Statistical Kinetics of Macromolecular Dynamics

Joshua W. Shaevitz,* Steven M. Block,[†‡] and Mark J. Schnitzer[†‡]
*Departments of Physics, [†]Biological Sciences, and [‡]Applied Physics, Stanford University, Stanford, California 94305-5020

ABSTRACT Fluctuations in biochemical processes can provide insights into the underlying kinetics beyond what can be gleaned from studies of average rates alone. Historically, analysis of fluctuating transmembrane currents supplied information about ion channel conductance states and lifetimes before single-channel recording techniques emerged. More recently, fluctuation analysis has helped to define mechanochemical pathways and coupling ratios for the motor protein kinesin as well as to probe the contributions of static and dynamic disorder to the kinetics of single enzymes. As growing numbers of assays are developed for enzymatic or folding behaviors of single macromolecules, the range of applications for fluctuation analysis increases. To evaluate specific biochemical models against experimental data, one needs to predict analytically the distribution of times required for completion of each reaction pathway. Unfortunately, using traditional methods, such calculations can be challenging for pathways of even modest complexity. Here, we derive an exact expression for the distribution of completion times for an arbitrary pathway with a finite number of states, using a recursive method to solve algebraically for the appropriate moment-generating function. To facilitate comparisons with experiments on processive motor proteins, we develop a theoretical formalism for the randomness parameter, a dimensionless measure of the variance in motor output. We derive the randomness for motors that take steps of variable sizes or that move on heterogeneous substrates, and then discuss possible applications to enzymes such as RNA polymerase, which transcribes varying DNA sequences, and to myosin V and cytoplasmic dynein, which may advance by variable increments.

INTRODUCTION

Stochastic fluctuations in biochemical processes reflect the kinetic structure of the underlying reaction pathways. For example, two processes with identical average rates may nevertheless display widely different variances in the time to completion. The variance in completion time depends on the numbers and lifetimes of intermediate states in the pathway. Studies of biochemical kinetics commonly seek to compare distributions of completion times to theoretical predictions from model pathways. Such comparisons can help to determine the rates for intermediate steps and to exclude incompatible models. However, for bulk solution studies, inferring the distribution of completion times requires some experimental means to initiate the reaction synchronously, and thereafter to monitor continual progress toward completion. As a practical matter, it is not always feasible to meet both of these requirements using traditional biochemical approaches.

In contrast, modern studies of single molecules (or small numbers of molecules) offer convenient experimental access to reaction time distributions. When individual turnovers can be resolved with adequate time resolution and in sufficient numbers, the distribution of times can be constructed directly. The advent of methods for observing single-molecule behavior in increasing detail has facilitated this direct approach. For example, single-channel recordings enable measurements of ionic conductance states with impressive precision ((1) and references therein). Advanced fluorescence techniques have enabled studies of folding in individual nucleic acids and proteins (2–6), of static and dynamic disorder in single enzyme kinetics (7,8), and of single chaperone function (9). For single-molecule studies of motor proteins, nanometric measurements have enabled the direct detection of enzyme translations of <1 nm and rotations of $<10°$ (10,11) with excellent temporal resolution, out to bandwidths exceeding 10 kHz (12). Experiments with optical and magnetic tweezers have shown that motor proteins move in discrete increments: linear motors, such as kinesin (13), dynein (14), and myosin (15,16), all take nanometer-sized steps, whereas rotary motors, such as the F_1-ATPase, are driven in distinct angular steps by ATP hydrolysis (17). For some of these mechanoenzymes, step-time distributions can be constructed directly from single-molecule records, then compared to predictions from kinetic models (11,18,19).

Even when individual events are obscured by noise and cannot be observed directly, stochastic fluctuations in molecular behavior can nevertheless supply information about the structure of the underlying kinetic pathways and intermediate reaction rates. This capacity to address kinetic issues without resolving single events provided the key to early estimates of ion channel conductance values and state lifetimes, based on statistical properties of the electrical noise in recordings from nerve cells (20). Likewise, under conditions where the steps taken by single-motor proteins cannot be detected reliably (e.g., at lower loads, or under large compliance), variance measurements have been used to evaluate candidate kinetic models (19,21–23).

Although fluctuation analysis is a valuable tool for experimental characterization of kinetics, analytical calculation of the distribution of completion times for a model

Submitted April 8, 2005, and accepted for publication July 5, 2005.

Address reprint requests to Steven M. Block, E-mail: sblock@stanford.edu.

0006-3495/05/10/2277/09 $2.00

doi: 10.1529/biophysj.105.064295

pathway with a complex reaction scheme can be an arduous task. Nonetheless, analytical solutions are often more valuable than answers achieved by numerical or Monte Carlo approaches because of the insight that analytical methods can provide into the dependence of kinetic quantities on input parameters. For example, analytical solutions can be especially good for examination of limiting cases.

To date, several mathematical approaches have been developed to find the distribution of reaction times. The first, a classical master-equation approach, involves solving the full set of coupled differential equations that describe the time-varying changes in concentration of each state in the pathway (this is an eigenvalue problem). An extensive literature on the formal solution of such equations has been developed for the study of reaction kinetics (7,24,26). In many cases, these must be solved numerically (26). A second approach involves finding the moment-generating function for the completion time distribution by forming a weighted sum of all possible trajectories between states in the kinetic pathway (27). For all but simple, unidirectional pathways, the generating-function method requires the evaluation of infinite sums over trajectories. Unfortunately, such sums are often nested, hindering the development of simple, diagrammatic rules for their separate computation. Here, we introduce a third approach that improves upon the previous methodology. We show that the desired moment-generating function can be calculated exactly, using a recursive definition that avoids infinite sums. This method yields a set of coupled algebraic equations that can be solved analytically for pathways containing finite numbers of states with arbitrarily complex reaction schemes. The ability to calculate the distribution of completion times analytically for specific models should facilitate quantitative comparisons with experimental data for a variety of biological processes, including ion channel dynamics, macromolecular folding, and enzymatic reactions.

A convenient variable characterizing fluctuations in enzymatic behavior is the randomness parameter, r, a dimensionless measure of the variance in kinetic output. For processive mechanoenzymes, the randomness parameter supplies a measure of the variance in the motor position about its average trajectory, which depends upon both the distribution of times required to complete each stepping cycle and on the distribution of displacements of the individual steps. However, previous theoretical work on the randomness parameter treated only a few comparatively simple cases analytically, involving kinetic pathways for motors moving along periodic substrates with mechanical displacements that are low multiples of an invariant step size, or else resorted to numerical solutions of the kinetics (27,28–31). Based on a growing body of experimental evidence, certain motor proteins seem poorly described by the simplest models, and may therefore require a more complex description of the relationship between mechanical step size and biochemical cycle time. Here, we develop a general expression for the randomness parameter for the case of both nonuniform step times and step sizes, and we discuss the implications for motors that move with variable-sized displacements. We also treat the case of motors that move along a heterogeneous substrate. Unlike kinesin or myosin motors, processive nucleic acid enzymes travel along a continuously varying DNA or RNA substrate. For such enzymes, which include polymerases, nucleases, and helicases, the step-time distributions are governed by reaction pathways that presumably depend upon details of the local nucleic acid sequence. We model the simplest scenario for substrate heterogeneity, by assuming that for each stepwise displacement, the enzyme traverses one of four distinct biochemical pathways corresponding to each of the four possible DNA bases encountered, and we explore the consequences of substrate heterogeneity on the step-time distribution and the randomness parameter.

THEORETICAL RESULTS

Recursive calculation of the step-time distribution

Consider the simple pathway shown in Fig. 1 *A*, where the system begins in State 1 and a cycle is completed when State 3 is reached. The cycle-time distribution is equivalent to the distribution of first-passage times to State 3. When the system is in State 2, kinetic partitioning governs whether the following transition will be forward to State 3 or backward to State 1. Assuming that all transitions are Markovian, each visit is independent of any previous visits. This lack of

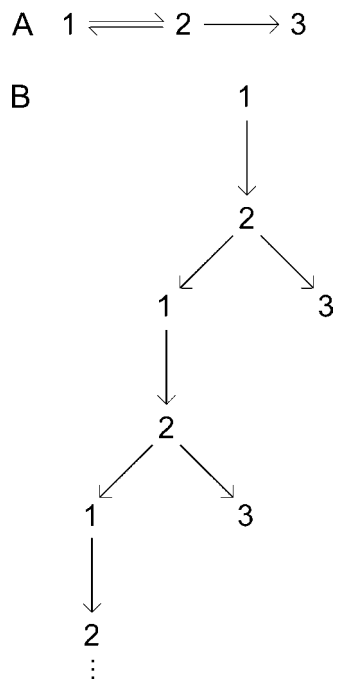

FIGURE 1 Unfolding a biochemical pathway. (*A*) A biochemical pathway with a single reversible transition between State 1 to State 2. (*B*) The unfolded representation of this pathway yields an infinite diagram containing only irreversible transitions.

memory allows us to unfold the reversible transition $1 \leftrightarrow 2$, by rewriting it as an infinite, branched pathway containing only irreversible transitions (Fig. 1 *B*). Because of the repeating character of Fig. 1 *B*, we can define the total cycle-time distribution recursively, given the branching probabilities at each vertex and the passage time for each transition.

This method can be generalized for an arbitrary biochemical pathway consisting of N states using a first-step analysis (32), in which the time taken to complete a cycle is related to the amount of time taken for the first step plus the time taken to complete the remainder of all other steps. We define the finishing-time distribution, $\pi_i(t)$, as the probability density that the system goes from State i to the final State N in time t. Let k_{ij} be the transition rate constant for going from State i to State j, such that the probability that the system makes this transition in a brief time, Δt, equals $k_{ij}\Delta t$. The probability of finishing the cycle in a time $t + \Delta t$ from State i equals the probability that the system remains in State i during the interval $[0, \Delta t]$, times the probability that it takes a further time t to complete the cycle from State i, plus the probability of the system making a transition to State j (not equal to i) during the interval $[0, \Delta t]$, times the probability it takes a further time t to complete the cycle from State j, plus a residual term of order Δt^2:

$$\pi_i(t+\Delta t) = \left(1 - \sum_{j:j\neq i} k_{ij}\Delta t\right) \times \pi_i(t) + \sum_{j:j\neq i} (k_{ij}\Delta t \times \pi_j(t)) + O(\Delta t^2). \quad (1)$$

In the limit $\Delta t \to 0$, Eq. 1 reduces to

$$\frac{d\pi_i(t)}{dt} = \sum_{j:j\neq i} k_{ij} \times (\pi_j(t) - \pi_i(t)). \quad (2)$$

Taking the Laplace transform of Eq. 2 yields

$$s\tilde{\pi}_i(s) - \pi_i(0) = \sum_{j:j\neq i} k_{ij} \times (\tilde{\pi}_j(s) - \tilde{\pi}_i(s)), \quad (3)$$

where the tilde indicates the Laplace transform. Equation 3 will allow us to solve for the set of $\tilde{\pi}_i(s)$, which serve as moment-generating functions for the corresponding distributions, $\pi_i(t)$. By definition, the probability of finishing the cycle from any State $i \neq N$ in time $t = 0$ is zero, so that $\pi_i(t = 0) = 0$. Solving for the first $N - 1$ generating functions, $\tilde{\pi}_i(s)$, we find

$$\tilde{\pi}_i(s) = \sum_{j:j\neq i} \frac{k_{ij}\tilde{\pi}_j(s)}{\sum_{m:m\neq i} k_{im} + s}, \text{ where } i = 1, 2, \ldots, N-1. \quad (4)$$

State N is defined to be the end of the cycle, hence $\pi_N(t) = \delta(t)$, which has Laplace transform $\tilde{\pi}_N(s) = 1$.

Equation 4 describes a system of of $N - 1$ coupled, linear algebraic equations for the moment-generating functions of the completion-time distributions. Such a system of equations can be solved algebraically by direct elimination or iterative methods. An expression similar to Eq. 4 was derived by Harrison and Knottenbelt (32) for modeling response times in computer communication systems, using an integral representation. In general, the total step-time distribution for a motor will be given by $\pi_{total}(t) = \sum_{i=1}^{N} p_i^0 \pi_i(t)$, where p_i^0 is the probability that the motor starts in State i at the beginning of each step. In this treatment, we consider the simplest case where the motor always starts the pathway in State 1, and therefore the step-time distribution is given by $\pi_{total}(t) = \pi_1(t)$.

Because Eq. 4 gives the solution to the generating functions, $\tilde{\pi}_i(s)$, all of the moments of $\pi_i(t)$ can easily be found through differentiation

$$\langle t^n \rangle = \left. \frac{(-1)^n d^n \tilde{\pi}_i(s)}{ds^n} \right|_{s=0}, \quad (5)$$

where the angle brackets, $\langle\ \rangle$, indicate the average. For many applications, having the moments of $\pi_i(t)$suffices. In some cases, it may be preferable to work with $\pi_i(t)$ directly, and the inverse Laplace transform of Eq. 4 can be found by factoring the denominator and using the Bromwich integral $f(t) = (1)/2\pi i \int_{\gamma-i\infty}^{\gamma+i\infty} e^{st}\tilde{f}(s)ds$.

The randomness parameter

The mechanical output of a molecular motor is governed by both the time taken to complete a full enzymatic cycle, referred to here as the cycle time, τ, and the distance moved during each cycle, the step size, d. The randomness parameter is derived from records of overall motor displacement, $x(t)$, and defined by the following limit (33):

$$r = \lim_{t\to\infty} \frac{\langle x^2(t)\rangle - \langle x(t)\rangle^2}{d\langle x(t)\rangle}. \quad (6)$$

Below, we derive general expressions for the first and second moments of the displacement $x(t)$, and thereby for the randomness, for motor mechanisms where the cycle times or step sizes are not constant.

Randomness for the case of nonuniform step sizes

To evaluate Eq. 6, we need to consider the result of a sequence of stochastic events where the step size and cycle time for each successive event are independent of each other, variable and uncorrelated. We define two probability distributions: $\rho(x)$ for the step sizes and $\pi(t)$ for the cycle times. The mean and variance in displacement from Eq. 6 can be found from the joint probability distribution, $P(x,t)$, that the system will be found at position x after time t:

$$P(x,t) = \sum_{N=0}^{\infty} P_x(N,x) \times P_t(N,t). \quad (7)$$

$P_x(N,x)$ describes the probability that N consecutive steps sum to a distance, x, and is therefore given by the N-fold convolution of $\rho(x)$ with itself. The probability of completing exactly N steps in time t, $P_t(N,t)$, comprises $N+1$ events: N steps are taken before t, and no more steps occur during the remaining time (27). When dealing with convolutions of probabilities, it is again convenient to work with the moment-generating function obtained using the Laplace transform. We can then write $\tilde{P}_x(N,q)$ and $\tilde{P}_t(N,s)$ in terms of $\tilde{\rho}(q)$and $\tilde{\pi}(s)$, where (t, s) and (x, q) are Laplace transform pairs. By the convolution theorem for Laplace transforms, we have

$$\tilde{P}_x(N,q) = \tilde{\rho}^N(q)$$
$$\tilde{P}_t(N,s) = \tilde{\pi}^N(s)\left(\frac{1-\tilde{\pi}(s)}{s}\right). \quad (8a, b)$$

From the Laplace transform of Eq. 7, it follows that

$$\tilde{P}(q,s) = \left(\frac{1-\tilde{\pi}(s)}{s}\right)\times\sum_{N=0}^{\infty}[\tilde{\rho}(q)\tilde{\pi}(s)]^N. \quad (9)$$

Equation 9 follows from the composition theorem of generating functions (34), which governs the probability distribution for the sum of a stochastic number of variables that are each distributed according to a random distribution. Constructing this probability distribution using the composition theorem is formally analogous to the construction of the grand partition function in statistical physics using the chemical potential and the partition function. Here, the total distance, x, traveled in time, t, results from the sum of a variable number of steps of variable size. To find the moment-generating function for x, one composes the generating function for the number of steps, N, taken in time t (constructed using Eq. 8b) with the moment-generating function for the distribution of step sizes, $\tilde{\rho}(q)$, yielding Eq. 9. Completing the geometric sum in Eq. 9 yields

$$\tilde{P}(q,s) = \frac{1}{s}\times\frac{1-\tilde{\pi}(s)}{[1-\tilde{\rho}(q)\tilde{\pi}(s)]}. \quad (10)$$

The first and second derivatives of $\tilde{P}(q,s)$ with respect to q are given by

$$\frac{\partial\tilde{P}(q,s)}{\partial q} = \frac{\tilde{\pi}(s)[1-\tilde{\pi}(s)]}{s[1-\tilde{\rho}(q)\tilde{\pi}(s)]^2}\times\frac{d\tilde{\rho}(q)}{dq} \quad (12)$$

and

$$\frac{\partial^2\tilde{P}(q,s)}{\partial q^2} = \frac{\tilde{\pi}(s)[1-\tilde{\pi}(s)]}{s[1-\tilde{\rho}(q)\tilde{\pi}(s)]^2}\times\left(\frac{d^2\tilde{\rho}(q)}{dq^2}\right) + \frac{2\tilde{\pi}^2(s)[1-\tilde{\pi}(s)]}{s[1-\tilde{\rho}(q)\tilde{\pi}(s)]^3}\times\left(\frac{d\tilde{\rho}(q)}{dq}\right)^2. \quad (13)$$

$\tilde{\rho}(0) = \int_0^\infty \rho(t)dt = 1$, because $\rho(x)$ is a normalized probability distribution. Solving for $\langle x(s)\rangle$ in Eq. 11 yields

$$\langle\tilde{x}(s)\rangle = \frac{\tilde{\pi}(s)}{s[1-\tilde{\pi}(s)]}\times\left(-1\times\left.\frac{d\tilde{\rho}(q)}{dq}\right|_{q=0}\right) = \langle\tilde{N}(s)\rangle\times\langle d\rangle, \quad (14)$$

where $\langle\tilde{N}(s)\rangle$ is the Laplace transform of the average number of steps taken after time t (33) and $\langle d\rangle$ is the average step size derived from the distribution $\rho(x)$. Similarly,

$$\langle\tilde{x}^2(s)\rangle = \frac{\tilde{\pi}(s)}{s[1-\tilde{\pi}(s)]}\times\left(\left.\frac{d^2\tilde{\rho}(q)}{dq^2}\right|_{q=0}\right) + \frac{2\tilde{\pi}^2(s)}{s[1-\tilde{\pi}(s)]^2}\times\left(\left.\frac{d\tilde{\rho}(q)}{dq}\right|_{q=0}\right)^2 = \langle\tilde{N}(s)\rangle\times\langle d^2\rangle + [\langle\tilde{N}^2(s)\rangle - \langle\tilde{N}(s)\rangle]\times\langle d\rangle^2. \quad (15)$$

Taking the inverse Laplace transform of Eqs. 14 and 15 yields

$$\langle x(t)\rangle = \langle N(t)\rangle\times\langle d\rangle$$
$$\langle x^2(t)\rangle = \langle N(t)\rangle\times\langle d^2\rangle + [\langle N^2(t)\rangle - \langle N(t)\rangle]\times\langle d\rangle^2. \quad (16)$$

Taking d to be the average step size, the randomness, as defined in Eq. 6, is

$$r = \underset{t\to\infty}{Lim}\frac{\langle N(t)\rangle\times\langle d^2\rangle + [\langle N^2(t)\rangle - \langle N(t)\rangle]\times\langle d\rangle^2 - \langle N(t)\rangle^2\times\langle d\rangle^2}{\langle N(t)\rangle\times\langle d\rangle^2}, \quad (17)$$

Computing the randomness according to Eq. 6 requires the calculation of $\langle x(s)\rangle$ and $\langle x(s)^2\rangle$. These can be found from Eq. 10 using the following relation:

$$\langle\tilde{x}^n(s)\rangle = (-1)^n\left.\frac{\partial^n\tilde{P}(q,s)}{\partial q^n}\right|_{q=0}. \quad (11)$$

which reduces to

$$r = \frac{\langle d^2\rangle - \langle d\rangle^2}{\langle d\rangle^2} + \underset{t\to\infty}{Lim}\frac{\langle N^2(t)\rangle - \langle N(t)\rangle^2}{\langle N(t)\rangle}. \quad (18)$$

Schnitzer and Block (27) showed that $Lim_{t\to\infty}(\langle N^2(t)\rangle - \langle N(t)\rangle^2)/\langle N(t)\rangle = (\langle\tau^2\rangle - \langle\tau\rangle^2)/\langle\tau\rangle^2$, where $\langle\tau\rangle$ and $\langle\tau^2\rangle$ are

the first and second moments of the step-time distribution $\pi(t)$. Therefore, Eq. 18 becomes

$$r = \frac{\langle d^2\rangle - \langle d\rangle^2}{\langle d\rangle^2} + \frac{\langle \tau^2\rangle - \langle \tau\rangle^2}{\langle \tau\rangle^2}. \quad (19)$$

We can now define two contributions to the overall randomness: one derived from the variability in the step size and one from the variability in step time,

$$r_{\text{step sizes}} = \frac{\langle d^2\rangle - \langle d\rangle^2}{\langle d\rangle^2} \quad (20)$$

and

$$r_{\text{step times}} = \frac{\langle \tau^2\rangle - \langle \tau\rangle^2}{\langle \tau\rangle^2}. \quad (21)$$

Equation 19 can then be rewritten as $r = r_{\text{step sizes}} + r_{\text{step times}}$.

In an experiment, the measured step size may not equal the average step size, $\langle d\rangle$, as calculated from $\rho(x)$. This can be seen by considering a distribution that includes, with finite probability, steps with sizes below the resolution limit of the apparatus. Any experimental measurements of motor steps would, by definition, miss these steps. Hence, the measured step size would be higher than $\langle d\rangle$. This experimental bias can be corrected by scaling the calculated randomness by the ratio of the predicted to the measured step size,

$$r_{\text{measured}} = \frac{\langle d\rangle}{d_{\text{measured}}}(r_{\text{step sizes}} + r_{\text{step times}}). \quad (22)$$

Randomness for the case of a heterogeneous substrate

When a motor moves along a heterogeneous substrate, such as DNA, the total step-time distribution is governed by a set of biochemical pathways related to details of the local sequence. We define M different step-time distributions, $\pi^{\text{i}}(t)$, corresponding to each of the M particular sequence motifs in the substrate, and probabilities, ϕ_{i}, of encountering the ith motif. To model the sequence-specific behavior of RNA (or DNA) polymerase, for example, we might begin by supposing that there are four such distributions, one for each possible base: $\pi^{\text{A}}(t)$, $\pi^{\text{T}}(t)$, $\pi^{\text{C}}(t)$, and $\pi^{\text{G}}(t)$, where ϕ_{A} represents the local fraction of A values in the DNA template (this simplest case ignores any longer-range sequence effects due to adjacent bases). The total step-time distribution is then

$$\pi_{\text{total}}(t) = \sum_{\text{i}=1}^{\text{M}} \phi_{\text{i}} \cdot \pi^{\text{i}}(t). \quad (23)$$

To calculate the step-time randomness as defined by Eq. 21, we solve for the moments of the ith step-time distribution. Averaging over all paths, Eq. 23 leads to

$$\langle \tau\rangle = \sum_{\text{i}=1}^{\text{M}} \phi_{\text{i}}\langle \tau_{\text{i}}\rangle, \quad \langle \tau^2\rangle = \sum_{\text{i}=1}^{\text{M}} \phi_{\text{i}}\langle \tau_{\text{i}}^2\rangle. \quad (24)$$

Inserting Eq. 24 into Eq. 21 yields the total randomness

$$r = \sum_{\text{i}=1}^{\text{M}} \left[\phi_{\text{i}} \frac{\langle \tau_{\text{i}}^2\rangle}{\langle \tau\rangle^2}\right] - 1, \quad (25)$$

assuming the step size is the same for each pathway. Rearrangement of Eq. 21 reveals that

$$\langle \tau_{\text{i}}^2\rangle = \langle \tau_{\text{i}}\rangle^2 + r_{\text{i}}\langle \tau_{\text{i}}\rangle^2, \quad (26)$$

which can be used to rewrite the total randomness in terms of the randomness from each pathway alone, r_{i}:

$$r = \left[\sum_{\text{i}=1}^{\text{M}} \phi_{\text{i}} \left(\frac{\langle \tau_{\text{i}}\rangle}{\langle \tau\rangle}\right)^2 r_{\text{i}}\right] + \left[\sum_{\text{i}=1}^{\text{M}} \phi_{\text{i}} \left(\frac{\langle \tau_{i}\rangle}{\langle \tau\rangle}\right)^2 - 1\right]. \quad (27)$$

The first term on the right is a weighted sum of the randomness values from each distribution. The weighting function includes terms for the probability of following a particular pathway as well as for the average step time from that path relative to the total average step time. The second term represents the randomness in step times generated by stochastically choosing pathways with different mean step times. If the mean step times are all equal, the total randomness, r, becomes simply the sum of the individual randomness values, r_{i}, weighted by the probability of choosing each path, ϕ_{i}. Conversely, even if each path is perfectly clocklike, i.e., $r_{\text{i}} = 0$ for all i, the total randomness will be >0 due to the variation in step times from each path.

WORKED EXAMPLES

Step-time calculations

Consider the enzymatic pathway described by Fig. 1 *A*, where the first of two transitions is reversible. The finishing-time distributions can be found using Eq. 4:

$$\begin{aligned}\tilde{\boldsymbol{\pi}}_1(s) &= \frac{k_{12}\tilde{\boldsymbol{\pi}}_2(s)}{k_{12}+s}\\ \tilde{\boldsymbol{\pi}}_2(s) &= \frac{k_{21}\tilde{\boldsymbol{\pi}}_1(s) + k_{23}\tilde{\boldsymbol{\pi}}_3(s)}{k_{21}+k_{23}+s}\\ \tilde{\boldsymbol{\pi}}_3(s) &= 1.\end{aligned} \quad (28)$$

Solving for the step-time distribution $\tilde{\boldsymbol{\pi}}_{\text{total}}(s) = \boldsymbol{\pi}_1(s)$ yields

$$\tilde{\boldsymbol{\pi}}_{\text{total}}(s) = \frac{k_{12}k_{23}}{s^2 + s(k_{12}+k_{21}+k_{23}) + k_{12}k_{23}}, \quad (29)$$

which is identical to the result derived by Schnitzer and Block (27) using the previous sum-of-all-paths method. Note that the recursive derivation introduced here does not require the evaluation of any infinite sums; the infinite number of paths that the system can take is embodied instead in the recursive nature of Eq. 28.

The example shown in Fig. 2 *A* describes a more complex pathway that consists of one reversible transition, between States 2 and 3, plus a loop, cycling between States 1 through 3. The sum-of-all-paths method would involve two interrelated infinite sums counting the number of times the

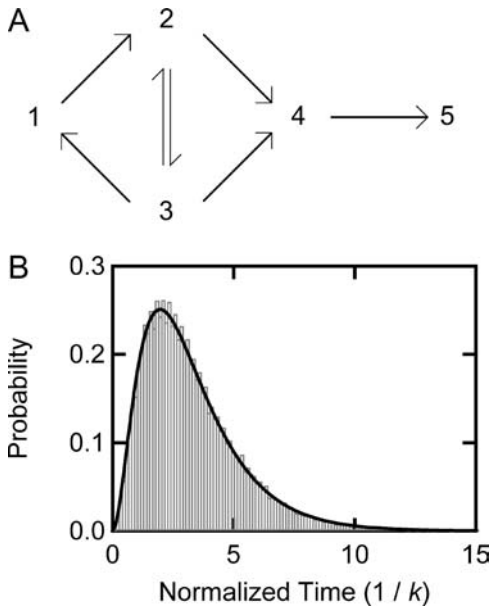

FIGURE 2 Analysis of a figure-eight-shaped biochemical pathway with five states, one loop, and both reversible and irreversible transitions. (*A*) A biochemical pathway with a single reversible transition between State 2 and State 3 and a loop from States 1–3. In this simplified example, each of the transitions is governed by the identical rate constant, k. (*B*) Probability density for the completion of the pathway described in *A*. The calculated probability (*solid curve*, from Eq. 33) is in excellent agreement (histogram, *shaded bars*) with the results of a Monte Carlo simulation ($N = 100{,}000$ trials).

system went back and forth from State 3 to 2, and the number of times the system went around the loop, as well as a term relating to whether the system arrived at State 4 from State 2 or from State 3. The recursive method is much simpler, and the set of equations defined by Eq. 4 reduces to

$$\tilde{\pi}_1(s) = \frac{k_{12}\tilde{\pi}_2(s)}{k_{12}+s} \quad \tilde{\pi}_2(s) = \frac{k_{23}\tilde{\pi}_3(s) + k_{24}\tilde{\pi}_4(s)}{k_{23}+k_{24}+s}$$
$$\tilde{\pi}_3(s) = \frac{k_{31}\tilde{\pi}_1(s) + k_{32}\tilde{\pi}_2(s) + k_{34}\tilde{\pi}_4(s)}{k_{31}+k_{32}+k_{34}+s}$$
$$\tilde{\pi}_4(s) = \frac{k_{45}\tilde{\pi}_5(s)}{k_{45}+s} \quad \tilde{\pi}_5(s) = 1. \tag{30}$$

Solving for the total step-time distribution yields

$$\tilde{\pi}_{\text{total}}(s) = \frac{k_{12}k_{45}(k_{23}k_{34} + k_{24}(k_{31}+k_{32}+k_{34}+s))}{(k_{45}+s)(k_{12}(k_{23}(k_{34}+s) + (k_{24}+s)(k_{31}+k_{32}+k_{34}+s)) + s(k_{23}(k_{31}+k_{34}+s) + (k_{24}+s)(k_{31}+k_{32}+k_{34}+s)))}. \tag{31}$$

For illustrative purposes, we consider the case when all transition rates happen to equal k. Then, $\tilde{\pi}_{\text{total}}(s)$ reduces to

$$\tilde{\pi}_{\text{total}}(s) = \frac{k^3(4k+s)}{(k+s)(2k+s)(2k^2+4ks+s^2)}. \tag{32}$$

The inverse Laplace transform reduces to

$$\pi_1(t) = \frac{k(4\cosh(\sqrt{2}kt) - 6e^{kt} + 3\sqrt{2}\sinh(\sqrt{2}kt) + 2)}{2e^{2kt}}. \tag{33}$$

Fig. 2 *B* shows the histogram resulting from of a Monte Carlo simulation of the figure-eight pathway of Fig. 2 *A*. The solid curve matching this histogram is supplied by Eq. 33.

Futile hydrolysis I: the step-time method

The recursive dwell-time method can be used to predict the effect of futile hydrolysis on randomness. If we define ε as the probability that the system does not finish the cycle when State N is reached, but instead must repeat the cycle again, beginning at State 1, then the completion-time distribution from State N becomes

$$\pi_N(t) = (1-\varepsilon)\delta(t) + \varepsilon \times \pi_1(t), \tag{34}$$

or equivalently

$$\tilde{\pi}_N(s) = (1-\varepsilon) + \varepsilon \times \tilde{\pi}_1(s). \tag{35}$$

Solving for the total cycle-time distribution using Eqs. 4 and 35 yields

$$\tilde{\pi}_{\text{total}}(s) = \frac{(1-\varepsilon)\tilde{\pi}^0(s)}{1-\varepsilon\tilde{\pi}^0(s)}, \tag{36}$$

where $\tilde{\pi}^0(s) \equiv \tilde{\pi}_{\text{total}}(s)$ for the case $\varepsilon = 0$. The moments of the cycle-time distribution can be found by differentiating $\tilde{\pi}_{\text{total}}(s)$:

$$\langle\tau\rangle = (-1)\frac{d\tilde{\pi}_{\text{total}}(s)}{ds}\bigg|_{s=0} = \frac{(1-\varepsilon)}{[1-\varepsilon\tilde{\pi}^0(s)]^2}\frac{d\tilde{\pi}^0(s)}{ds}\bigg|_{s=0} = \frac{\langle\tau^0\rangle}{1-\varepsilon} \tag{37}$$

and

$$\langle\tau^2\rangle = \frac{d^2\tilde{\pi}_{\text{total}}(s)}{ds^2}\bigg|_{s=0} = \frac{2\varepsilon(1-\varepsilon)}{[1-\varepsilon\tilde{\pi}^0(s)]^3}\left(\frac{d\tilde{\pi}^0(s)}{ds}\right)^2\bigg|_{s=0} + \frac{(1-\varepsilon)}{[1-\varepsilon\tilde{\pi}^0(s)]^2}\frac{d^2\tilde{\pi}^0(s)}{ds^2}\bigg|_{s=0} = \frac{2\varepsilon}{(1-\varepsilon)^2}\langle\tau^0\rangle^2 + \frac{\langle\tau^{02}\rangle}{1-\varepsilon}, \tag{38}$$

where $\langle\tau^0\rangle$ and $\langle\tau^{02}\rangle$ are the first and second moments of the cycle-time distribution for $\varepsilon = 0$. Inserting Eqs. 37 and 38 into Eq. 21 yields a randomness of $r = \varepsilon + (1-\varepsilon)r^0$, where

$$r^0 \equiv \frac{\langle\tau^{02}\rangle - \langle\tau^0\rangle^2}{\langle\tau^0\rangle^2} \tag{39}$$

is the randomness of the pathway for $\varepsilon = 0$, in agreement with the findings of Svoboda et al. (33).

Futile hydrolysis II: the step-size method

The identical result can be obtained using Eq. 19 and a judicious choice of step-size distribution. We define a step-size distribution such that with probability ε, the motor takes a step of size zero, and with probability $(1 - \varepsilon)$, the motor takes a step of size d_0:

$$\rho(x) = (1 - \varepsilon) \times \delta(x - d_0) + \varepsilon \times \delta(x). \tag{40}$$

The moments of this distribution are given by

$$\langle d \rangle = (1 - \varepsilon) d_0 \ \langle d^2 \rangle = (1 - \varepsilon) d_0^2. \tag{41}$$

We then calculate the randomness due to the variation in step sizes, defined in Eq. 20,

$$r_{\text{step sizes}} = \frac{\varepsilon}{1 - \varepsilon}. \tag{42}$$

Taking into account that the measured step size, d_0, differs from the average step size $\langle d \rangle = (1 - \varepsilon) d_0$, we find, using Eq. 22, that

$$\begin{aligned} r_{\text{measured}} &= \frac{(1 - \varepsilon) d_0}{d_0} \left(\frac{\varepsilon}{1 - \varepsilon} + r^0 \right) \\ &= \varepsilon + (1 - \varepsilon) r^0, \end{aligned} \tag{43}$$

where (again) r^0 represents the randomness of the pathway for $\varepsilon = 0$, in agreement with the recursive derivation above.

A similar analysis leads to the value for randomness in the presence of backward steps. If the motor takes a step backward with probability, P_-, the step-size distribution takes the form

$$\rho(x) = (1 - P_-) \times \delta(x - d_0) + P_- \times \delta(x + d_0). \tag{44}$$

Equation 22 then yields a measured randomness of

$$r_{\text{measured}} = (r^0 - 1)(1 - 2P_-) + \frac{1}{1 - 2P_-}, \tag{45}$$

in agreement with Schnitzer and Block (19).

A heterogeneous substrate

Now consider the progress of a motor that, at each step, chooses between one of two distinct pathways. This might correspond, in a highly simplified model, to a polymerase enzyme moving along a DNA template consisting of a random distribution of just two of the four bases. For illustrative purposes, we further suppose that each pathway has a biochemical randomness of zero, i.e., that the step times are fixed (nonstochastic) but may be different from one another, and that the probability of taking either pathway is identical, i.e., $\phi_1 = \phi_2 = 1/2$. For such a case, Eq. 27 reduces to

$$r = \frac{1}{2} \left(\frac{\langle \tau_1 \rangle}{\langle \tau \rangle} \right)^2 + \frac{1}{2} \left(\frac{\langle \tau_2 \rangle}{\langle \tau \rangle} \right)^2 - 1. \tag{46}$$

If we define N as the ratio between the two step times, then $\langle \tau_2 \rangle = N \langle \tau_1 \rangle$ and $\langle \tau \rangle = \langle \tau_1 \rangle (N+1)/2$. Solving for the randomness, we find

$$r = \frac{1}{2} \left(\frac{2}{N + 1} \right)^2 + \frac{1}{2} \left(\frac{2N}{N + 1} \right)^2 - 1, \tag{47}$$

which simplifies to

$$r = \left(\frac{N - 1}{N + 1} \right)^2. \tag{48}$$

If one type of step takes twice the time as the other, then the measured randomness will have a value of only 1/9, due solely to the variability in choosing between the different pathways.

In the limit that one step is much faster than the other, i.e., $N \gg 1$, then the randomness reduces to a value of 1. This result is readily understood, since this scenario is analogous to taking a composite step comprised of a variable number of short fast steps taken before one long slow step. The probability of taking a composite step of size $d = n \times d_0$, where d_0 is the short step size, is given by

$$\begin{aligned} P(\text{step size of } d) &= P(n - 1 \text{ ``fast'' steps}) \\ &\quad \times P(\text{one ``slow'' step}) \\ &= \left(\frac{1}{2} \right)^{d/d_0}. \end{aligned} \tag{49}$$

Because the biochemical randomness for the slow path is equal to zero, we can solve for the total randomness by solving for the first and second moments of Eq. 49

$$\langle d \rangle = 2 d_0 \tag{50}$$

$$\langle d^2 \rangle = 6 d_0^2, \tag{51}$$

which yields a randomness of 1 using Eq. 22.

DISCUSSION

Recent advances in biophysics have led to the ability to measure completion time distributions for enzymatic reactions and folding pathways. Comparisons of such distributions with theory can reveal intermediate biochemical states that might otherwise remain hidden in an analysis based on traditional, bulk solution studies. However, analytical calculation of completion time distributions for all but the simplest pathways remains a difficult proposition. The recursive method derived here extends kinetic analysis to include reversible, branched, and other, more complicated kinetic schemes. This formalism has potential applications in studies of molecular motor mechanochemistry, enzyme dynamics, protein and nucleic-acid folding, and ion channel conductance.

The ability to measure every step that an individual protein takes has led to a number of discoveries in the field of molecular motors. For example, Rief et al. (18) found that the step-time distribution for myosin V could be described by a convolution of two exponentials with different time constants, where just one time constant varied with the ATP concentration. This finding allowed them to distinguish both ATP-dependent and ATP-independent rates in the mechanochemical cycle. A comparable analysis was performed for rotational stepping data on the F_1-ATPase motor (35). In general, to learn about enzymatic behavior, one studies the influence of control parameters thought to affect specific rates in the biochemical pathway. Traditionally, this might involve changing the concentrations of substrates, inhibitors, or end products and measuring their influence in a biochemical assay. More recently, with the development of instruments that can manipulate single molecules, such as optical and magnetic traps, force and torque have also been established as useful control parameters that can selectively modify rates of biochemical transitions, particularly those involving physical motions (22). The ability to predict the velocity and randomness for an arbitrary model pathway allows one to perform global fits to an entire series of experimental measurements to judge the goodness-of-fit along with the most likely rate parameters (22).

Optical trapping studies of processive nucleic acid-based motors are now approaching single basepair resolution (10), and enzymes such as RNA polymerase and phage λ exonuclease have been shown to enter into long-lived, sequence-dependent pause states as they travel along DNA (8,36–38). As the experimental measurements become more precise, kinetic modeling will need to take into account sequence-dependent variations in enzyme behavior. Further information about the elongation cycle of RNA polymerase can be gained by measuring its randomness in the presence of limiting amounts of free nucleotide, or by using a template with a reduced representation of the four bases. The expression derived here for the effect of sequence on the randomness may have applicability to such studies.

Recent experiments on myosin V and cytoplasmic dynein suggest that these mechanoenzymes may take steps of a variable size. During stepping, the heads of myosin V are thought to be able to bind to a variety of alternative, nearby sites along an actin filament, yielding a broad step-size distribution with an average displacement of 36 nm ((39) and references therein). The intrinsic width of the step size distribution is currently uncertain, however, because experimental measurements thus far have been dominated by various types of noise. This width could be determined, in principle, by measuring the overall randomness for myosin V motors under limiting ATP conditions (40), where the biochemical contribution to the randomness takes on a value of 1. Under such conditions, any deviations from unity would be correlated to the width of the step-size distribution using Eq. 22. Dynein, on the other hand, has recently been reported to take steps (up to 32 nm) that are always integral multiples of 8 nm, in a load-dependent manner (14). The effect of this step-size distribution on the randomness can be calculated, and a statistical analysis can be used, in principle, to characterize the step size distribution without the need to identify every individual step.

As experimental methods advance, the questions one can usefully ask about the events driving enzyme function, or about the processes involved in macromolecular folding and unfolding, become increasingly sophisticated. The formalism developed here may prove useful in untangling questions about the potentially complex mechanisms underlying these systems, particularly those described by nontrivial kinetic schemes.

Note added in proof: Dr. Hongyun Wang (University of California, Santa Cruz, Dept. Applied Mathematics and Statistics) has furnished an elegant alternative derivation of our Eq. 18 using conditional probabilities (unpublished).

REFERENCES

1. Sakmann, B., and E. Neher. 1984. Patch clamp techniques for studying ionic channels in excitable membranes. *Annu. Rev. Physiol.* 46: 455–472.
2. Rueda, D., G. Bokinsky, M. M. Rhodes, M. J. Rust, X. W. Zhuang, and N. G. Walter. 2004. Single-molecule enzymology of RNA: Essential functional groups impact catalysis from a distance. *Proc. Natl. Acad. Sci. USA.* 101:10066–10071.
3. Lipman, E. A., B. Schuler, O. Bakajin, and W. A. Eaton. 2003. Single-molecule measurement of protein folding kinetics. *Science.* 301:1233–1235.
4. Bokinsky, G., D. Rueda, V. K. Misra, M. M. Rhodes, A. Gordus, H. P. Babcock, N. G. Walter, and X. Zhuang. 2003. Single-molecule transition-state analysis of RNA folding. *Proc. Natl. Acad. Sci. USA.* 100: 9302–9307.
5. Schuler, B., E. A. Lipman, and W. A. Eaton. 2002. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature.* 419:743–747.
6. Zhuang, X., L. E. Bartley, H. P. Babcock, R. Russell, T. Ha, D. Herschlag, and S. Chu. 2000. A single-molecule study of RNA catalysis and folding. *Science.* 288:2048–2051.
7. Xie, X. S. 2002. Single-molecule approach to dispersed kinetics and dynamic disorder: Probing conformational fluctuation and enzymatic dynamics. *J. Chem. Phys.* 117:11024–11032.
8. van Oijen, A. M., P. C. Blainey, D. J. Crampton, C. C. Richardson, T. Ellenberger, and X. S. Xie. 2003. Single-molecule kinetics of lambda exonuclease reveal base dependence and dynamic disorder. *Science.* 301:1235–1238.
9. Ueno, T., H. Taguchi, H. Tadakuma, M. Yoshida, and T. Funatsu. 2004. GroEL mediates protein folding with a two successive timer mechanism. *Mol. Cell.* 14:423–434.
10. Shaevitz, J. W., E. A. Abbondanzieri, R. Landick, and S. M. Block. 2003. Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature.* 426:684–687.
11. Yasuda, R., H. Noji, M. Yoshida, K. Kinosita Jr., and H. Itoh. 2001. Resolution of distinct rotational substeps by submillisecond kinetic analysis of F1-ATPase. *Nature.* 410:898–904.
12. Nishiyama, M., E. Muto, Y. Inoue, T. Yanagida, and H. Higuchi. 2001. Substeps within the 8-nm step of the ATPase cycle of single kinesin molecules. *Nat. Cell Biol.* 3:425–428.

13. Svoboda, K., C. F. Schmidt, B. J. Schnapp, and S. M. Block. 1993. Direct observation of kinesin stepping by optical trapping interferometry. *Nature*. 365:721–727.

14. Mallik, R., B. C. Carter, S. A. Lex, S. J. King, and S. P. Gross. 2004. Cytoplasmic dynein functions as a gear in response to load. *Nature*. 427:649–652.

15. Mehta, A. D., R. S. Rock, M. Rief, J. A. Spudich, M. S. Mooseker, and R. E. Cheney. 1999. Myosin-V is a processive actin-based motor. *Nature*. 400:590–593.

16. Rock, R. S., S. E. Rice, A. L. Wells, T. J. Purcell, J. A. Spudich, and H. L. Sweeney. 2001. Myosin VI is a processive motor with a large step size. *Proc. Natl. Acad. Sci. USA*. 98:13655–13659.

17. Yasuda, R., H. Noji, K. Kinosita Jr., and M. Yoshida. 1998. F1-ATPase is a highly efficient molecular motor that rotates with discrete 120 degree steps. *Cell*. 93:1117–1124.

18. Rief, M., R. S. Rock, A. D. Mehta, M. S. Mooseker, R. E. Cheney, and J. A. Spudich. 2000. Myosin-V stepping kinetics: a molecular model for processivity. *Proc. Natl. Acad. Sci. USA*. 97:9482–9486.

19. Schnitzer, M. J., and S. M. Block. 1997. Kinesin hydrolyses one ATP per 8-nm step. *Nature*. 388:386–390.

20. Neher, E., and C. F. Stevens. 1977. Conductance fluctuations and ionic pores in membranes. *Annu. Rev. Biophys. Bioeng*. 6:345–381.

21. Visscher, K., M. J. Schnitzer, and S. M. Block. 1999. Single kinesin molecules studied with a molecular force clamp. *Nature*. 400:184–189.

22. Block, S. M., C. L. Asbury, J. W. Shaevitz, and M. J. Lang. 2003. Probing the kinesin reaction cycle with a 2D optical force clamp. *Proc. Natl. Acad. Sci. USA*. 100:2351–2356.

23. Samuel, A. D., and H. C. Berg. 1995. Fluctuation analysis of rotational speeds of the bacterial flagellar motor. *Proc. Natl. Acad. Sci. USA*. 92: 3502–3506.

24. Colquhoun, D., and F. J. Sigworth. 1983. Fitting and statistical analysis of single-channel records. *In* Single-Channel Recordings. B. Sakmann and E. Neher, editors. Plenum, New York. 191–263.

25. Hill, T. 1989. Free Energy Transduction and Biochemical Cycle Kinetics. Springer-Verlag, New York.

26. Milescu, L. S., G. Akk, and F. Sachs. 2005. Maximum likelihood estimation of ion channel kinetics from macroscopic currents. *Biophys. J*. 88: 2494–2515.

27. Schnitzer, M. J., and S. M. Block. 1995. Statistical kinetics of processive enzymes. *Cold Spring Harb. Symp. Quant. Biol.* 60:793–802.

28. Derrida, B. 1983. Velocity and diffusion of a periodic one-dimensional hopping model. *J. Stat. Phys*. 31:433–450.

29. Wang, H., C. S. Peskin, and T. C. Elston. 2003. A robust numerical algorithm for studying biomolecular transport processes. *J. Theor. Biol*. 221:491–511.

30. Elston, T. C. 2000. A macroscopic description of biomolecular transport. *J. Math. Biol*. 41:189–206.

31. Fisher, M. E., and A. B. Kolomeisky. 2001. Simple mechanochemistry describes the dynamics of kinesin molecules. *Proc. Natl. Acad. Sci. USA*. 98:7748–7753.

32. Harrison, P. G., and W. J. Knottenbelt. 2002. Passage time distributions in large Markov models. *ACM SIGMETRICS Performance Evaluation Review*. 30:77–85.

33. Svoboda, K., P. P. Mitra, and S. M. Block. 1994. Fluctuation analysis of motor protein movement and single enzyme kinetics. *Proc. Natl. Acad. Sci. USA*. 91:11782–11786.

34. Feller, W. 1971. An Introduction to Probability Theory and Its Applications, 2nd ed. J. Wiley and Sons, New York.

35. Shimabukuro, K., R. Yasuda, E. Muneyuki, K. Y. Hara, K. Kinosita Jr., and M. Yoshida. 2003. Catalysis and rotation of F1 motor: cleavage of ATP at the catalytic site occurs in 1 ms before 40 degree substep rotation. *Proc. Natl. Acad. Sci. USA*. 100:14731–14736.

36. Perkins, T. T., R. V. Dalal, P. G. Mitsis, and S. M. Block. 2003. Sequence-dependent pausing of single lambda exonuclease molecules. *Science*. 301:1914–1918.

37. Artsimovitch, I., and R. Landick. 2000. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci. USA*. 97:7090–7095.

38. Shundrovsky, A., T. J. Santangelo, J. W. Roberts, and M. D. Wang. 2004. A single-molecule technique to study sequence-dependent transcription pausing. *Biophys. J*. 87:3945–3953.

39. Mehta, A. 2001. Myosin learns to walk. *J. Cell Sci.* 114:1981–1998.

40. Kolomeisky, A. B., and M. E. Fisher. 2003. A simple kinetic model describes the processivity of myosin-V. *Biophys. J*. 84:1642–1650.