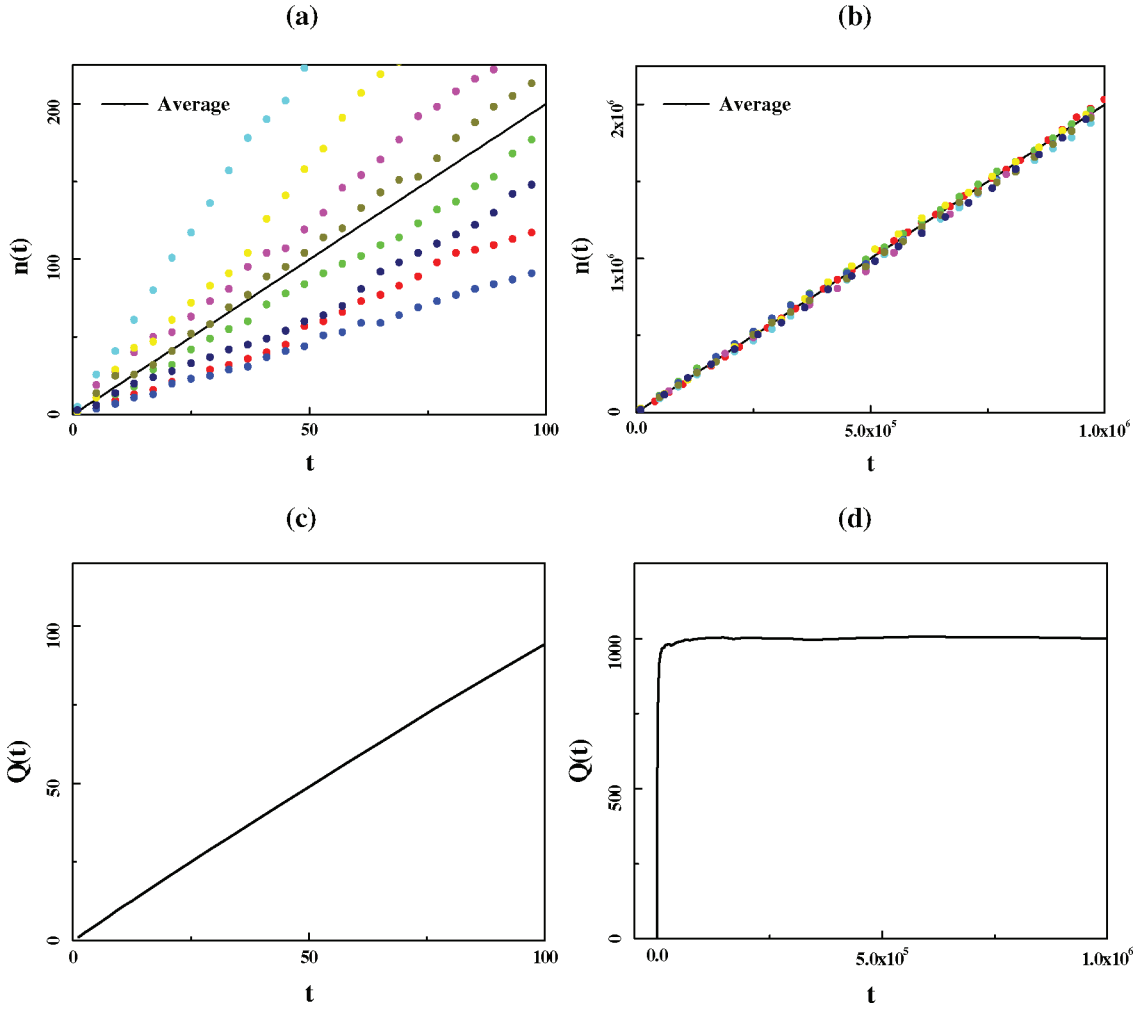

Figure 2. (Upper panels) Typical reaction event trajectories showing the number $n(t)$ of reaction events performed by an individual single biopolymer (a) in a short time regime with $t \ll \xi$ and (b) the same trajectories but in a long time regime with $t \gg \xi$. In the short time regime, single biopolymer reaction trajectories exhibit strong heterogeneity among the single biopolymers; in contrast, the single biopolymer reaction trajectories become statistically equivalent in time scales much longer than ξ. (Lower panels) Time dependence $Q(t)$ of Mandel's Q parameter $[\equiv -1 + (\langle n^2(t)\rangle - \langle n(t)\rangle^2)/\langle n(t)\rangle]$ (c) in the short time regime and (d) in the long time regime. Mandel's Q parameter is linearly proportional to the measurement time t in the short time regime, but it assumes a constant value at long times. Displayed are stochastic simulation results for the elementary biopolymer reaction model considered in Figure 1, with characteristic time scale $\xi(=\lambda^{-1})$ of the rate coefficient fluctuation being $500k_0^{-1}$. The unit of time, t, is k_0^{-1} as in Figure 1.

$r(t)$ of each single biopolymer can hardly span the entire conformational space of the biopolymer; instead, it is highly localized around its initial value $r(0)$ for each single biopolymer. In this sense, the single biopolymer system is a strongly nonergodic and heterogeneous system at the short time regime, where the value of rate coefficient $k(t)$ is nearly time-independent constant, approximately given by $k_0 + \kappa r^2(0)$, with the value of $r(0)$ being heterogeneous and distributed over single biopolymers. This is why each single molecule reaction trajectory has a seemingly different mean reaction number from the others in the short time regime shown in Figure 2a. In the latter heterogeneous and nonergodic reaction system, Mandel's Q defined by eq 5 or by $\Delta(t)/\langle n(t)\rangle$ is linearly proportional to measurement time t, $Q(t) \propto \langle n(t)\rangle \propto t$, according to eq 15a, but it does not approach a constant value, in contrast to the prediction, eq 6, of renewal theory. Such deviation from renewal statistics would emerge for any biological reaction system with a heterogeneous rate coefficient distribution when our measurement time is shorter than the relaxation time scale of the rate coefficient fluctuation of the system.

On the other hand, when our measurement time, t, is much greater than the relaxation time scale, ξ, of the rate coefficient fluctuation, the value of Mandel's Q becomes constant according to eq 15b, as shown in Figure 2d, which conforms to eq 6 obtained from renewal theory. In the long time regime where $t \gg \xi$, the distribution of the conformational state variable, $r(t)$, of each single biopolymer spans the entire conformational space of the single biopolymer and relaxes to the same equilibrium distribution irrespective of the initial state of the single biopolymer; consequently, the distribution of the conformation-dependent rate coefficient fluctuation becomes the equivalent for every single molecule in the system. In the long time region where single biopolymers become an ergodic and homogeneous system, heterogeneity among single biopolymer reaction trajectories becomes insignificant, as shown in Figure 2b, so that RECS extracted from one single molecule reaction trajectory should be equivalent to that extracted from another.

Relaxation of the reactivity fluctuation of biomolecules can occur in a wide range of time scales.[9] When the biopolymer reaction is coupled to a number of dynamic degrees of freedom,

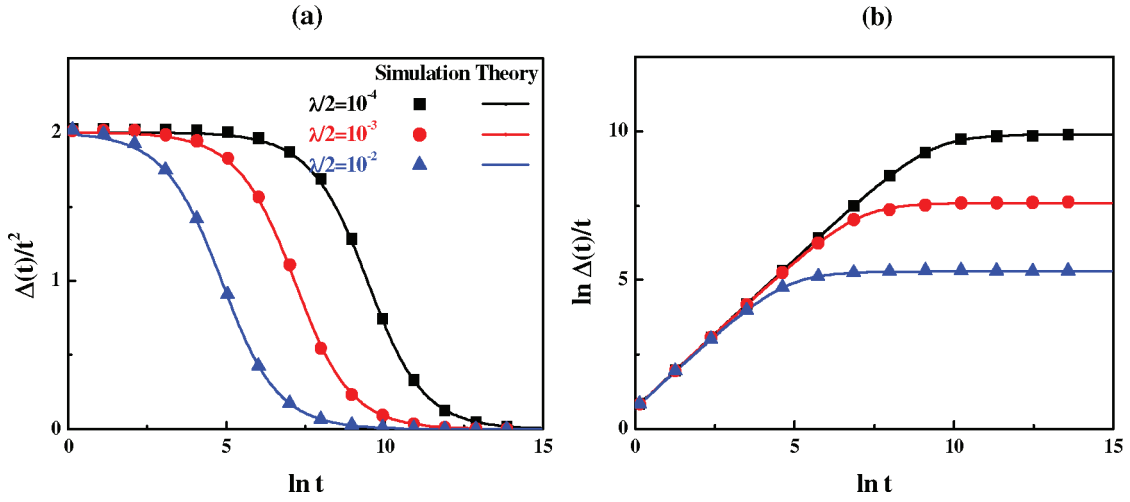

Figure 3. Time dependence of $\Delta(t)/t^2$ and $\Delta(t)/t$. The displayed results and data are the same as those shown in Figure 1b. The short time asymptotic value of $\Delta(t)/t^2$ coincides with the variance $\langle \delta k^2 \rangle$ of the rate coefficient fluctuation, marked by diamonds in Figure 1b. The long time asymptotic value of $\Delta(t)/t$ is the same as that of 2 $\langle \delta k^2 \rangle \langle \xi \rangle$ with $\langle \xi \rangle$ the mean relaxation time of the rate coefficient fluctuation.

the relaxation of the rate coefficient fluctuation can occur on multiple time scales and the time correlation function $\phi_k(t)[= \langle \delta k(t)\ \delta k(0) \rangle_{eq}/\langle \delta k^2 \rangle_{eq}]$ of the rate coefficient fluctuation is given by a multiexponential function: $\phi_k(t) = \sum_{j=1}^{M} f_j \exp(-\lambda_j t)$, where M is the number of the rate coefficient fluctuation modes and f_j accounts for the relative contribution of the jth rate coefficient fluctuation mode with relaxation rate λ_j to the variance of the rate coefficient fluctuation, δk. f_j satisfies the normalization condition, $\sum_{j=1}^{M} f_j = 1$. When the reactivity fluctuation occurs on multiple time scales, the expression for $\Delta(t)$ in eq 14 becomes

$$\Delta(t) = \langle \delta k^2 \rangle t^2 \sum_j f_j g(\lambda_i t) \tag{19}$$

Noting that $g(x) \cong 1$ for $x \ll 1$ but $g(x) \cong 2/x$ for $x \gg 1$, we obtain the following expression for $\Delta(t)$ from eq 19:

$$\Delta(t) = \begin{cases} \langle \delta k^2 \rangle t^2 & (t \ll \xi_M) \\ 2t \langle \delta k^2 \rangle \langle \xi \rangle & (t \gg \xi_1) \end{cases} \tag{20}$$

Here, ξ_j denotes the characteristic relaxation time λ_j^{-1} of the jth rate coefficient fluctuation mode ($\lambda_1 \le \lambda_2 \le ... \le \lambda_M$), and $\langle \xi \rangle$ denotes the average relaxation time defined by $\langle \xi \rangle = \sum_{j=1}^{M} f_j \xi_j$. Equation 20 tells us that $\langle \delta k^2 \rangle$ can be estimated from the initial value of $\Delta(t)/t^2$, i.e., $\lim_{t \to 0} \Delta(t)/t^2 = \langle \delta k^2 \rangle$, and $\langle \xi \rangle$ can be estimated by the long time limit value of $\Delta(t)/t$, i.e., $\lim_{t \to \infty} \Delta(t)/t = 2 \langle \delta k^2 \rangle \langle \xi \rangle$.

In Figure 3, we display the time dependence of $\Delta(t)/t^2$ and $\Delta(t)/t$ for the same model considered in Figure 1. As shown in Figure 3a, the value of $\langle \delta k^2 \rangle$ is the same as the intercept of the curves showing the time dependence of $\Delta(t)/t^2$. In addition, the long-time asymptotic value of $\Delta(t)/t$ coincides with the value of 2 $\langle \delta k^2 \rangle \langle \xi \rangle$ as shown in Figure 3b. We note that, in the presence of time scale separation between rate relaxation modes, $\Delta(t)/t^2[\equiv F(t)]$ can be interpreted as the fraction of the rate coefficient fluctuation modes in which relaxation occurs at times longer than t. In this case, eq 19 can be approximated as

$$F(t)/F(0) = \sum_{j=1}^{M} f_j g(t/\xi_j) \approx \begin{cases} 1 & (t \ll \xi_M) \\ \sum_{j=1}^{m^*} f_j + \frac{2}{t} \sum_{j=m^*+1}^{M} f_j \xi_j & (\xi_{m^*+1} \ll t \ll \xi_{m^*}) \\ 0 & (t \gg \xi_1) \end{cases} \tag{21}$$

where m^* denotes the number of rate coefficient fluctuation modes of which the relaxation time constant is greater than the size t of measurement time. Equation 20 indicates, given that t is much greater than ξ_{m^*+1} but much smaller than ξ_{m^*}, $F(t)/F(0)$ becomes a good approximation for $\sum_{j=1}^{m^*} f_j$, the fraction of the rate coefficient fluctuation modes of which relaxation does not occur significantly in time t. We also note that the time dependence of the rate−rate autocorrelation function $\langle \delta k(t)\ \delta k(0) \rangle_{eq}$ can be obtained from the variance, $\sigma_n^2(t)$, in the RECS:

$$\left\langle \delta k(t)\ \delta k(0) \right\rangle_{eq} = 2^{-1} d^2 \Delta(t)/dt^2 = 2^{-1} d^2 \sigma_n^2(t)/dt^2 \tag{22}$$

which follows from eq 14.

Up to now, we have focused on the first two moments of the RECS of the elementary biopolymer reaction. For a higher moment of the RECS, we obtain the following expression from the generalized master equation given in eq 12:[25,26]

$$\langle n(n-1)...(n-l+1)(t) \rangle_{eq} = \int_0^t d\tau C_{l-1}(t) \tag{23}$$

where

$$C_0(t) = k_{eq}$$

$$C_1(t) = \int_0^t d\tau_1 \langle k(\tau_1) k(0) \rangle_{eq}$$

$$C_l(t) = \int_0^t d\tau_l \int_0^{\tau_l} d\tau_{l-1} ... \int_0^{\tau_2} d\tau_1 \langle k(\tau_l)\ k(\tau_{l-1}) ... k(\tau_1)\ k(0) \rangle_{eq} \qquad (l \ge 2) \tag{24}$$

dx.doi.org/10.1021/ct200785q | J. Chem. TheoryComput. 2012, 8, 1415−1425

Here, $\langle k(\tau_l)\, k(\tau_{l-1})...k(\tau_1)\, k(0)\rangle_{eq}$ designates the multitime rate autocorrelation function defined by

$$\langle k(\tau_l)\, k(\tau_{l-1})...k(\tau_1)\, k(0)\rangle_{eq} = \int d\mathbf{r}_l...d\mathbf{r}_0 \left[\prod_{j=1}^{l} k(\mathbf{r}_j)\, G(\mathbf{r}_j, \tau_j - \tau_{j-1}|\mathbf{r}_{j-1})\right] k(\mathbf{r}_0)\, P_{eq}(\mathbf{r}_0) \quad (25)$$

with τ_0 being equal to 0.

The result of the present work is exact only when the reaction process of biopolymers in system state $\mathbf{r}$ is a rate process or a Poisson process with conformation dependent reaction rate $k(\mathbf{r})$. When the reaction process of the biopolymer in a given system state $\mathbf{r}$ involves multiple reaction channels or multiple transition states, its stochastic property can be different from that of the Poisson process. In this case, the results of the present work are only approximate.[25] The approximation would work better when the time scale of the conformational dynamics of the biopolymer is longer than that of the individual reaction event. Generalization of the present work for the non-Poisson biopolymer reaction involving reversible binding of substrate molecules is currently undergoing.

So far, we have assumed that the initial distribution $f_0(\mathbf{r})$ of the system state $\mathbf{r}$ is the equilibrium distribution $f_{eq}(\mathbf{r})$ that satisfies $L(\mathbf{r})\, f_{eq}(\mathbf{r}) = 0$, so the results presented above are applicable to the RECS of a reaction system composed of a number of catalytic biopolymers in the equilibrium state. They are also applicable to the RECS of a single biopolymer reaction system if the initial state of the single biopolymer reaction system is sampled according to the equilibrium distribution. Upon the analysis of RECS of a single molecule reaction trajectory, $\langle n^k(t)\rangle$ should be identified as $\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} n^k(t;\, t_i)$ where N and $n(t;t_i)$ respectively denote the number of the measurement intervals and the number of single molecule reaction events occurring in the ith measurement interval $(t_i, t_i + t)$. If the single molecule system is ergodic and we sample initial measurement time t_i uniformly over a sufficiently long reaction trajectory, the system state at initial measurement time t_i would distribute according to the equilibrium distribution.

Upon analysis of a single molecule reaction trajectory, distribution $f_0(\mathbf{r})$ of the system state at initial measurement time t_i can deviate from the equilibrium distribution, unless t_i is sampled uniformly in time. For the reaction system with a nonequilibrium initial state distribution $f_0(\mathbf{r})$, the following expression for the moments of reaction event number distribution can be obtained from the generalized master equation given in eq 12:[25]

$$\langle n(n-1)...(n-l+1)(t)\rangle_0 = \mathcal{L}^{-1}\left\{\frac{l!}{u}\int d\mathbf{r}_l\, d\mathbf{r}_{l-1}...d\mathbf{r}_0 \left[\prod_{j=1}^{l} k(\mathbf{r}_j)\, \hat{G}(\mathbf{r}_j, u|\mathbf{r}_{j-1})\right] f_0(\mathbf{r}_0)\right\} \quad (26)$$

Here, $\mathcal{L}^{-1}$ denotes the inverse Laplace transform operator, and $\hat{G}(\mathbf{r}_j, u|\mathbf{r}_{j-1})$ denotes the Laplace transform of propagator $G(\mathbf{r}_j, t|\mathbf{r}_{j-1})$ defined below eq 14. A particularly important nonequilibrium initial state distribution is the distribution $f_0^*(\mathbf{r})$ of the single molecule system state at time t_i with t_i being sampled only at the moment where each single molecule reaction begins. The expression for $f_0^*(\mathbf{r})$ is given by[26,40]

$$f_0^*(\mathbf{r}) \equiv k(\mathbf{r})\, f_{eq}(\mathbf{r}) / \int d\mathbf{r}'\, k(\mathbf{r}')\, f_{eq}(\mathbf{r}') \quad (27)$$

For the latter initial condition, eq 26 yields

$$\left\langle n(n-1)...(n-l+1)(t)\right\rangle_0^* = k_{eq}^{-l}\, C_l(t) \quad (28)$$

where $C_l(t)$ is given in eq 24. Comparing eq 23 and eq 28, we obtain the following remarkable relation

$$\langle n(n-1)...(n-l+1)(t)\rangle_0^* = k_{eq}^{-l}\frac{d}{dt}\left\langle n(n-1)...(n-l)(t)\right\rangle_{eq} \quad (29)$$

between the RECS of the elementary biopolymer reaction with two different initial state distributions, $f^*_0(\mathbf{r})$ and $f_{eq}(\mathbf{r})$. Equation 29 with l being equal to 1 indicates that the information about the rate−rate autocorrelation function contained in $\langle n(n-1)(t)\rangle_{eq}$ can be obtained directly from the mean reaction number $\langle n(t)\rangle_0^*$ of RECS for the case with nonequilibrium initial state distribution $f^*_0(\mathbf{r})$:

$$\begin{aligned}\langle n(t)\rangle_0^* &= k_{eq}^{-1}\frac{d}{dt}\langle n(n-1)(t)\rangle_{eq} \\ &= k_{eq}^{-1}\int_0^t d\tau\left\langle k(\tau)\, k(0)\right\rangle_{eq} \\ &= k_{eq}\left(t + k_{eq}^{-2}\int_0^t d\tau\left\langle \delta k(\tau)\delta k(0)\right\rangle_{eq}\right)\end{aligned} \quad (30)$$

Note that $\langle n(t)\rangle_0^*$ is nonlinear in time in contrast to the mean reaction number $\langle n(t)\rangle_{eq}$ given in eq 13; the time derivative of $\langle n(t)\rangle_0^*$ is a monotonically decreasing function of the size t of measurement time:

$$\frac{d\langle n(t)\rangle_0^*}{dt} = k_{eq}\left[1 + \frac{\left\langle \delta k(t)\, \delta k(0)\right\rangle_{eq}}{k_{eq}^2}\right] \quad (31)$$

which varies from the initial value, $k_{eq} + \langle \delta k^2\rangle/k_{eq}$, to the final value, k_{eq} the equilibrium reaction rate. The time derivative of $\langle n(t)\rangle_0^*$ is the same as the exact result for a two reaction event density, previously reported by Yang and Cao.[24]

CONCLUSION

We present the exact expression for the mean and the higher moments of the reaction event number distribution for a biomolecule reaction in which rate coefficient $k(\mathbf{r})$ fluctuates in line with its conformational dynamics, $\mathbf{r}(t)$. We also generalize Gillespie's stochastic simulation method for numerical investigation of RECS of a reaction with slow reactivity fluctuation and confirm that the simulation results for RECS are in good agreement with the prediction of the analytic results obtained for a general biopolymer reaction model. It is found that the difference $\Delta(t)$ of the variance $\sigma_n^2(t)$ from the mean $\langle n(t)\rangle$ of the reaction event number occurred in measurement time t can serve as a useful probe of the conformation dependent rate coefficient fluctuation of a biopolymer. Our results predict a universal kinetic phase transition in the fluctuation of reaction event statistics of a biopolymer. Given that the initial state distribution of the reaction system is the equilibrium distribution, difference $\Delta(t)$ of the variance $\sigma_n^2(t)$ from the mean $\langle n(t)\rangle$ of the number of reaction events occurring in time t is proportional to t^2 or $\langle n(t)\rangle^2$ at short times, which does not conform to renewal theory. However, it becomes proportional

to t or $\langle n(t)\rangle$ at time scale much longer than the time scale of the rate coefficient fluctuation. The long time asymptotic behavior of the mean and the variance in the reaction event counting statistics for the reaction with fluctuating rate $k(\mathbf{r}(t))$ is the same as that of the mean and the variance of the renewal reaction process whose mean reaction time $\langle\tau\rangle$ and the randomness parameter $R = (\langle\tau^2\rangle - 2\langle\tau\rangle^2)/\langle\tau\rangle^2$ are given by k_{eq}^{-1} and $2k_{eq}^{-1}\int_0^\infty dt\langle\delta k(t)\ \delta k(0)\rangle$, respectively. From the short time behavior of $\Delta(t)$, one can extract the variance $\langle\delta k^2\rangle_{eq}$ of the rate coefficient fluctuation; in comparison, one can obtain the relaxation time $\xi(\equiv \int_0^\infty dt\ \langle\delta k(t)\ \delta k(0)\rangle/\langle\delta k^2\rangle)$ of the rate coefficient fluctuations from the long time behavior of $\Delta(t)$. The time profile of the time correlation function $\langle\delta k(t)\ \delta k(0)\rangle_{eq}$ of the rate coefficient fluctuation can be obtained from $\Delta(t)$ with use of the following relation: $\langle\delta k(t)\ \delta k(0)\rangle_{eq} = 2^{-1}d^2\Delta(t)/dt^2$. The value of $\Delta(t)/t^2$ scaled by its initial value estimates the fraction of the rate coefficient fluctuation modes of which relaxation does not occur in time t. On the basis of the obtained results, we propose a method of quantitative analysis for the first two moments of the RECS of a reaction system with an arbitrary rate coefficient fluctuation, which enables one to extract object information about the magnitude of fluctuation and the relaxation times of the fluctuating reaction rate, without a bias that can be introduced by assuming a particular model of conformational dynamics and the conformation dependent reactivity. The general relationship between the higher moments of RECS and the time correlation function of the reaction rate coefficient is also established for a nonequilibrium initial state.

APPENDIX

Here, we present an account of our stochastic simulation method for a reaction system of which the reaction rate coefficient is a function, $k(\mathbf{r})$, of system state $\mathbf{r}$. In general, system state $\mathbf{r}$ is a vector in multidimensional state space so that an explicit simulation of the system state dynamics in addition to the reaction process costs a lot of computation time. However, when the knowledge about the propagator $G(\mathbf{r},t|\mathbf{r}_0)$ of the system state is available, one can circumvent the problem by focusing on the reaction process of the system with the time sequence of the system state being generated from the conditional probability, $G(\mathbf{r},t|\mathbf{r}_0)$. The following is the algorithm of our stochastic simulation method.

1. Sample the initial system state $\mathbf{r}_0$ according to initial distribution $f_0(\mathbf{r}_0)$.
2. Calculate the rate coefficient $k(\mathbf{r}_0)$ at time 0.
3. Sample the reaction time t_1 for the first reaction event from the reaction time distribution $\phi_1(t)$ with rate coefficient $k(\mathbf{r}_0)$, i.e., $\phi_1(t_1) = k(\mathbf{r}_0)\exp[-k(\mathbf{r}_0)t_1]$.
4. Sample the system state $\mathbf{r}_1$ at time t_1 according to the conditional probability $G(\mathbf{r}_1,t_1|\mathbf{r}_0)$.
5. Calculate the reaction rate $k(\mathbf{r}_1)$ at time t_1.
6. Sample the reaction time t_j for the jth reaction event from the reaction time distribution $\phi_j(t)$ with rate coefficient $k(\mathbf{r}_{j-1})$, i.e., $\phi_j(t_j) = k(\mathbf{r}_{j-1})\exp[-k(\mathbf{r}_{j-1})t]$ (for $j \geq 2$).
7. Sample the system state $\mathbf{r}_j$ at time $t_1 + t_2 + \ldots + t_j$ according to the conditional probability $G(\mathbf{r}_j t_j|\mathbf{r}_{j-1})$.
8. Calculate the reaction rate coefficient $k(\mathbf{r}_j)$ at time $t_1 + t_2 + \ldots + t_j$.

From the above-mentioned algorithm, we obtain a sequence of stochastic reaction times series, $\{t_1,t_2, \ldots, t_j, \ldots\}$, from which the number $n(t)$ of reaction events occurring in measurement time interval $(0, t)$ can be determined, i.e., $n(t) = j$ if $t_j \leq t < t_{j+1}$. Then, the lth moment of the number of reaction events occurring in time interval $(0, t)$ can be estimated by performing average of $n^l(t)$ over the different reaction time trajectories.

This algorithm allows one to investigate the RECS of a reaction system with a state-dependent rate coefficient efficiently, without having to spend much computational time in direct simulation of system state dynamics, $\mathbf{r}$. However, the simulation algorithm is applicable to the case where the time scale of the rate coefficient fluctuation is much longer than that of the individual reaction event. For the case where the time scale of the rate coefficient fluctuation is comparable to that of the individual reaction event, one has to generalize the present simulation method by employing the suitable reaction time distribution $\phi_j(t_j)$ for the jth reaction process, different from the Poisson reaction time distribution $k(\mathbf{r}_{j-1})\exp[-k(\mathbf{r}_{j-1})t]$ used in the above algorithm, which should be dependent not only on the system state $\mathbf{r}_{j-1}$ at the time the jth reaction begins but also on the system dynamics during the reaction event. Generalization of the present simulation algorithm to a state-dependent non-Poisson reaction process is underway.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jaeyoung@cau.ac.kr, jianshu@mit.edu.

Present Addresses

[§]On sabbatical leave at Massachusetts Institute of Technology.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

J. Sung acknowledges helpful discussions with Dr. C. Hyeon. This work was supported by the Korea Research Foundation Grant (KRF-C00180), the National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST) (2011-0016412), and Priority Research Centers Program through the National Research Foundation of Korea (NRF) (2009-0093817). J. Cao acknowledges the support by the Singapore-MIT Alliance for Research and Technology (SMART) and DOD ARO grant W911NF-09-0480. J. Sung gratefully acknowledges Professors Robert J. Silbey, Irwin Oppenheim, and Jianshu Cao for their hospitality and support.

DEDICATION

[‖]Deceased on October 27, 2011. We dedicate this work to his memory.

REFERENCES

(1) Balazsi, G.; Oudenaarden, A. v.; Collins, J. J. Cellular Decision Making and Biological Noise: From Microbes to Mammals. *Cell* **2011**, *144* (6), 910−925.

(2) Raj, A.; Oudenaarden, A. v. Nature, Nurture or Chances: Stochastic Gene Expression and Its Consequences. *Cell* **2008**, *135*, 216−226.

(3) Maamar, H.; Raj, A.; Dubnau, D. Noise in gene expression determines cell fate in Bacillus subtillis. *Science* **2007**, *317*, 526−529.

(4) Maheshri, N.; O'Shea, E. K. Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 413−434.

(5) Rao, C. V.; Wolf, D. M.; Arkin, A. P. Control, exploitation and tolerance of intracellular noise. *Nature* **2002**, *420*, 231−237.

(6) Kampen, N. G. V. *Stochastic Processes in Physics and Chemistry*; North-Holland: Amsterdam, 2007; pp 96−192.
(7) Gillespie, D. T. Exact Stochastic Simulation of Coupled Chemical Reactions. *J. Phys. Chem.* **1977**, *81*, 2340−2361.
(8) Velonia, K.; Flomenbom, O.; Loos, D.; Masuo, S.; Cotlet, M.; Engelborghs, Y.; Hofkens, J.; Rowan, A. E.; Klafter, J.; Nolte, R. J. M.; De Schryver, F. C. Single-Enzyme Kinetics of CALB-Catalyzed Hydrolysis. *Angew. Chem., Int. Ed.* **2005**, *44*, 560−564.
(9) English, B. P.; Min, W.; van Oijen, A. M.; Lee, K. T.; Luo, G.; Sun, H.; Cherayil, B. J.; Xie, X. S. Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. *Nat. Chem. Biol.* **2006**, *2*, 87−94.
(10) Elowitz, M. B.; Levine, A. J.; Siggia, E. D.; Swain, P. S. Stochastic gene expression in a single cell. *Science* **2002**, *297*, 1183−1186.
(11) Asbury, C. L.; Fehr, A. N.; Block, S. M. Kinesin Moves by an Asymmetric Hand-Over-Hand Mechanism. *Science* **2003**, *302*, 2130−2134.
(12) Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B.; Bibillo, A.; Bjornson, K.; Chaudhuri, B.; Christians, F.; Cicero, R.; Clark, S.; Dalal, R.; deWinter, A.; Dixon, J.; Foquet, M.; Gaertner, A.; Hardenbol, P.; Heiner, C.; Hester, K.; Holden, D.; Kearns, G.; Kong, X.; Kuse, R.; Lacroix, Y.; Lin, S.; Lundquist, P.; Ma, C.; Marks, P.; Maxham, M.; Murphy, D.; Park, I.; Pham, T.; Phillips, M.; Roy, J.; Sebra, R.; Shen, G.; Sorenson, J.; Tomaney, A.; Travers, K.; Trulson, M.; Vieceli, J.; Wegener, J.; Wu, D.; Yang, A.; Zaccarin, D.; Zhao, P.; Zhong, F.; Korlach, J.; Turner, S. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **2009**, *323*, 133−138.
(13) Clarke, J.; Wu, H. C.; Jayasinghe, L.; Patel, A.; Reid, S.; Bayley, H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* **2009**, *4* (4), 265−270.
(14) Visscher, K.; Schnitzer, M. J.; Block, S. M. Single kinesin molecules studied with a molecular force clamp. *Nature* **1999**, *400*, 184−189.
(15) Min, W.; Gopich, I. V.; English, B. P.; Kou, S. C.; Xie, X. S.; Szabo, A. When Does the Michaelis-Menten Equation Hold for Fluctuating Enzymes? *J. Phys. Chem. B* **2006**, *110*, 20093−20097.
(16) Cao, J.; Silbey, R. J. Generic Schemes for Single-Molecule Kinetics. 1: Self-Consistent Pathway Solutions for Renewal Processes. *J. Phys. Chem. B* **2008**, *112*, 12867−12880.
(17) Jung, W.; Yang, S.; Sung, J. Novel Chemical Kinetics for a Single Enzyme Reaction: Relationship between Substrate Concentration and the Second Moment of Enzyme Reaction Time. *J. Phys. Chem. B* **2010**, *114*, 9840−9847.
(18) Cao, J. Michaelis-Menten Equation and Detailed Balance in Enzymatic Networks. *J. Phys. Chem. B* **2011**, *115*, 5493−5498.
(19) Fisher, M. E.; Kolomeisky, A. B. Simple mechanochemistry describes the dynamics of kinesin molecules. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 7748−7753.
(20) Yang, S.; Cao, J.; Silbey, R. J.; Sung, J. Quantitative Interpretation of the Randomness in Single Enzyme Turnover Times. *Biophys. J.* **2011**, *101*, 519−524.
(21) Sung, J.; Silbey, R. J. Exact Dynamics of a Continuous Time Random Walker in the Presence of a Boundary: Beyond the Intuitive Boundary Condition Approach. *Phys. Rev. Lett.* **2003**, *91*, 160601.
(22) Lu, H. P.; Xun, L.; Xie, X. S. Single-Molecule Enzymatic Dynamics. *Science* **1998**, *282*, 1877−1882.
(23) Yang, H.; Xie, X. S. Statistical Approaches for Probing single-molecule dynamics photon by photon. *Chem. Phys.* **2002**, *284*, 423−437.
(24) Yang, S.; Cao, J. Direct measurements of memory effects in single-molecule kinetics. *J. Chem. Phys.* **2002**, *117*, 10996−11009.
(25) Sung, J.; Silbey, R. J. Counting statistics of single molecule reaction events and reaction dynamics of a single molecule. *Chem. Phys. Lett.* **2005**, *415*, 10−14.
(26) Cao, J. Correlations in single molecule photon statistics: renewal indicator. *J. Phy. Chem. B* **2006**, *110* (38), 19040−19043.
(27) Gopich, I. V.; Szabo, A. Theory of the statistics of kinetic transitions with application to single-molecule enzyme catalysis. *J. Chem. Phys.* **2006**, *124*, 154712−154733.
(28) Jung, Y.; Barkai, E.; Silbey, R. J. A stochastic theory of single molecule spectroscopy. *Adv. Chem. Phys.* **2002**, *123*, 199−266.
(29) Zheng, Y.; Brown, F. L. H. Single-Molecule Photon Counting Statistics via Generalized Optical Bloch Equations. *Phys. Rev. Lett.* **2003**, *90*, 238305.
(30) Barkai, E.; Jung, Y.; Silbey, R. J. Theory of single molecule spectroscopy: Beyond the ensemble average. *Annu. Rev. Phys. Chem.* **2004**, *55*, 457−507.
(31) Shikerman, F.; Barkai, E. Photon Statistics for Single-Molecule Nonlinear Spectroscopy. *Phys. Rev. Lett.* **2007**, *99*, 208302.
(32) Gopich, I.; Szabo, A. Theory of photon statistics in single-molecule Forster resonance energy transfer. *J. Chem. Phys.* **2005**, *122*, 014707.
(33) Bel, G.; Brown, F. L. Theory for Wavelength-Resolved Photon Emission Statistics in Single-Molecule Fluorescence Spectroscopy. *Phys. Rev. Lett.* **2009**, *102*, 018303.
(34) Brown, F. L. H. Single-Molecule Kinetics with Time-Dependent Rates: A Generating Function Approach. *Phys. Rev. Lett.* **2003**, *90*, 028302.
(35) Zwanzig, R. Dynamical disorder: Passage through a fluctuating bottleneck. *J. Chem. Phys.* **1992**, *97*, 3587−3589.
(36) Cox, D. R. *Renewal Theory*; Methuen: London, 1962; pp 25−60.
(37) Zwanzig, R. Rate Processes with Dynamical Disorder. *Acc. Chem. Res.* **1990**, *23*, 148−152.
(38) Uhlenbeck, G. E.; Ornstein, L. S. On the theory of Brownian Motion. *Phys. Rev.* **1930**, *36*, 823−841.
(39) Schnitzer, M. J.; Block, S. M. Statistical Kinetics of Processive Enzymes. *Cold Spring Harbor Symp. Quant. Biol.* **1995**, *60*, 793−802.
(40) Cao, J. Event-averaged measurements of single molecule kinetics. *Chem. Phys. Lett.* **2000**, *327*, 38−44.