CHAPTER TEN

Methods in Statistical Kinetics

Jeffrey R. Moffitt,* Yann R. Chemla,† *and* Carlos Bustamante*,‡

Contents

Abstract

A variety of recent advances in single-molecule methods are now making possible the routine measurement of the distinct catalytic trajectories of individual enzymes. Unlike their bulk counterparts, these measurements directly reveal the fluctuations inherent to enzymatic dynamics, and statistical measures of these fluctuations promise to greatly constrain possible kinetic mechanisms. In this chapter, we discuss a variety of advances, ranging from theoretical to practical, in the new and growing field of statistical kinetics. In particular, we formalize the connection between the hidden fluctuations in the kinetic states that compose a full kinetic cycle and the measured fluctuations in

* Department of Physics and Jason L. Choy Laboratory of Single-Molecule Biophysics, University of California, Berkeley, California, USA
† Department of Physics and Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
‡ Departments of Molecular and Cell Biology, and Chemistry, Howard Hughes Medical Institute, University of California, Berkeley, California, USA

Methods in Enzymology, Volume 475
ISSN 0076-6879, DOI: 10.1016/S0076-6879(10)75010-2

the time to complete this cycle. We then discuss the characterization of fluctuations in a fashion that permits kinetic constraints to be easily extracted. When there are multiple observable enzymatic outcomes, we provide the proper way to sort events so as not to bias the final statistics, and we show that these classifications provide a first level of constraint on possible kinetic mechanisms. Finally, we discuss the basic substrate dependence of an important function of the statistical moments. The new kinetic parameters of this expression, akin to the Michaelis–Menten parameters, provide model-independent constraints on the kinetic mechanism.

1. Introduction

Enzyme dynamics are naturally stochastic. While the directionality of catalyzed reactions is driven by the energy stored in chemical or electrochemical potentials, it is not this energy that drives the internal conformational changes and chemical transformations that compose the kinetic cycle of the enzyme. Rather these transitions are driven by the energy of the surrounding, fluctuating thermal bath. The electrochemical driving potential simply biases these conformational fluctuations along the reaction pathway. As a result, kinetic transitions are stochastic, and the time to complete one full enzymatic cycle is a random quantity. Thus, measures of enzyme dynamics must naturally be statistical.

For much of the twentieth century such fluctuations were ignored due simply to the difficulty in detecting them in large ensembles of unsynchronized copies of enzyme. However, with the recent advances in single-molecule techniques and synchronized ensemble methods, it is now possible to observe these fluctuations directly (Cornish and Ha, 2007; Greenleaf *et al.*, 2007; Moffitt *et al.*, 2008; Sakmann and Neher, 1984). These powerful experimental advances necessarily raise new theoretical questions. In particular, what type of mechanistic information is contained within fluctuations, and how can this kinetic information be extracted in an accurate and unbiased fashion? Moreover, can fluctuations be classified—characterized, perhaps, by quantities in analogy to the Michaelis–Menten constants K_M and k_{cat}? And, if so, what are the implications of such classification, and what do the new kinetic parameters reveal about possible models for the enzymatic reaction?

These are the types of basic questions that face the new field of *statistical kinetics*—the extension of enzyme kinetics from the mean rate of product formation to measures of the inherent fluctuations in this rate. In this chapter, we discuss several recent advances in this field, both theoretical and practical. Our purpose is not to provide a comprehensive discussion of the theoretical foundation of this field nor of the various techniques and methods that are being developed, but rather to complement discussions of

a variety of topics, some of them treated extensively in the literature (Charvin *et al.*, 2002; Chemla *et al.*, 2008; Fisher and Kolomeisky, 1999; Kolomeisky and Fisher, 2007; Neuman *et al.*, 2005; Qian, 2008; Schnitzer and Block, 1995; Shaevitz *et al.*, 2005; Svoboda *et al.*, 1994; Xie, 2001). In the first section, we revisit the foundational ideas of statistical kinetics. We adopt a slightly different perspective from other authors, one based on lifetimes rather than kinetic rates, and derive a formal connection between the statistical properties of enzymatic reactions and the statistical properties of their composite kinetic states. In the second section, we turn to more practical matters and discuss different methods of quantifying fluctuations. We explain why, with current methods, fitting the full distribution of lifetimes likely introduces greater risk of bias than extracting kinetic information from statistical moments. We also present simple statistical tests that allow different lifetimes from different kinetic mechanisms to be properly sorted when an enzyme has multiple observable outcomes. In the final section, we discuss methods for extracting kinetic information from the measured moments. In particular, we discuss a newly developed technique for classifying enzymatic fluctuations and the mechanistic constraints provided by the new kinetic parameters introduced in the classification.

2. The Formalism of Statistical Kinetics

2.1. From steady-state kinetics to statistical kinetics

To illustrate how fluctuations arise in kinetic processes, consider the canonical kinetic model—the Michaelis–Menten mechanism (Michaelis and Menten, 1913):

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1[\mathrm{S}]}{\leftrightarrow}} \mathrm{ES} \xrightarrow{k_2} \mathrm{E} + \mathrm{P}. \tag{10.1}$$

Here an enzyme, E, binds substrate S to form the bound form ES with the pseudo-first-order rate constant $k_1[\mathrm{S}]$. This bound form can then produce product P, or it can unbind the substrate unproductively, returning to E. These processes have rate constants k_2 and k_{-1}, respectively. In general, we will treat the formation of "product" as the only detectable event, though this event could be any kinetic transition that generates an experimentally measurable signal. All other kinetic transitions are hidden from detection. Our measured quantity will be the time between subsequent product formation events or detectable events, the *cycle completion time*. In the case of molecular motors, the formation of "product" is often the generation of a physical motion, such as a discrete step along a periodic track (Fig. 10.1A).

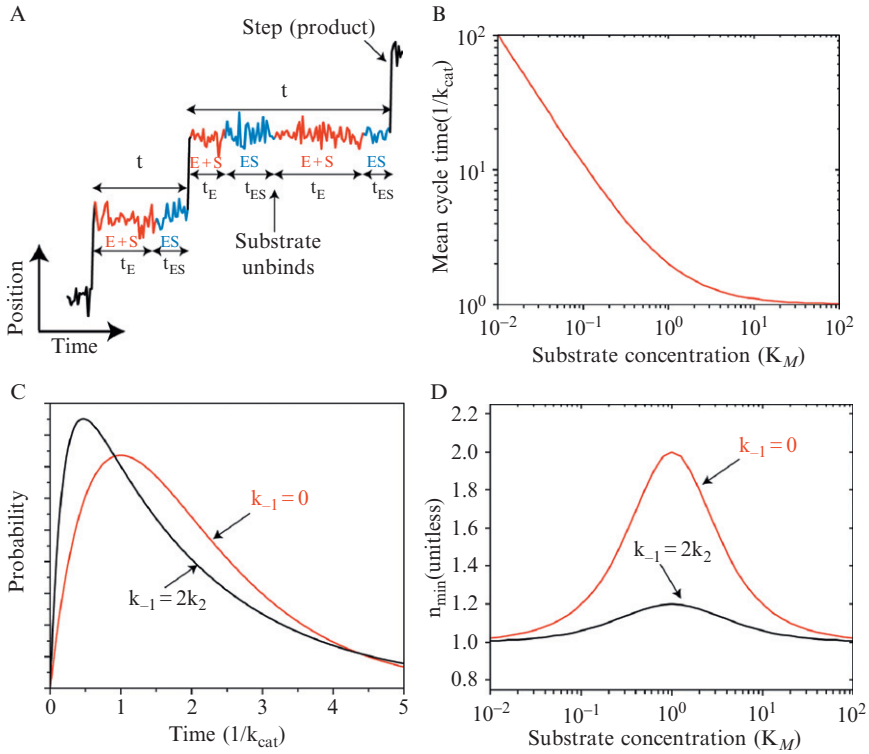

Figure 10.1 Statistical kinetics. (A) A simulated molecular motor trace for the canonical Michaelis–Menten mechanism. Here product formation corresponds to the generation of an observable step—an increase in position. Each full dwell time or cycle completion time, t, can be divided into the lifetimes of the individual kinetic states, t_E and t_{ES}. Fluctuations in t come from both the variations in the individual lifetimes or in different number of visits to each state. (B) Mean cycle completion time versus substrate concentration, measured in units of the maximum rate ($1/k_{cat}$) and the substrate concentration at which the rate is half maximal (K_M). (C) Probability distribution of cycle completion times at a substrate concentration equal to K_M for two different choices of rate constants. Light gray (red online) corresponds to a system in which binding of substrate is irreversible while black correspond to a system in which, on average, two out of three substrate molecules are unbound before catalysis. Despite having the same mean cycle completion time in (B), the fluctuations for these two different systems are distinct. The functional form for these distributions can be found in Chemla *et al.* (2008) and Xie (2001). (D) A statistical measure of enzymatic fluctuations, n_{min}, as a function of substrate concentration for two different choices of rate constants. Again, these curves are distinct, indicating that while measurements of the mean could not distinguish these mechanisms, measurements of fluctuations can. In contrast to the distributions in (C) simple features of the curves in (D) reveal the difference in binding properties between the two models. The functional form for these curves is described below.

In this case, the cycle completion time is often called a *dwell time* or a *residency time* since during this time the motor resides or dwells at a single place along the molecular track. (Technically speaking the step itself can take some time to complete, just as any kinetic transition requires a small, but finite time to be completed; however, we will ignore this time since it is typically much smaller than the lifetime of the general kinetic state.) This formalism also applies to systems in which there are multiple detectable events per cycle, such as the switching between two enzymatic conformations revealed by fluorescence resonance energy transfer (FRET).

In traditional steady-state kinetics, one would write out the set of coupled, first-order differential equations for the concentration of each of the species, E and ES—equations that describe how the concentration of one species changes continuously into the next species—and then assume that the concentrations of one or various intermediate forms have reached a steady state, that is, are constant in time. This assumption allows differential equations to be changed into algebraic equations, and quantities such as the average rate of product formation to be calculated in terms of substrate concentrations and individual kinetic rates (for a comprehensive discussion, see Segel, 1975). From a single-molecule perspective, however, this picture is flawed. Continuous changes of one species into another are nonphysical since chemical transformations that lead to product formation are discrete, punctate events. Moreover, the system can never be in steady state: at any given time a single enzyme is either in one state or another, not in some constant fraction of both.

Rather than considering a continuous flow of one species to the next, it is more useful, on a single-molecule level, to think of a series of discrete paths through the kinetic cycle—paths that consist of consecutive and discrete transformations of one species into another (i.e., discrete hops between different kinetic states). For example, the above kinetic scheme implies that the following two diagrams represent valid microscopic paths to product formation:

$$\begin{aligned} &\mathrm{E} + \mathrm{S} \rightarrow \mathrm{ES} \rightarrow \mathrm{E} + \mathrm{P}, \\ &\mathrm{E} + \mathrm{S} \rightarrow \mathrm{ES} \rightarrow \mathrm{E} + \mathrm{S} \rightarrow \mathrm{ES} \rightarrow \mathrm{E} + \mathrm{P}. \end{aligned} \tag{10.2}$$

In the first path the enzyme binds substrate and then immediately forms product, whereas in the second path it unbinds this substrate unproductively and has to rebind substrate before making product. Figure 10.1A depicts one possible way in which these two paths might produce an experimental signal. The first dwell represents the first path in Eq. (10.2), while the second represents the second path. It is clear that even for this simple example, there are an infinite number of microscopic paths, representing the infinite number of times, in principle, that the enzyme could release substrate unproductively before completing catalysis.

Now instead of considering the rate at which each species is transformed into the next, we consider the time the enzyme exists as each species. The advantage of this subtle shift in perspective is that the lifetime for each of the above pathways is just the sum of the individual lifetimes of the states visited. For example, the total cycle completion time for the first pathway is $t = t_E + t_{ES}$, where t_E and t_{ES} are the individual lifetimes of the empty and substrate bound states, respectively. The cycle completion time for the second pathway is $t = t_E + t_{ES} + t'_E + t'_{ES}$, where t'_E and t'_{ES} are the distinct lifetimes of these states during the second visit. Despite this shift in perspective, a formal connection can be made between this picture and the first-order differential equations that govern the concentrations of each species by replacing the concentration of any given species with the probability of being in that state (Chemla *et al.*, 2008; Qian, 2008; Schnitzer and Block, 1995).

2.2. Basic statistics of the cycle completion time

In this formalism, it is clear that the statistical nature of the cycle completion time arises in two fashions. First, the individual lifetimes of the states are themselves stochastic; thus, their sum will also be a random variable. Second, the number of times that a given kinetic state is visited in a complete cycle is variable as well. Thus, the number of times that a given lifetime contributes to the total cycle completion time may vary each cycle. Stated simply, the basic statistical problem that arises in enzyme dynamics is one in which the total cycle completion time is a sum of random variables in which the number of terms in the sum is itself variable. Mathematically, if a given kinetic scheme has N states, the cycle completion time, τ, is

$$\tau = \sum_{i=0}^{n_1} t_{1,i} + \sum_{i=0}^{n_2} t_{2,i} + \cdots + \sum_{i=0}^{n_j} t_{j,i} + \cdots + \sum_{i=0}^{n_N} t_{N,i}, \tag{10.3}$$

where n_j is the number of times that kinetic state j was visited during the specific cycle and $t_{j,i}$ represents the stochastic lifetime of state j during the ith visit to that state. While the individual lifetimes are a property of distinct kinetic states, the number of visits to each state is, in general, a function of how all of the kinetic states are interconnected; thus, in general, the lifetimes will be independently distributed random variables while the state visitation numbers will not.

Despite its simplicity, Eq. (10.3) completely determines the relationship between the statistical properties of the cycle completion time and the statistical properties of the hidden kinetic states that compose the cycle. This connection is what forms the basis for the basic premise of statistical kinetics: Measurements of the statistics of the total cycle completion time can provide insight into the properties of the hidden states that

compose the cycle. Figure 10.1B–D illustrates this principle. Two different kinetic mechanisms may have the same mean cycle completion time with the same substrate concentration dependence (Fig. 10.1B), yet these two mechanisms are clearly distinguishable via different statistical measures of the fluctuations in the cycle completion time (Fig. 10.1C and D).

Because of its generality, Eq. (10.3) allows the derivation of some basic relationships between the statistics of the cycle completion time and the statistical properties of the individual kinetic states, properties that are independent of the specifics of a given kinetic mechanism. These relations will elucidate how the statistical properties of the individual states generate the statistical properties of the total cycle completion time. In the Appendix, we show that, unsurprisingly, the mean cycle completion time for the arbitrary N state kinetic mechanism is simply the sum of the average amount of time spent in each kinetic state:

$$\langle \tau \rangle = \langle n_1 \rangle \langle t_1 \rangle + \langle n_2 \rangle \langle t_2 \rangle + \cdots + \langle n_N \rangle \langle t_N \rangle. \tag{10.4}$$

We then extend this calculation, in the Appendix, to the next statistical moment, the variance. We find that an N state kinetic cycle has a cycle completion time variance of

$$\begin{aligned} \langle \tau^2 \rangle - \langle \tau \rangle^2 = {} & (\langle t_1^2 \rangle - \langle t_1 \rangle^2)\langle n_1 \rangle + \cdots + (\langle t_N^2 \rangle - \langle t_N \rangle^2)\langle n_N \rangle \\ & + (\langle n_1^2 \rangle - \langle n_1 \rangle^2)\langle t_1 \rangle^2 + \cdots + (\langle n_N^2 \rangle - \langle n_N \rangle^2)\langle t_N \rangle^2 \\ & + 2(\langle n_1 n_2 \rangle - \langle n_1 \rangle \langle n_2 \rangle)\langle t_1 \rangle \langle t_2 \rangle + 2(\langle n_1 n_3 \rangle - \langle n_1 \rangle \langle n_3 \rangle) \\ & \langle t_1 \rangle \langle t_3 \rangle + \cdots + 2(\langle n_{N-1} n_N \rangle - \langle n_{N-1} \rangle \langle n_N \rangle)\langle t_{N-1} \rangle \langle t_N \rangle. \end{aligned} \tag{10.5}$$

This expression indicates that the fluctuations in the total cycle completion time arise from both the inherent fluctuations in the individual lifetimes, that is, terms that go as $\langle t_i^2 \rangle - \langle t_i \rangle^2$, and fluctuations in the number of visits to a given kinetic state, that is, terms that go as $\langle n_i^2 \rangle - \langle n_i \rangle^2$. As discussed above, the number of visits to a given state may depend on the number of visits to neighboring kinetic states. The correlation terms in the final two lines capture these interstate relationships. The fact that there are correlation terms in state visitation but not in lifetimes again reflects the fact that lifetimes are properties of individual states while the number of visitations are a function of how states are connected to one another—the topology of the kinetic mechanism.

2.3. The "memory-less" enzyme

At this point, the connection between the statistical properties of the cycle completion time and the statistical properties of the individual kinetic states is independent of the properties of these states. Now we must insert some basic assumptions about enzymatic dynamics in order to determine the

statistical properties of the individual kinetic states. First, we require that the properties of a given kinetic state are independent of time—the enzyme has no memory of how long it has lived in a specific state. Second, we require that the properties of a given kinetic state are independent of the past trajectory of the enzyme—the enzyme has no memory of the states from which it came. Thus, we assume that the transitions between kinetic states represent a Markov process, that is, the enzyme is completely "memory-less."

These assumptions arise naturally out of the basic physical properties of the energy landscape that determines the dynamics of enzymes. Protein conformational dynamics can be thought of as diffusion on a complex, multidimensional, energy landscape (Henzler-Wildman and Kern, 2007). In general, the exact position of the system on such a landscape provides a form of memory—the local potential determines the probability of fluctuating in one direction or another, and positions closer to a barrier are more likely to result in fluctuations across that barrier than positions further away. However, it is generally thought that the energy landscapes of most enzymes are characterized by a variety of local minima (Fig. 10.2). If these

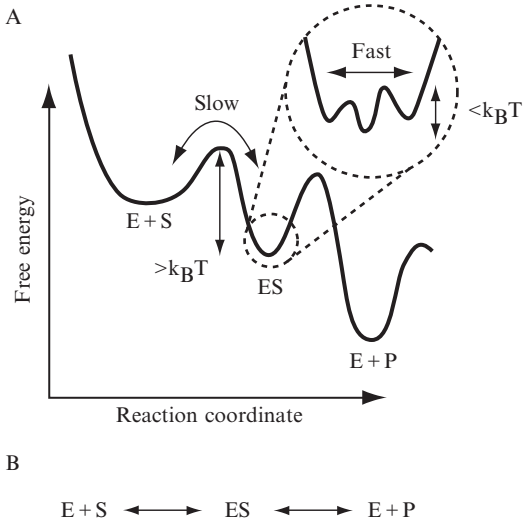

Figure 10.2 The "memory-less" enzyme. (A) One-dimensional projection of the energy landscape of an enzyme. Kinetic states correspond to local minima or energy wells in this landscape. Wells are typically characterized by properties of the enzyme, such as empty (E + S) or substrate bound (ES), that are common to all conformations within a given well. Relatively large barriers between wells ($> k_B T$) creates a separation of timescales in which the enzyme fluctuations many times within a single well before successfully transitioning out of the state. These internal fluctuations "erase" any memory of the amount of time spent in a given well or the past trajectory that brought the system to that well. (B) Because of these physical properties of the energy landscape, the dynamics can often be modeled as discrete hops between distinct kinetic states.

minima are surrounded by energy barriers that are few times the thermal energy available in the bath ($k_B T$), then there is a natural separation of timescales in the diffusional dynamics (Henzler-Wildman and Kern, 2007): Fluctuations can be divided into fast fluctuations within a well and slow fluctuations between wells. Because fluctuations within a well are much faster than fluctuations between wells, the enzyme will fluctuate many times within a well before leaving it. These fluctuations effectively average over differences in transition rates due to the position of the enzyme within a well, erasing any "memory" of its lifetime in the well or the state from which it came. These local minima, or collections of local minima, can be identified as *kinetic states* of the enzyme, and the goal of kinetic modeling is to count the number of these wells, determine their interconnectedness and the height of barriers between them, and identify the chemical state of substrates or products in these wells.

There are a few situations in which these assumptions may not be valid, and discrete hops between kinetic states with no memory may not be a suitable description of enzymatic dynamics (Kolomeisky and Fisher, 2007). For example, if the barriers between local minima are not large, the lifetime of a given kinetic state may not be significantly larger than natural relaxation time of the conformational dynamics within that state; and it may display transition probabilities and lifetimes that depend on the identity of the previous state—assuming distinct conformational states can even be identified. As a rule of thumb, the natural timescales for internal fluctuations within a kinetic state are $\sim$ns–μs (Henzler-Wildman and Kern, 2007); thus, it seems highly unlikely that enzymatic dynamics on the ms and larger timescales will be significantly affected by these complications. However, as experimental techniques push the limit of temporal and spatial resolution, the observation of such transient states may become increasingly common. In such cases, it is likely that more complicated modeling efforts that seek to include this inherent diffusive process will be needed (Karplus and McCammon, 2002; Xing *et al.*, 2005).

2.4. Lifetime statistics

These basic Markovian assumptions completely determine the statistical properties of both the lifetimes and visitation numbers. Since each kinetic state has no memory of when the system arrived in this state, or from where it arrived, the rate at which the system leaves the state (the transition probability per unit time) should be constant and, therefore, the probability of finding the system in that state should decrease exponentially in time. Then, the probability density of observing a lifetime t for that state or, equivalently, its lifetime distribution should be

$$\psi(t) = \langle t \rangle^{-1} \mathrm{e}^{-t/\langle t \rangle}. \tag{10.6}$$

Note that the only number that characterizes the distribution of lifetimes, $\psi(t)$, is the mean lifetime, $\langle t \rangle$. Thus, all of the statistical moments of the distribution are completely specified by this mean. For example, the variance in the lifetime is simply the mean squared:

$$\left\langle t^2 \right\rangle - \langle t \rangle^2 = \int_0^\infty (t - \langle t \rangle^2)\psi(t)\mathrm{d}t = \langle t \rangle^2. \tag{10.7}$$

This relationship between the variance and the mean has physical implications. States with longer mean lifetimes naturally have larger fluctuations in the lifetime. As we will discuss below, this fundamental relationship between the mean and variance of kinetic lifetimes will prove useful in placing limits on possible mechanisms from statistical measures of fluctuations.

2.5. State visitation statistics

Just as the lack of memory determined the basic statistics of the lifetimes, this assumption also determines the statistics of the number of times an enzyme visits a specific kinetic state during a single cycle. To illustrate this point, consider the options available to an enzyme in a given kinetic state. The enzyme can leave that state and complete the cycle without returning to that state. Let this event occur with a probability, p. Alternatively, the enzyme can return to that state without completing the cycle. Since there are only two options, this event will occur with a probability $1 - p$. With no memory of how it arrived in a given state, once the enzyme returns to this state, the probabilities of completing the cycle must be the same, and completing the cycle must again occur with the same probability p. Thus, each subsequent visit to a given kinetic state occurs with a geometrically decreasing probability.

This argument assumes that the enzyme visits a given kinetic state every cycle, that is, the given state is *on-pathway*. However, there are kinetic pathways in which the enzyme can complete the cycle without visiting a given state—that is, that state is off-pathway or the enzyme has multiple parallel pathways to cycle completion (see Figs. 10.3B or 10.5D and E). Once the system visits this state, however, the above argument applies, so we need only to include a probability, p_0, for visiting the state for the first time. For on-pathway states, $p_0 = 1$.

Combining these arguments yields the visitation statistics for a given kinetic state, that is, the probability that the system completes its cycle with only n visitations to a given kinetic state:

$$P(n) = p_0(1 - p)^{n-1}p. \tag{10.8}$$

The first term is the probability that the system visits the specific state for the first time. The second term represents the probability of visiting the given state $n-1$ times without completing the cycle while the final term represents the probability of actually completing the cycle from this state. Equation (10.8) applies only for $n \geq 1$, that is, the system visits the state at least once. For no visits to the state, $n = 0$, $P(n = 0) = 1 - p_0$, the probability of not visiting that state for the first time. Summing Eq. (10.8) over all possible visitation numbers and including this $P(0)$ term produces 1, as expected for a normalized distribution. Interestingly, this is the discrete form of the exponential distribution above, and represents the only "memory-less" discrete distribution.

Given the discrete distribution in Eq. (10.8), we can determine the statistics of visiting a given state. The average number of times that a system visits a given kinetic state is

$$\langle n \rangle = \sum_{n=0}^{\infty} nP(n) = \frac{p_0}{p} \tag{10.9}$$

and the variance is

$$\left\langle n^2 \right\rangle - \langle n \rangle^2 = \sum_{n=0}^{\infty} (n - \langle n \rangle^2) P(n) = p_0 \frac{(1 - p_0 + 1 - p)}{p^2}. \tag{10.10}$$

If a given state is on-pathway, that is, a mandatory state, then $p_0 = 1$ for this state and these expressions simplify. In this case the mean number of visitations completely determines the higher statistical moments of the visitation number. In particular, the variance becomes

$$\left\langle n^2 \right\rangle - \langle n \rangle^2 = \langle n \rangle (\langle n \rangle - 1). \tag{10.11}$$

This expression is almost the mean squared, as is the case with the lifetime statistics. The distinguishing term, $\langle n \rangle - 1$, reflects the discrete nature of the number of visits: If $\langle n \rangle = 1$, the system visits the state once and only once, and the variance in the visitation number must be zero. Again, these statistical properties have clear physical implications. States that are visited on average more frequently will naturally have larger fluctuations in the number of visits.

For off-pathway states or states that compose parallel catalytic pathways, $p_0 < 1$, and the variance in visitation number is not uniquely determined by the mean number of visits. Rather

$$\left\langle n^2 \right\rangle - \langle n \rangle^2 = \frac{1}{p_0} \langle n \rangle (\langle n \rangle - p_0 + \langle n \rangle (1 - p_0)). \tag{10.12}$$

This expression captures the distinction between the visitation statistics of on-pathway and off-pathway states. First, it is important to note that because $0 \leq p \leq 1$, Eq. (10.9) implies that $p_0 \leq \langle n \rangle$, and this expression is always positive, as expected. More interestingly, by comparing Eqs. (10.11) and (10.12), it becomes clear that, for all permissible values of p_0, the variance in the visitation number for an off-pathway state is always *larger* than the visitation number for an on-pathway state with the same average number of visits. Thus, the number of visits to off-pathway states is naturally more stochastic than on-pathway states, and the presence of an off-pathway state or parallel pathways will increase the fluctuations in a system.

3. Characterizing Fluctuations

In Section 2, we provide the connection between the statistical properties of the hidden kinetic states and the statistical properties of the total cycle completion time. However, before statistical measures of the cycle completion time can be used to constrain the properties of the kinetic states that compose the kinetic mechanism, care must be taken to characterize these fluctuations in a useful and unbiased way. In particular, not all statistical measures may be as convenient in constraining kinetic mechanisms. Moreover, the characterization of cycle completion or dwell times becomes more subtle when the enzyme has multiple observable outcomes, as is becoming increasingly common. In this section, we discuss these issues.

3.1. Fitting distributions

In the above formalism, we considered only the statistical moments of the cycle completion time; however, one can also compute the full probability distribution for the observation of different cycle completion times. This quantity is often referred to as a *dwell time distribution*, and because all of the statistical moments can be calculated from this distribution, it clearly contains more kinetic information than a subset of the moments. For example, we show in Fig. 10.1 that with the proper choice of rate constants, two different kinetic models (irreversible binding or reversible binding) can have the same mean time to complete the enzymatic cycle (Fig. 10.1B) but very different dwell time distributions (Fig. 10.1C). However, while the differences between these two distributions are clear when they are compared directly, imagine that one has only a single distribution, as would be the case for real data. How would mechanistic information be extracted from this distribution?

One obvious possibility is to derive a function that describes the dwell time distribution and fit this expression to the measured distribution to

extract information about the kinetic mechanism. It turns out that, under very general assumptions, it can be shown that a general kinetic model with N states will have a dwell time distribution that is described by a sum of N exponentials with different relative weights and decay rates (Chemla *et al.*, 2008). Thus, any dwell time distribution can be expressed as

$$\psi(t) = a_1 e^{-\lambda_1 t} + a_2 e^{-\lambda_2 t} + \cdots + a_N e^{-\lambda_N t}, \tag{10.13}$$

where a_i are the weights for each exponential—these can be positive or negative—and λ_i are the eigenvalues of the system—these values set the natural timescales for the cycle completion time and must be positive. If multiple eigenvalues happen to be equal, the m terms in Eq. (10.13) that correspond to these m equivalent eigenvalues are replaced by the term $a_i t^{m-1} e^{-\lambda_i t}$. Distributions of this form are known as *phase-type distributions*, and there is a body of literature on the properties of this class of distribution (cf. Neuts, 1975, 1994; O'Cinneide, 1990, 1999).

As an interesting aside, Eq. (10.13) provides a measure of the maximum kinetic information contained in a dwell time distribution. Since there are N eigenvalues and $N - 1$ free weights (one weight is fixed by the fact that the distribution must be normalized), Eq. (10.13) implies that even the most accurate fit can only extract $2N - 1$ constraints on the system. However, it is easy to imagine an N state kinetic model that has many more than this number of kinetic rates (if every state is connected to every other state—the maximum connectivity—then there are $N(N - 1)$ kinetic rates). Thus, even though much can be determined about these hidden kinetic transitions from enzymatic fluctuations, information is lost in the process of examining only the total cycle completion time. However, some of this information can be restored if there are multiple observables in each cycle. In such a situation, there will be more than one dwell time distribution, and it is possible that additional information can be determined from fits to these multiple distributions.

With Eq. (10.13), we have the appropriate distribution to fit to any measured cycle completion time distribution. However, in practice, there are several problems with using this expression. First, the general dwell time distribution has too many free parameters to be well constrained by typical amounts of data. This is complicated by the fact that the number of terms in the sum, that is, the number of kinetic states in the system, is generally not known *a priori*; thus, one must in principle fit the measured distribution using different numbers of exponentials, comparing the results to determine which number better fits the data. While there are statistically sound methods for performing this comparison (Yamaoka *et al.*, 1978), there is no guarantee that the data will constrain the fits well enough to determine the appropriate number of exponentials. Moreover, if the appropriate number of exponentials is not uniquely determined, then it is not clear how the fit values should be interpreted.

One common solution to this problem is to assume a functional shape for the dwell time distribution that contains far fewer parameters. For example, it is becoming increasingly common to use the gamma distribution to describe such distributions:

$$\varphi(t) = \frac{k^N t^{N-1}}{\Gamma(N)} \mathrm{e}^{-kt}, \tag{10.14}$$

where $\Gamma(N)$ is the gamma function. Since the gamma distribution only has two free parameters—an average rate, k, and the "number" of kinetic states, N—it is often quite well constrained by the data. However, the gamma distribution is the correct functional form for the dwell time distribution only when the underlying kinetic model has N states with equal lifetimes $1/k$ which are connected via irreversible transitions. In other words, the gamma distribution is the correct distribution *only* when the kinetic mechanism is of the form

$$1 \xrightarrow{k} 2 \xrightarrow{k} 3 \xrightarrow{k} 4 \xrightarrow{k} \cdots \xrightarrow{k} N. \tag{10.15}$$

Unfortunately, it is often not the case that such a mechanism applies to an average kinetic model. First, it is extremely rare that all kinetic rates are identical, and even if this happened to be true for one substrate concentration, it would not be true for arbitrary substrate concentrations. Second, an irreversible transition requires a large energy input, and it is rare for a kinetic mechanism to have only transitions that involve such large energies. Thus, while the gamma distribution is an excellent way to characterize the "shape" of a distribution, it is unlikely that this is often the correct functional form for the distribution, and, thus, it is unclear how the fit values should be interpreted.

Despite these caveats, it is clear that the full cycle completion time distribution contains the most kinetic information available; thus, efforts should be taken to improve methods for extracting this information. There is a recent technique that attempts to extract rate information from dwell time distributions without fitting the distribution. With this method one calculates a rate spectrum—the amplitude of exponentials at each decay rate—via direct numerical manipulation of the distribution (Zhou and Zhuang, 2006). This method is analogous to the Fourier method for disentangling different frequency components from complicated time series. While initial results appear promising, this technique is only reported to perform well when the decay rates are separated by an order of magnitude (Zhou and Zhuang, 2006), which is typically not the case in real experimental data. However, future developments in this direction seem promising.

There is one final issue with fitting distributions. Ignoring the complications with actually extracting the various decay rates, that is, the eigenvalues λ_i and the different weights for each exponential a_i, it turns out that it is difficult, and in some cases impossible, to analytically relate these values to the kinetic rates of a specific kinetic model. The problem is a mathematical one. Calculating analytical expressions for these eigenvalues and weights corresponds to solving for the roots of a polynomial expression with order equal to the number of kinetic states in the model (Chemla *et al.*, 2008). However, Abel's impossibility theorem (Abel, 1826) states that there is no general analytical solution for this problem if the polynomial is of order five or higher. Thus, if there are five or more kinetic states in a kinetic mechanism, the eigenvalues and exponential weights *cannot* in general be expressed in terms of the individual kinetic rates analytically. Of course, this does not mean that a numerical connection cannot be made—there are many techniques capable of calculating these values for any numeric choice of kinetic rates for any mechanism (Liao *et al.*, 2007)—however, since the proper choice of rate constants (and number of rate constants for that matter) are typically not known *a priori*, it is unclear how useful numerical solutions would be.

3.2. Calculating moments

Many of the problems associated with extracting information from the dwell time distribution directly can be relaxed by first calculating properties of this distribution such as its statistical moments or, as we will show below, properties such as its "shape." While it is clear that the moments of the distribution will contain a subset of the information contained in the full distribution, the advantage is that these moments are *model-independent*, that is, a basic kinetic model or form for the distribution does not need to be assumed to calculate the mean dwell time or the variance in the dwell times. Moreover, calculation of these moments is simple and straightforward, and there are well-established techniques to estimate the stochastic uncertainty in these moments directly from the measured data itself (Efron, 1981; Efron and Tibshirani, 1986). Since the uncertainty in the moments can be easily calculated, one can use only the moments that are well constrained by the data, conveniently circumventing issues with unconstrained fits to poorly determined distributions. Finally, methods now exist for calculating analytical expressions for the moments of the cycle completion time for any kinetic mechanism, no matter how complex (Chemla *et al.*, 2008; Shaevitz *et al.*, 2005).

There is one notable disadvantage to characterizing fluctuations via statistical moments. In all measurements, there is a natural dead-time, a time below which events are too quick to observe experimentally. When fitting distributions, one can address this problem simply by fitting only over

time durations that are known to be measured accurately. However, a similar method does not exist for calculating moments, and dead-times can introduce bias into the estimation of statistical moments. Fortunately, it is relatively simple to estimate the relative size of this error directly from the data itself with only a few assumptions. In the Appendix, we derive expressions for estimating this systematic bias.

3.3. Multiple pathways and multiple steps

The above discussion makes an implicit assumption: that each random dwell time is derived from the same kinetic mechanism, that is, stochastic passage through the same kinetic states. This is an innocuous assumption when the enzyme takes a single type of step, for example, forward steps of uniform size, since it is likely that identical steps are produced by the same kinetic pathway. However, it is becoming increasingly clear that real enzymes display more complicated behaviors. For example, filament based cargo transport proteins such as kinesin, myosin, and dynein have now been observed to take both forward and *backward* steps and steps of varying size (Cappello *et al.*, 2007; Carter and Cross, 2005; Clemen *et al.*, 2005; Gennerich and Vale, 2009; Gennerich *et al.*, 2007; Mallik *et al.*, 2004; Reck-Peterson *et al.*, 2006; Rief *et al.*, 2000; Yildiz *et al.*, 2008). Moreover, multiple observable events may be produced within a single kinetic cycle, creating in effect multiple classes of dwell times, as is observed for the packaging motor of the bacteriophage φ 29 (Moffitt *et al.*, 2009). Finally, some single-molecule measurements naturally observe multiple events within a single cycle, such as the transitions of a system between multiple FRET states (Cornish and Ha, 2007; Greenleaf *et al.*, 2007) or the folding and refolding of nucleic acids (Li *et al.*, 2008; Woodside *et al.*, 2008) or proteins (Cecconi *et al.*, 2005).

When there are multiple types of steps or observable states, there may also be multiple classes of dwell times, that is, cycle completion times that originate from different kinetic pathways. Clearly, the combined statistical analysis of dwell times derived from different kinetic pathways will provide little insight into each pathway individually. Thus, before one can extract any kind of mechanistic information from fluctuations, one must be sure that dwell times that are generated by the same kinetic mechanism and result in the same basic type of step are properly sorted. However, it is not immediately obvious how this should be done. Does one simply sort dwells based on the type of the step following the dwell? Or is it possible that enzymatic dynamics may have some memory of more distant steps, perhaps the type of step before the dwell as well? Moreover, how does one distinguish between these possibilities; are there simple statistics that can be calculated directly from the data that would allow these different possibilities to be addressed and the appropriate classification scheme to be determined?

It has been recently recognized that there are three basic statistical classes of enzymatic dynamics (Chemla *et al.*, 2008; Linden and Wallin, 2007; Tsygankov *et al.*, 2007) in which there are multiple enzymatic outcomes. For simplicity, we consider different outcomes to represent steps of different sizes, but this discussion can be easily generalized to any measurable enzymatic outcome, for example, off-pathway pauses or dissociation events. In the first statistical class, the size of the step or its direction have no relationship to the hidden kinetic events that must occur before that step (Shaevitz *et al.*, 2005). In this case, the statistics of the dwell times will be uncorrelated to the type of the subsequent step or event. We term such statistics *uncorrelated*. Figure 10.3A contains an example kinetic scheme that would display these statistics.

There are two different classes of correlated statistics. First, the statistics of the dwell time may be related to the type of the subsequent step but not the identity of the previous step. We term such statistics *unconditional* because the statistical properties of the dwells depend only on the identity of the following step and are not conditional on the identity of the previous step. For this statistical class the dwells must be sorted by the identity of the subsequent step. To generate these statistics, each different step must have a distinct kinetic pathway (or subset of a kinetic pathway) in order to generate the distinct dwell statistics, yet all kinetic pathways must start in the same kinetic state. See Fig. 10.3B for an example of a kinetic mechanism which would display such statistics.

Finally, it is possible for an enzyme to actually remember the type of its previous step. This emergent memory arises because different steps need not place the enzyme in the same hidden kinetic state. In this case, the dwell times are *conditional* on both the type of step before and after the dwell, and these durations must be sorted by the identity of both steps. This case is likely for enzymes in which formation of product generates a forward step but the reverse reaction, the recatalysis of substrate from product, corresponds to a backward step (Linden and Wallin, 2007; Tsygankov *et al.*, 2007). See Fig. 10.3C for an example of an enzymatic mechanism that generates correlated, conditional statistics. In this example, a forward step ($+d$) coincides with the formation of product and places the enzyme in the empty state (E) whereas a backward step ($-d$) coincides with the reformation of substrate from product and places the enzyme not in the empty state (E) but in the substrate bound state (ES). Because forward and backward steps place the enzyme into different kinetic states, the microscopic kinetic trajectories (as in Eq. (10.2)) that end in the next step will be different. Thus, the kinetics of a single dwell time will depend on the initial type of step.

Figure 10.3 shows that the individual stepping traces can appear quite similar by eye despite the different mechanisms that produce these traces. Fortunately, there are well-defined statistical properties of such traces that

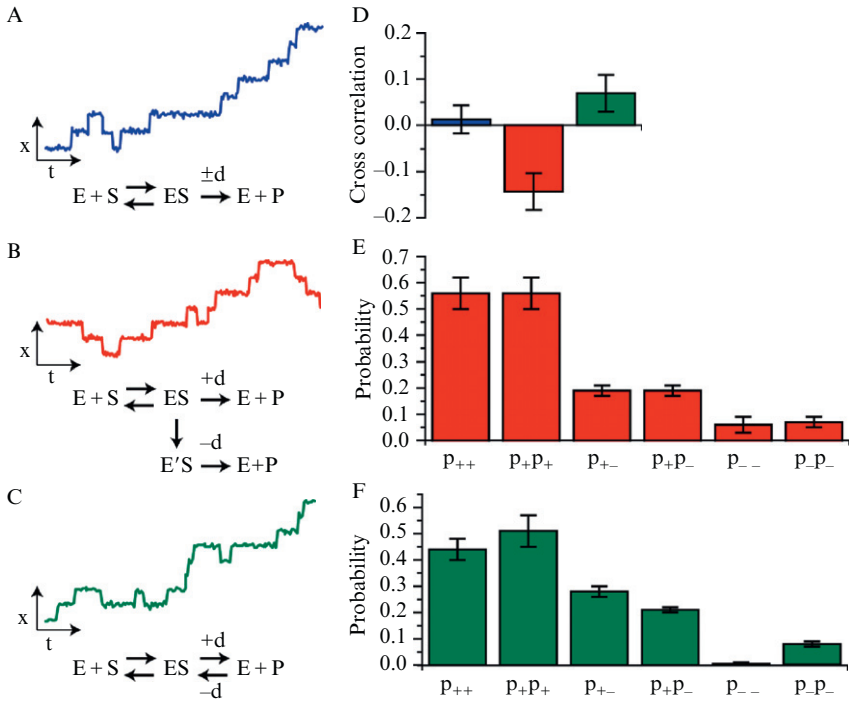

Figure 10.3 Statistical classes of enzymatic dynamics. (A–C) Simulated stepping traces for different mechanisms capable of generating forward and backward steps. (A) The steps and dwell times are uncorrelated because the same kinetic mechanism generates both forward ($+d$) and backward ($-d$) steps. (B) The steps and dwell times are correlated but the dwells are unconditional on the type of the previous step because each step returns the system to the same kinetic state (E). (C) The steps and dwells are correlated and the dwells are conditional on the type of the previous step because forward and backward steps return the enzyme to different kinetic states, E and ES, respectively. (D) Cross-correlation between steps and dwell times (arbitrary units) for the different stepping traces. Correlation values correspond to the different mechanisms in (A) - (C) from left to right, respectively (color online). (E) Single-step probabilities and pair probabilities for the mechanism in (B). (F) Single-step probabilities and pair probabilities for the mechanism in (C). Despite the fact that the individual stepping traces are very similar, the statistical measures in (D)–(F) can clearly distinguish the differences between these mechanisms. Values in (D)–(F) were calculated from kinetic simulations containing 1000 steps with rates set such that the probability of taking a forward step is 0.75 and the average velocity is the same between the different mechanisms. Error bars represent the standard deviation in the various statistics estimated from 100 repetitions of the simulations.

can clearly distinguish the statistical class of the enzyme. For example, if the type of the subsequent step is uncorrelated with the subsequent dwell time, then the appropriate statistical class is *uncorrelated*. In this case, step size is

independent of the kinetic pathway, and dwell times can be analyzed together. Mathematically, this can be tested in a variety of ways, but if there is no cross-correlation between step type or size and dwell time, then the steps are uncorrelated, e.g.

$$\langle dt \rangle - \langle d \rangle \langle t \rangle = 0. \tag{10.16}$$

Here d is the step size (or the type of the enzymatic outcome) and t is the dwell or cycle completion time. Figure 10.3D shows that this statistic clearly indicates that the stepping traces for the mechanism in Fig. 10.3A are uncorrelated while the other two traces are correlated.

Often experimental noise broadens step size distributions, and it can be difficult to determine if multiple steps are actually present. In addition, a portion of the steps may have dwells too small for direct detection, giving the appearance that the enzyme can generate steps of multiple sizes. In both cases, the steps will be uncorrelated with the preceding dwell times. Thus, the violation of Eq. (10.16) is a strong criterion for establishing that an enzyme generates multiple types of steps and that such steps are not experimental artifacts.

If the steps are correlated, that is Eq. (10.16) is violated, there is a simple statistical test to determine if the statistical class is conditional or unconditional. Simply compute the probability of observing each pair of events. For the forward and backward step examples considered in Fig. 10.3, this will produce four probabilities, for example, the probability of observing two forward steps in a row,p_{++}, a forward step followed by a backward step, p_{+-}, a backward step followed by a forward step, p_{-+}, and finally two backward steps in a row, p_{--}. These probabilities are computed by simply counting the number of each type of event and dividing by the total number of events, though care must be taken in computing the exact number of events if statistics are small, as pointed out by Tsygankov *et al.* (2007). These quantities can then be compared to the probability of observing a given outcome independent of the previous type of step. For the forward and backward stepping example, this would be the probability of taking a forward step, p_+, and a backward step, p_-, which are again calculated from the number of each step type divided by the total number of steps. If the statistics of the system are unconditional, then the probability for observing two types of events in a row will be equal to the product of the probabilities of observing each of these events individually. If the statistics of the system are conditional, this relation will not be true. For the forward and backward stepping example, there are four equalities to test:

$$p_{++} = p_+ p_+, \quad p_{+-} = p_+ p_-, \quad p_{-+} = p_- p_+, \quad \text{and} \quad p_{--} = p_- p_-. \tag{10.17}$$

If these equalities are violated, then the statistics of the system are conditional. If these equalities are upheld by the data, then it is likely that the statistics of the system are unconditional. Figure 10.3D and E illustrates the ability of these statistics to distinguish the unconditional and conditional statistics of the mechanisms in Fig. 10.3B and C. These equalities can be easily extended for systems that have more than two outcomes or observable states.

In addition to providing the correct method for sorting cycle completion times, the specific statistical class of an enzyme places clear constraints on the underlying kinetic mechanism (Chemla *et al.*, 2008). For example, a lack of correlation between steps and dwell times can only occur if these processes are determined independently; thus, the observation of uncorrelated statistics implies that the kinetic pathway for each type of step is *identical*. The converse is also true: If the statistics are correlated, then at least a portion of the kinetic pathway that leads to each kinetic event cannot be the same. Moreover, the type of correlated statistics—conditional or unconditional—provides further constraint on the kinetic mechanism. In general, the use of different kinetic pathways to develop different enzymatic outcomes creates an emergent enzymatic "memory." The partial loss of this memory, as is the case in unconditional statistics, requires that all kinetic pathways share at least one common kinetic state after the generation of each step. The memory-less properties of this state is what decouples the statistics of the subsequent dwell time from the identity of the preceding step. While there have been many examples of enzymes that take multiple types and sizes of steps (Cappello *et al.*, 2007; Clemen *et al.*, 2005; Gennerich and Vale, 2009; Gennerich *et al.*, 2007; Kohler *et al.*, 2003; Mallik *et al.*, 2004; Reck-Peterson *et al.*, 2006; Rief *et al.*, 2000; Rock *et al.*, 2001; Sellers and Veigel, 2006; Yildiz *et al.*, 2004), the statistical properties of these steps has so far been underutilized in the analysis of the kinetic mechanism of these enzymes.

4. Extracting Mechanistic Constraints from Moments

Once fluctuations have been properly characterized and measured, the question arises: What can be learned from these statistics? Should candidate models be selected, their properties calculated via the variety of techniques available (Chemla *et al.*, 2008; Derrida, 1983; Fisher and Kolomeisky, 1999; Koza, 1999, 2000; Shaevitz *et al.*, 2005), and then these properties compared to the measured statistics to reject or accept models? This scheme is certainly a viable approach; however, it turns out that, as was seen with the statistical class of an enzyme above, certain properties of the statistics of

enzymatic fluctuations can place constraints on candidate models, even before the properties of these models are calculated. In this section, we discuss a method in which statistical measures of fluctuations can constrain kinetic mechanisms without the assumption of candidate models.

4.1. The randomness parameter and $n_{\min}$

In 1994, Schnitzer and Block (Schnitzer and Block, 1995, 1997; Svoboda *et al.*, 1994) introduced a kinetic parameter related to the first and second moments of enzymatic fluctuations: the randomness parameter, $r = 2D/vd$, where D is the effective diffusion constant of the enzyme, v is the average rate of the enzyme, and d is a normalization constant that determines the amount of product each cycle. For molecular motors, where this expression was first introduced, d is the step size. In this context, the diffusion constant, D, is not the rate at which an enzyme diffuses freely through solution, but a measure of how quickly two synchronized enzymes will drift apart from one another. For example, if two identical motors are started at the same location at the same time, they will gradually separate, due to fluctuations, with a squared distance that increases linearly with time. D is a measure of this diffusive-like behavior (Schnitzer and Block, 1995, 1997; Svoboda *et al.*, 1994).

In the limit that the motor takes a single type of step of uniform size and direction (Schnitzer and Block, 1995; Svoboda *et al.*, 1994), the randomness parameter reduces to a quantity that is a function only of the statistics of the dwell times:

$$r = \frac{\langle \tau^2 \rangle - \langle \tau \rangle^2}{\langle \tau \rangle^2} = \frac{1}{n_{\min}}, \tag{10.18}$$

where $\langle \tau \rangle$ is the mean of the cycle completion time distribution and $\langle \tau^2 \rangle - \langle \tau \rangle^2$ is the second moment, the variance. This quantity is known as the squared coefficient of variation, and is used to characterize fluctuations in a wide range of stochastic systems. As we will see below, it is often more convenient to work with the inverse of this parameter—a quantity that we term $n_{\min}$. This new parameter can be thought as a shape parameter for the dwell time distribution (akin to the parameter N of the Gamma distribution). The smaller the variance, the more "sharply peaked" the distribution is, and the larger the value of $n_{\min}$.

We term this parameter, $n_{\min}$, because it has been shown (Aldous and Shepp, 1987) that it provides a strict lower bound on the number of kinetic states that compose the underlying kinetic model, n_{actual}:

$$n_{\min} \leq n_{\text{actual}}. \tag{10.19}$$

This inequality is worth a moment's inspection. Equation (10.19) states that a weighted measure of enzymatic fluctuations places a *firm* limit on the minimum number of kinetic states in the underlying kinetic model. The implication is that kinetic schemes with different numbers of kinetic states have fundamentally different statistical properties, and these properties can be used to discriminate between these models. Intuitively, this remarkable property arises from the fact that the variance of an exponentially distributed process is simply the mean squared, as discussed above. Thus, for a kinetic system with a single kinetic state (or dominated by one particularly long-lived state), the ratio of the mean squared to the variance is 1, the number of kinetic states in the system. As additional kinetic states are added (or their lifetimes become comparable), the mean increases more quickly than the variance, and the ratio of the mean squared to the variance increases. While Eq. (10.19) was first introduced as a conjecture in the single-molecule literature (Schnitzer and Block, 1995; Svoboda *et al.*, 1994), it has been formally proven in the context of phase-type distributions (Aldous and Shepp, 1987).

One significant advantage of the randomness parameter is that it can be measured even when the individual steps are obscured by noise, and individual cycle completion times cannot be measured (Schnitzer and Block, 1995; Svoboda *et al.*, 1994). Unfortunately, in recent years, several researchers (Chemla *et al.*, 2008; Shaevitz *et al.*, 2005; Wang, 2007) have shown that the randomness parameter is not always equal to the inverse of $n_{\min}$. In particular, variation in the step size or the stepping pathway will result in correction terms that must be added to r in order to reconstitute $n_{\min}$. Moreover, it does not appear that these terms can be measured from trajectories in which the individual turnovers are obscured by noise, limiting the applicability of the randomness parameter. In addition, in the case of steps of differing size, it is no longer unambiguous what value of d should be used to normalize this parameter (Chemla *et al.*, 2008; Shaevitz *et al.*, 2005; Tsygankov *et al.*, 2007). However, if one can observe the individual cycle completion events, as has been assumed here, different step sizes (or reaction outcomes) can be sorted as above, and it is straightforward to calculate the moments, and, thus, $n_{\min}$, for each of the different classes of dwell times.

4.2. Classifying fluctuations

Under any given experimental conditions, $n_{\min}$ places a firm lower limit on the number of kinetic events in the specific kinetic pathway. However, the degree to which $n_{\min}$ varies from the actual number of kinetic events depends on the relative lifetimes and visitation statistics of the different kinetic events. If one or more kinetic states tend to produce larger

fluctuations, that is, because they have longer lifetimes or smaller probabilities of completing the cycle (see Eqs. (10.6)–(10.12) above), then these states will dominate the measured statistics and will tend to lower $n_{\min}$. However, by changing experimental conditions such as substrate concentration or force, it is possible to change the lifetime and visitation statistics of the different states, making different kinetic states the dominate contributors to fluctuations. In this way, it should be possible to vary the concentration of substrate or force and determine limits on classes of kinetic states, such as the number of substrate binding states or the number of force sensitive kinetic states.

This general dependence of $n_{\min}$ or the related randomness parameter on substrate concentration has been widely recognized, and several different expressions for this dependence have been derived for a variety of specific kinetic mechanisms (Chemla *et al.*, 2008; Garai *et al.*, 2009; Goedecke and Elston, 2005; Kolomeisky and Fisher, 2003; Kou *et al.*, 2005; Moffitt *et al.*, 2009; Schnitzer and Block, 1997; Tinoco and Wen, 2009; Xu *et al.*, 2009). However, we have recently shown (Moffitt *et al.*, 2010) that most if not all of these expressions can be combined into a single expression for the substrate dependence of $n_{\min}$. In particular, it appears that most kinetic mechanisms for which the mean cycle completion time follows the Michaelis–Menten expression:

$$\langle \tau \rangle = \frac{K_M + [S]}{k_{cat}[S]} \tag{10.20}$$

have a substrate concentration dependence of $n_{\min}$ described by

$$n_{\min} = \frac{N_L N_S \left(1 + \frac{[S]}{K_M}\right)^2}{N_S + 2\alpha \frac{[S]}{K_M} + N_L \left(\frac{[S]}{K_M}\right)^2}. \tag{10.21}$$

Here k_{cat} and K_M are the Michaelis–Menten parameters, which set the maximum rate of a reaction and the substrate concentration at which the rate is half-maximal, respectively. Just as these constants contain the specifics of each kinetic mechanism, the new macroscopic constants for $n_{\min}$, N_L, N_S, and α, contain all of the details of each kinetic mechanism—that is, the specific kinetic rates. Figure 10.4 illustrates the possible shapes permitted by Eq. (10.21) and provides a geometric interpretation of these new kinetic parameters. N_S is the value of $n_{\min}$ at saturating substrate concentrations, N_L is the value of $n_{\min}$ at limiting substrate concentrations, and α controls the height of the peak between these two limits. If $\alpha = 0$, then the peak value is the sum of the two limits whereas if $\alpha > 0$, the peak value is smaller.

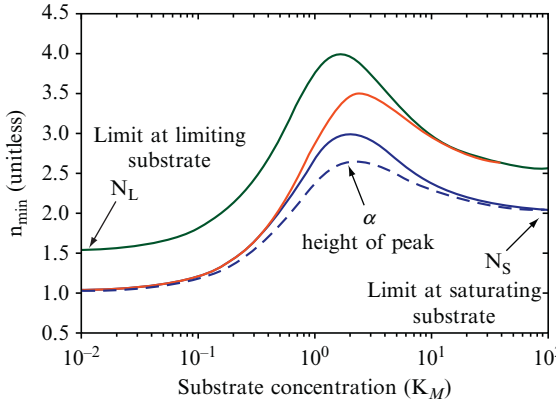

Figure 10.4 n_{min} versus substrate concentration. Different potential curves for n_{min} versus substrate concentration (measured in units of K_M). N_L controls the value of n_{min} at asymptotically low substrate concentration while N_S sets the value of n_{min} at saturating substrate concentration. The value of α controls the height of the peak between these two limits. For $\alpha = 0$, the maximum value of n_{min} is $N_L + N_S$. For $\alpha > 0$, the peak value is less than the sum of the asymptotic limits. The bottom most curves (blue online) correspond to $N_L = 1$, $N_S = 2$, and $\alpha = 0$ (solid) or $\alpha = 0.2$ (dashed). The light gray curve (red online) corresponds to $N_L = 1$, $N_S = 2.5$, and $\alpha = 0$, while the top most curve (green online) corresponds to $N_L = 1.5$, $N_S = 2.5$, and $\alpha = 0$.

4.3. Mechanistic constraints

These new kinetic parameters allow the classification of enzymatic dynamics based on fluctuations, just as the Michaelis–Menten parameters allow such a classification based on measurements of the mean rate versus substrate concentration. However, in contrast to the Michaelis constants, these new parameters also provide clear constraints on the underlying kinetic mechanism. To illustrate these constraints, Fig. 10.5 lists a variety of common kinetic mechanisms and Tables 10.1 and 10.2 list the different kinetic parameters as a function of the individual kinetic rates.

The trends for the kinetic parameters of n_{min} in Table 10.2 can be summarized as follows. First, depending on the kinetic model, each of the different kinetic parameters can be complicated functions of the kinetic rates, just as the Michaelis–Menten parameters (Table 10.1). It is in this fashion that the details of a given kinetic model are "hidden" in these parameters. Moreover, by changing the relative values of the kinetic rates, these parameters can be changed continuously, which implies that N_L and N_S need not be integers. However, this observation does not imply that these parameters can take any value for a specific kinetic mechanism. Rather inspection of these expressions reveals that they have clear upper and lower limits. These limits are included in Table 10.2.

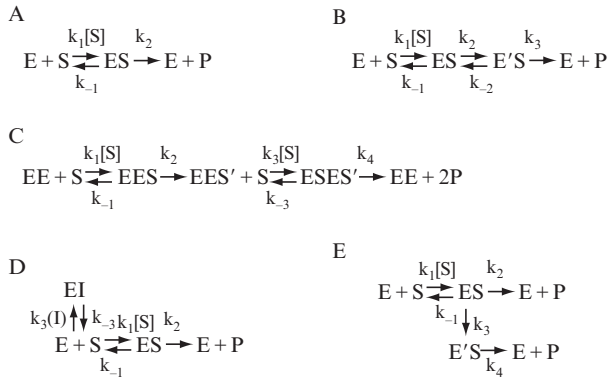

Figure 10.5 Example kinetic mechanisms with a Michaelis–Menten dependence on the substrate concentration. (A) The classic Michaelis–Menten mechanism. (B) One additional intermediate state (E′S). (C) Two substrate binding events, separated by irreversible transitions. (D) The Michaelis–Menten mechanism in the presence of a competitive inhibitor (I). (E) A parallel catalytic pathway. The values for the Michaelis–Menten parameters, k_{cat} and K_M, in terms of the pictured rate constants are listed in Table 10.1. Similarly, the values for the new parameters of n_{min}, N_L, N_S, and α, are listed in Table 10.2. The values in these tables can be calculated with a variety of techniques (Chemla *et al.*, 2008; Shaevitz *et al.*, 2005).

Table 10.1 Michaelis-Menten parameters for the example mechanisms in Fig. 10.6

Panel	k_{cat}	K_M
A	k_2	$\frac{k_2+k_{-1}}{k_1}$
B	$\frac{k_2 k_3}{k_2+k_3+k_{-2}}$	$\frac{k_2 k_3+k_3 k_{-1}+k_{-1} k_{-2}}{k_1(k_2+k_3+k_{-2})}$
C	$\frac{k_2 k_4}{k_2+k_4}$	$\frac{k_1 k_2 k_4+k_2 k_3 k_4+k_3 k_4 k_{-1}+k_1 k_2 k_{-3}}{k_1 k_2 k_3+k_1 k_3 k_4}$
D	k_2	$\frac{k_2+k_{-1}}{k_1}\left(1+\frac{k_3}{k_{-3}}[\mathrm{I}]\right)$
E	$k_4\frac{k_2+k_3}{k_3+k_4}$	$\frac{k_4}{k_1}\frac{k_2+k_3+k_{-1}}{k_3+k_4}$

Investigation of the limits listed in Table 10.2 illustrates the mechanistic constraints provided by each of the different kinetic parameters of n_{min}. Compare, for example, the value of N_S for the two-state mechanism in Fig. 10.5A with the upper limit N_S for the mechanism with one additional state in Fig. 10.5B. This parameter is strictly 1 for the example in Fig. 10.5A, while it is bounded from below by 1 and from above by 2 for the example in Fig. 10.5B. This upper bound is particularly suggestive as it is the number of states that do not involve the binding of substrate. In this sense, the value of

Table 10.2 $n_{\min}$ parameters for the example mechanisms in Fig. 10.6

Panel	N_L	N_S	α
A	1	1	$\frac{k_{-1}}{k_{-1}+k_2}$
B	1	$1 \leq \frac{(k_2+k_3+k_{-2})^2}{k_2^2+2k_2k_{-2}+(k_3+k_{-2})^2} \leq 2$	$\frac{k_{-1}(k_2+k_3+k_{-2})\,(k_2k_{-2}+(k_3+k_{-2})^2)}{(k_2k_3+k_{-1}k_3+k_{-1}k_{-2})(k_2^2+2k_2k_{-2}+(k_3+k_{-2})^2)}$
C	$1 \leq \frac{(k_1k_2(k_{-3}+k_4)+k_3k_4(k_{-1}+k_2))^2}{(k_1k_2(k_{-3}+k_4))^2+(k_3k_4(k_{-1}+k_2))^2} \leq 2$	$1 \leq \frac{(k_2+k_4)^2}{k_2^2+k_4^2} \leq 2$	$(k_{-3}k_1k_2^2 + k_{-1}k_3k_4^2)\frac{k_2+k_4}{k_2^2+k_4^2} \times \frac{k_1k_2(k_{-3}+k_4)+k_3k_4(k_{-1}+k_2)}{(k_1k_2(k_{-3}+k_4))^2+(k_3k_4(k_{-1}+k_2))^2}$
D	1	1	$\frac{k_{-1}}{k_{-1}+k_2} + \frac{k_3}{k_{-3}}[\mathrm{I}]\frac{k_2/k_{-3}}{[\mathrm{I}]k_3/k_{-3}+1}$
E	1	$0 \leq \frac{(k_3+k_4)^2}{k_3^2+k_4^2+2k_2k_3} \leq 2$	$\frac{k_{-1}k_4}{k_{-1}+k_2+k_3}\ \frac{k_3+k_4}{k_3^2+k_4^2+2k_2k_3}$

N_S provides a lower limit on the number of nonsubstrate binding states in the kinetic model, and this value can be used to limit possible models. For example, a value larger than 1 immediately rules out the simple Michaelis–Menten mechanism, Fig. 10.5A, among others.

To illustrate the constraints imposed by N_L, note that all mechanisms that have only one substrate binding state have a value of 1 for this parameter. However, this value can be larger than 1 when there are additional substrate binding states in the system (see Fig. 10.5C). In this case, the upper limit is again set by the number of such states in the mechanism; thus, the value of N_L provides a strict lower limit on the number of kinetic states that bind substrate in a given cycle. Remarkably, this statistic can indicate multiple binding events even when the substrate dependence of the mean shows no evidence for cooperativity in binding, as was recently observed for the packaging motor of the bacteriophage φ 29 (Moffitt *et al.*, 2009).

If the kinetic mechanism has no parallel catalytic pathways, then the smallest possible value of N_L and N_S is 1. However, including a parallel catalytic pathway allows this value to be less than 1 as illustrated by the example in Fig. 10.5E. Thus, a measured value of N_L or N_S less than 1 immediately implies that there are multiple catalytic pathways. Multiple catalytic pathways are one possible explanation for enzymes that display dynamics disorder (English *et al.*, 2006; Kou *et al.*, 2005; Min *et al.*, 2006), and it has been predicted that an $n_{\min}$ value less than 1 should be possible for these systems.

The constraints placed on the kinetic mechanism by the parameter α are slightly different. Note that in all of the example systems in Fig. 10.5 and Table 10.2, the value of α is proportional to the rate at which substrate unbinds from the enzyme (k_{-1} and k_{-3} for the different examples). Thus, if these rates are zero, $\alpha = 0$. The presence of a competitive inhibitor, Fig. 10.5D, provides a notable exception. In this case, even with irreversible binding $\alpha > 0$ if there is a nonzero concentration of inhibitor. These observations can be summarized in two basic requirements for $\alpha = 0$: (1) the binding of substrate molecules must be irreversible, and (2) the binding competent state cannot be in equilibrium with a nonsubstrate bound state such as an inhibitor-bound state. If these restrictions are met, then $\alpha = 0$; otherwise $\alpha > 0$.

The implications for mechanistic constraints are clear. If the measured value of α is zero, that is, the maximum value of $n_{\min}$ is the sum of the two limits $N_L + N_S$, this observation indicates that the binding state has the above properties. This remarkable result stems from the fact that lifetime statistics and visitation statistics have different effects on the statistics of the total cycle completion time, as illustrated above. A binding state not in equilibrium with any other state, that is, with only irreversible transitions out of this state, has only one visit per cycle and, thus, has no fluctuations in the state visitation number. This fact is the reason why such a state has fundamentally different fluctuations—fluctuations that are revealed by

$\alpha = 0$. It is, perhaps, surprising that such clear mechanistic features can be observed from the statistics of fluctuations in which no single binding or unbinding event has been directly observed.

While the constraints described here were developed by considering only a handful of kinetic mechanisms, these constraints have been rigorously proven for all nearest neighbor kinetic models, that is, no off-pathway states or parallel pathways (Moffitt *et al.*, 2010). More importantly, no kinetic model has yet been found which violates these constraints, though a general proof of these properties is still lacking for the arbitrary kinetic model.

5. Conclusions and Future Outlook

Enzymatic dynamics are naturally stochastic, and experimental techniques capable of revealing these natural fluctuations are becoming increasingly commonplace. Thus, it is now time to formalize methods for characterizing these fluctuations and to develop techniques for extracting the full mechanistic information from these statistics. In this chapter, we have contributed to this effort in several ways. First, we have provided an additional theoretical perspective on the relationship between the statistical fluctuations of a total enzymatic cycle and the natural fluctuations of the hidden kinetic states that compose these states. While this connection is of little immediate use to experimentalists, it should help formalize the development of more fundamental connections between the statistical properties that can be measured and the underlying kinetic mechanism of the enzyme, permitting the extraction of more subtle mechanistic details from features of fluctuations. Second, we have discussed methods for characterizing fluctuations. In particular, we have provided statistical quantities that can be calculated for enzymes with multiple measurable outcomes, such as steps of different sizes or directions. These statistics allow the statistical class of the enzymatic dynamics to be identified easily, and from this statistical class, we have demonstrated that powerful constraints can be placed on the underlying mechanism. Finally, we have discussed a novel method for classifying enzymatic fluctuations. We demonstrate that, akin to the Michaelis–Menten expression for the substrate dependence of the mean, there is a general expression for the substrate dependence of a useful measure of enzymatic fluctuations, $n_{\min}$. This expression introduces three new kinetic parameters, and by deriving the expressions for these parameters for a variety of example mechanisms, we illustrate the mechanistic constraints provided by these parameters.

While it is now clear that statistical measures of enzymatic fluctuations cannot uniquely determine a kinetic mechanism, these measures can provide powerful mechanistic constraints, constraints that are not possible from measurements of the mean alone. Given the ease with which statistical

moments are calculated and the remarkable additional information provided by the second moment, we expect that generalizations of $n_{\min}$ that include higher statistical moments such as the skew or kurtosis, or more exotic functions of the data (Zhou and Zhuang, 2007), should provide equally powerful constraints. The growing numbers of enzymes for which fluctuations are being accurately measured, not to mention all previously published examples, now await the discovery of these statistics.

Appendix

A.1. Calculating the moments of the cycle completion time

Recall that the total cycle completion time is the sum of a random number of random lifetimes for each kinetic state, Eq. (10.3). For convenience, we define the quantity T_i which is the total time spent in the kinetic state i. Thus

$$T_i = \sum_{j=0}^{n_i} t_{i,j} \tag{10.A1}$$

and the total cycle completion time is

$$\tau = \sum_{i=1}^{N} T_i. \tag{10.A2}$$

To compute the statistical properties of τ, we start by computing the statistical properties of the individual T_i. To compute the mean, we use the tower property of averages. This property states that the mean of a two variable distribution is the average of the mean of the distribution evaluated with a constant value for one parameter averaged again over that parameter. Evaluating the average of $\langle T_i \rangle$ with $n_i = M$, yields:

$$\begin{aligned} \langle\langle T_i(n_i = M)\rangle\rangle &= \left\langle \left\langle \sum_{j=1}^{M} t_{i,j} \right\rangle \right\rangle = \left\langle \sum_{j=1}^{M} \langle t_{i,j} \rangle \right\rangle \\ &= \left\langle \sum_{j=1}^{M} \langle t_i \rangle \right\rangle = \langle M \langle t_i \rangle \rangle = \langle n_i \rangle \langle t_i \rangle. \end{aligned} \tag{10.A3}$$

Since the average of a sum of random variables is simply the sum of the averages, this implies that

$$\langle \tau \rangle = \sum_{i=1}^{N} \langle T_i \rangle, \tag{10.A4}$$

which is the result stated in the main text.

The variance of a sum of random variables is just the sum of the variance of the individual variables and the covariance of all pairs of variables; thus

$$\mathrm{var}(\tau) = \sum_{i=1}^{N} \mathrm{var}(T_i) + 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathrm{cov}(T_i, T_j). \tag{10.A5}$$

To compute the variance of the individual T_i, we again use the tower property of averages to calculate the second moment:

$$\begin{aligned} \left\langle \left\langle T_i^2 (n_i = M) \right\rangle \right\rangle = \left\langle \left\langle \sum_{j=1}^{M} t_{i,j} \sum_{k=1}^{M} t_{i,k} \right\rangle \right\rangle = \left\langle \sum_{j=1}^{M} \left\langle t_{i,j}^2 \right\rangle \right\rangle \\ + 2 \left\langle \sum_{j=1}^{M} \sum_{k=j+1}^{M} \left\langle t_{i,j} t_{i,k} \right\rangle \right\rangle, \end{aligned} \tag{10.A6}$$

where we have divided the terms into squared terms and cross terms. Since the individual lifetimes are independent and identically distributed, $\langle t_{i,j} t_{i,k} \rangle = \langle t_{i,j} \rangle \langle t_{i,k} \rangle = \langle t_i \rangle^2$. Counting terms in Eq. (10.A6) yields

$$\left\langle T_i^2 (n_i = M) \right\rangle = M \left\langle t_i^2 \right\rangle + M(M-1) \langle t_i \rangle^2 \tag{10.A7}$$

and averaging over M yields the second moment:

$$\left\langle T_i^2 \right\rangle = \langle n_i \rangle \left\langle t_i^2 \right\rangle + \left(\left\langle n_i^2 \right\rangle - \langle n_i \rangle \right) \langle t_i \rangle^2. \tag{10.A8}$$

Subtracting the mean squared from above yields the variance:

$$\mathrm{var}(T_i) = \left\langle T_i^2 \right\rangle - \langle T_i \rangle^2 = \mathrm{var}(n_i) \langle t_i \rangle^2 + \mathrm{var}(t_i) \langle n_i \rangle. \tag{10.A9}$$

The covariance of the different T_i is defined as

$$\mathrm{cov}(T_i, T_j) = \left\langle T_i T_j \right\rangle - \langle T_i \rangle. \tag{10.A10}$$

The final terms have already been calculated, so we need only the first term T_j.

Again, we calculate this by exploiting the tower property, evaluating the individual $T_{i,j}$ at $n_i = M_i$ and $n_j = M_j$. Thus

$$\left\langle T_i(n_i = M_i)\, T_j(n_j = M_j)\right\rangle = \left\langle \sum_{k=1}^{M_i} t_{i,k} \sum_{l=1}^{M_j} t_{j,l} \right\rangle = M_i M_j \langle t_i\rangle \left\langle t_j\right\rangle. \tag{10.A11}$$

Averaging over M_i and M_j yields

$$\left\langle T_i T_j\right\rangle = \left\langle n_i n_j\right\rangle \langle t_i\rangle \left\langle t_j\right\rangle. \tag{10.A12}$$

Combining this expression with the averages computed above yields the covariance:

$$\operatorname{cov}(T_i, T_j) = \left\langle n_i n_j\right\rangle \langle t_i\rangle \left\langle t_j\right\rangle - \langle n_i\rangle \left\langle n_j\right\rangle \langle t_i\rangle \left\langle t_j\right\rangle = \operatorname{cov}(n_i n_j) \langle t_i\rangle \left\langle t_j\right\rangle. \tag{10.A13}$$

Combining this result with the variances derived above yields the result in the main text.

A.2. Estimation of systematic errors in statistical moments

Nearly all experimental methods for detecting enzymatic cycle completion times or dwell times of molecular motors have a dead time—a minimum dwell time required for detection of a given event. In the ideal situation, this dead time is much smaller than the typical cycle completion time, and the effects of the dead-time can be ignored. However, in situations in which this is not the case, it would be useful to have quantitative measures of the bias introduced into moments and other statistical properties by the dead-time. Here we provide such an estimate.

If the measurement technique has a finite dead-time t_0, then the cycle completion times that are measured will not follow the actual dwell time distribution, $\varphi(t)$. Rather, these times will be distributed via the modified distribution $\varphi'(t)$:

$$\varphi'(t) = \begin{cases} 0 & t < t_0, \\ \alpha\varphi(t) & t \geq t_0. \end{cases} \tag{10.A14}$$

Below the dead-time, there is no probability of observing a cycle completion time; thus, the distribution is zero. Above the dead-time, we assume that we can measure all cycle completion times with equal fidelity; thus, the

dwell time distribution is the original distribution. An additional factor of α must be included to properly normalize this distribution. A similar technique has been used to produce unbiased estimators of DNA shortening events in single-molecule measurements (Koster *et al.*, 2006).

The measured moments are determined from this distribution via

$$\langle t' \rangle = \int_0^\infty t\varphi'(t)\mathrm{d}t \quad \text{and} \quad \langle t'^2 \rangle = \int_0^\infty t^2\varphi'(t)\mathrm{d}t. \tag{10.A15}$$

Thus, the systematic errors arise from both the region of the distribution that is not measured, $t < t_0$, and the fact that not measuring the probability in this region improperly weights the importance of the portion that is measured, that is, $\alpha \neq 1$.

To estimate the systematic errors, we will assume a shape for the distribution for $t < t_0$. As long as the final estimates for the systematic errors are small, then the specific function assumed will not be important. For simplicity, and to allow integrals to be computed directly, we assume that the distribution is well approximated by the gamma distribution during the dead-time:

$$\varphi(t) = \frac{k^n t^{n-1}}{\Gamma(n)} \mathrm{e}^{-kt}, \tag{10.A16}$$

where $\Gamma(n)$ is the gamma function. Moreover, since this distribution is only used for a small region of time, its parameters can be estimated from the measured parameters, again, without introducing significant errors into the final estimates of the systematic errors. Namely, the parameters of the gamma distribution are estimated via

$$n \approx \frac{\langle t' \rangle^2}{\langle t'^2 \rangle - \langle t' \rangle^2} \quad \text{and} \quad k \approx \frac{n}{\langle t' \rangle}. \tag{10.A17}$$

The systematic errors in the moments can now be related to integrals of this assumed distribution over the dead-time of the measurement. The error in the mean is

$$\begin{aligned} \langle t \rangle - \langle t' \rangle &= \int_0^\infty t\varphi(t)\mathrm{d}t - \int_0^\infty t\varphi'(t)\mathrm{d}t = \int_0^{t_0} t\varphi(t)\mathrm{d}t \\ &\quad + (1-\alpha)\int_{t_0}^\infty t\varphi'(t)\mathrm{d}t = \int_0^{t_0} t\varphi(t)\mathrm{d}t + (1-\alpha)\langle t' \rangle. \end{aligned} \tag{10.A18}$$

In the first line, we took advantage of the fact that the first integral, over $\varphi(t)$, can be split into the region below and above the dead-time, and the region above the dead time can be combined with the same region of the integral over $\varphi'(t)$ as long as the normalization factor α is considered. However, this final integral is by definition the measured moment, $\langle t' \rangle$. Thus, the final expression involves only integrals over the small region of the dead time—the first term and the normalization constant α—and parameters that have already been measured, $\langle t' \rangle$. Since we have assumed a gamma distribution for the distribution during the dead time, the first integral can be performed:

$$\int_0^{t_0} t\varphi(t)\mathrm{d}t = \frac{\Gamma(n+1) - \Gamma(n+1, kt_0)}{k\Gamma(n)}. \tag{10.A19}$$

Similarly the normalization constant, α, can be determined analytically

$$\alpha = \left(\int_{t_0}^{\infty} \varphi(t)\mathrm{d}t\right)^{-1} = \left(1 - \int_0^{t_0} \varphi(t)\mathrm{d}t\right)^{-1} = \frac{\Gamma(n)}{\Gamma(n, kt_0)}. \tag{10.A20}$$

Here $\Gamma(n, kt_0)$ is the upper incomplete gamma function. Thus, the final estimate for the systematic error in the mean introduced by a dead-time in the measurement is

$$\langle t \rangle - \langle t' \rangle = \frac{\Gamma(n+1) - \Gamma(n+1, kt_0)}{k\Gamma(n)} + \left(1 - \frac{\Gamma(n)}{\Gamma(n, kt_0)}\right)\langle t' \rangle. \tag{10.A21}$$

Similar arguments can be applied to determine the error in the second moment. Namely

$$\begin{aligned} \langle t^2 \rangle - \langle t'^2 \rangle &= \int_0^{\infty} t^2\varphi(t)\mathrm{d}t - \int_0^{\infty} t^2\varphi'(t)\mathrm{d}t \\ &= \int_0^{t_0} t^2\varphi(t)\mathrm{d}t + (1-\alpha)\int_{t_0}^{\infty} t^2\varphi'(t)\mathrm{d}t \\ &= \int_0^{t_0} t^2\varphi(t)\mathrm{d}t + (1-\alpha)\langle t'^2 \rangle, \end{aligned} \tag{10.A22}$$

where the integral over the dead-time can again be evaluated analytically:

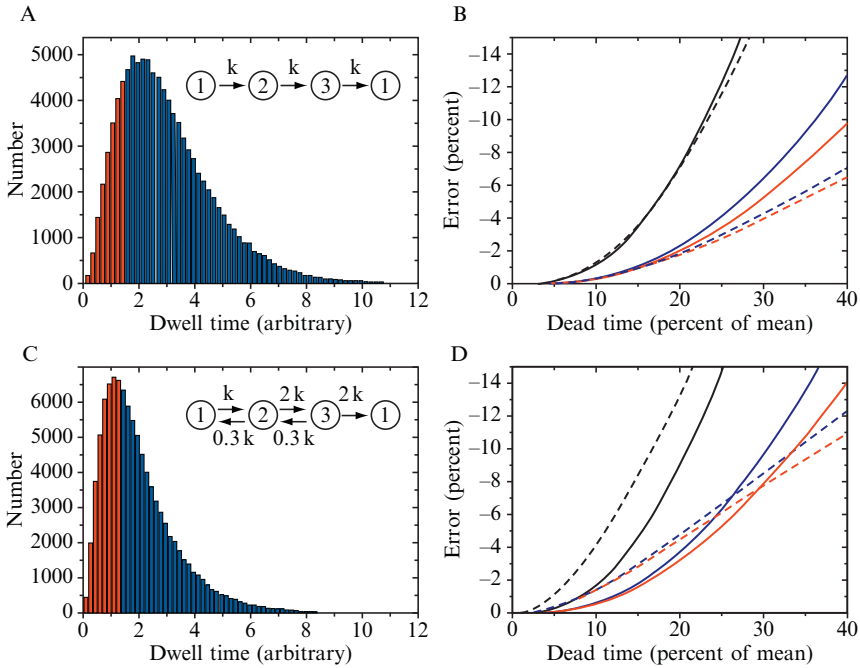

Figure 10.6 Estimating systematic errors from dead-times. (A) Histogram of simulated dwell times for the kinetic mechanism pictured: A three-state irreversible mechanism. The time axis is arbitrary units. Red corresponds to the portion of the distribution which would not be observed if there was a dead-time of 40% of the mean dwell time. (B) Systematic error in the mean (red), second moment (blue), and $n_{\min}$ (black) as a function of dead-time measured in percentage of the mean dwell time. The solid lines correspond to the actual error introduced by the dead-time while the dashed lines correspond to the estimates using the method described here. (C) Histogram of simulated dwell times for the kinetic mechanism pictured: A three-state system with reversible transitions. (B) Systematic errors for this mechanism as a function of dead-time as in panel (B). Data in panels (B) and (D) were calculated from stochastic simulations with 10^6 dwell times.

$$\int_0^{t_0} t^2 \varphi(t) \mathrm{d}t = \frac{\Gamma(n+2) - \Gamma(n+2, kt_0)}{k^2 \Gamma(n)}. \tag{10.A23}$$

Thus, the systematic error made in the measurement of the second moment is

$$\left\langle t^2 \right\rangle - \left\langle t'^2 \right\rangle = \frac{\Gamma(n+2) - \Gamma(n+2, kt_0)}{k^2 \Gamma(n)} + \left(1 - \frac{\Gamma(n)}{\Gamma(n, kt_0)}\right) \left\langle t'^2 \right\rangle. \tag{10.A24}$$

The systematic error in functions of the moments, such as the variance or $n_{\min}$, can be calculated by first correcting the measured moments with the estimated systematic errors and then calculating the function of the moments. The systematic error in this function can then be computed by comparing the new estimate to the function evaluated with the measured values. Figure 10.6 shows that these expressions provide reasonable estimates of the systematic errors due to a dead-time as long as the dead-time is relatively small fraction of the mean duration. All statistical properties have errors less than $\sim$10% when the dead-time is less than $\sim$20% of the mean.

In general, it is unlikely that the method used to detect cycle completion events will have a sharp cut-off below which events are never detected and above which events are always detected. If it is necessary to work in conditions in which the dead-time is relatively large, then it may be necessary to include a more representative description of the ability to detect events of different durations. In this case, the general approach described here can be used to estimate systematic errors, and perhaps correct measurements. The discrete distribution in Eq. (10.A14) simply needs to be replaced with the measured response of the detection technique, perhaps estimated from processing of simulated data. The necessary integrals in the subsequent steps can then be performed numerically.

REFERENCES

Abel, N. H. (1826). Beweis der Unmoglichkeit, algebraische Gleichungen von hoheren Graden als dem vierten allgemein aufzulosen. *J. fur die reine und Angewandte Math. (Crelle's Journal)* **1,** 65–84.

Aldous, D., and Shepp, L. (1987). The least variable phase type distribution is erlang. *Stochastic Models* **3,** 467–473.

Cappello, G., *et al.* (2007). Myosin V stepping mechanism. *Proc. Natl. Acad. Sci. USA* **104,** 15328–15333.

Carter, N. J., and Cross, R. A. (2005). Mechanics of the kinesin step. *Nature* **435,** 308.

Cecconi, C., *et al.* (2005). Direct observation of the three-state folding of a single protein molecule. *Science* **309,** 2057–2060.

Charvin, G., *et al.* (2002). On the relation between noise spectra and the distribution of time between steps for single molecular motors. *Single Molecules* **3,** 43–48.

Chemla, Y. R., *et al.* (2008). Exact solutions for kinetic models of macromolecular dynamics. *J. Phys. Chem. B.* **112,** 6025–6044.

Clemen, A. E., *et al.* (2005). Force-dependent stepping kinetics of myosin-V. *Biophys. J.* **88,** 4402–4410.

Cornish, P. V., and Ha, T. (2007). A survey of single-molecule techniques in chemical biology. *ACS Chem. Biol.* **2,** 53–61.

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

Efron, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika* **68,** 589–599.

Efron, B., and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* **1,** 54–77.

English, B. P., *et al.* (2006). Ever-fluctuating single enzyme molecules: Michaelis–Menten equation revisited (vol. 2, p. 87, 2006). *Nat. Chem. Biol.* **2,** 168.
Fisher, M. E., and Kolomeisky, A. B. (1999). Molecular motors and the forces they exert. *Physica a—Stat. Mech. Appl.* **274,** 241–266.
Garai, A., *et al.* (2009). Stochastic Kinetics of Ribosomes: Single Motor Properties and Collective Behavior. Arxiv preprint arXiv:0903.2608.
Gennerich, A., and Vale, R. D. (2009). Walking the walk: How kinesin and dynein coordinate their steps. *Curr. Opin. Cell Biol.* **21,** 59–67.
Gennerich, A., *et al.* (2007). Force-induced bidirectional stepping of cytoplasmic dynein. *Cell* **131,** 952–965.
Goedecke, D. M., and Elston, T. C. (2005). A model for the oscillatory motion of single dynein molecules. *J. Theor. Biol.* **232,** 27–39.
Greenleaf, W. J., *et al.* (2007). High-resolution, single-molecule measurements of biomolecular motion. *Annu. Rev. Biophys. Biomol. Struct.* **36,** 171–190.
Henzler-Wildman, K., and Kern, D. (2007). Dynamic personalities of proteins. *Nature* **450,** 964–972.
Karplus, M., and McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nat. Struct. Mol. Biol.* **9,** 646–652.
Kohler, D., *et al.* (2003). Different degrees of lever arm rotation control myosin step size. *J. Cell Biol.* **161,** 237–241.
Kolomeisky, A. B., and Fisher, M. E. (2003). A simple kinetic model describes the processivity of myosin-v. *Biophys. J.* **84,** 1642–1650.
Kolomeisky, A. B., and Fisher, M. E. (2007). Molecular motors: A theorist's perspective. *Annu. Rev. Phys. Chem.* **58,** 675–695.
Koster, D. A., *et al.* (2006). Multiple events on single molecules: Unbiased estimation in single-molecule biophysics. *Proc. Natl. Acad. Sci. USA* **103,** 1750–1755.
Kou, S. C., *et al.* (2005). Single-molecule Michaelis–Menten equations. *J. Phys. Chem. B* **109,** 19068–19081.
Koza, Z. (1999). General technique of calculating the drift velocity and diffusion coefficient in arbitrary periodic systems. *J. Phys. A Math. Gen.* **32,** 7637–7651.
Koza, Z. (2000). Diffusion coefficient and drift velocity in periodic media. *Physica A* **285,** 176–186.
Li, P. T. X., *et al.* (2008). How RNA unfolds and refolds. *Annu. Rev. Biochem.* **77,** 77–100.
Liao, J.-C., *et al.* (2007). Extending the absorbing boundary method to fit dwell-time distributions of molecular motors with complex kinetic pathways. *Proc. Natl. Acad. Sci. USA* **104,** 3171–3176.
Linden, M., and Wallin, M. (2007). Dwell time symmetry in random walks and molecular motors. *Biophys. J.* **92,** 3804–3816.
Mallik, R., *et al.* (2004). Cytoplasmic dynein functions as a gear in response to load. *Nature* **427,** 649–652.
Michaelis, L., and Menten, M. L. (1913). The kinetics of the inversion effect. *Biochem. Z.* **49,** 333–369.
Min, W., *et al.* (2006). When does the Michaelis–Menten equation hold for fluctuating enzymes? *J. Phys. Chem. B* **110,** 20093–20097.
Moffitt, J. R., *et al.* (2008). Recent advances in optical tweezers. *Annu. Rev. Biochem.* **77,** 205–228.
Moffitt, J. R., *et al.* (2009). Intersubunit coordination in a homomeric ring ATPase. *Nature* **457,** 446–450.
Moffitt, J. R., *et al.* (2010). Mechanistic constraints from the substrate concentration dependence of enzymatic fluctuations. (submitted).
Neuman, K. C., *et al.* (2005). Statistical determination of the step size of molecular motors. *J. Phys.: Condens. Matter* **17,** S3811.

Neuts, M. F. (1975). Probability distributions of phase type. Liber Amicorum Prof. Emeritus H. Florin, Leuven, pp. 173–206.
Neuts, M. F. (1994). Matrix-geometric solutions in stochastic models: An algorithmic approach. Dover Publications, New York.
O'Cinneide, C. A. (1990). Characterization of phase-type distributions. *Stochastic Models* **6,** 1–57.
O'Cinneide, C. A. (1999). Phase-type distributions: Open problems and a few properties. *Commun. Stat.—Stochastic Models* **15,** 731–758.
Qian, H. (2008). Cooperativity and specificity in enzyme kinetics: A single-molecule time-based perspective. *Biophys. J.* **95,** 10–17.
Reck-Peterson, S. L., *et al.* (2006). Single-molecule analysis of dynein processivity and stepping behavior. *Cell* **126,** 335–348.
Rief, M., *et al.* (2000). Myosin-V stepping kinetics: A molecular model for processivity. *Proc. Natl. Acad. Sci. USA* **97,** 9482–9486.
Rock, R. S., *et al.* (2001). Myosin VI is a processive motor with a large step size. *Proc. Natl. Acad. Sci. USA* **98,** 13655.
Sakmann, B., and Neher, E. (1984). Patch clamp techniques for studying ionic channels in excitable membranes. *Annu. Rev. Physiol.* **46,** 455–472.
Schnitzer, M. J., and Block, S. M. (1995). Statistical kinetics of processive enzymes. *Cold Spring Harb. Symp. Quant. Biol.* **60,** 793–802.
Schnitzer, M. J., and Block, S. M. (1997). Kinesin hydrolyses one ATP per 8-nm step. *Nature* **388,** 386–390.
Segel, I. H. (1975). Enzyme Kinetics. John Wiley & Sons Inc., New Jersey.
Sellers, J. R., and Veigel, C. (2006). Walking with myosin V. *Curr. Opin. Cell Biol.* **18,** 68–73.
Shaevitz, J. W., *et al.* (2005). Statistical kinetics of macromolecular dynamics. *Biophys. J.* **89,** 2277–2285.
Svoboda, K., *et al.* (1994). Fluctuation analysis of motor protein movement and single enzyme kinetics. *PNAS* **91,** 11782–11786.
Tinoco, I., and Wen, J. D. (2009). Simulation and analysis of single-ribosome translation. *Phys. Biol.* **6,** 025006.
Tsygankov, D., *et al.* (2007). Back-stepping, hidden substeps, and conditional dwell times in molecular motors. *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.* **75,** 021909.
Wang, H. (2007). A new derivation of the randomness parameter. *J. Math. Phys.* **48,** 103301.
Woodside, M. T., *et al.* (2008). Folding and unfolding single RNA molecules under tension. *Curr. Opin. Chem. Biol.* **12,** 640–646.
Xie, S. N. (2001). Single-molecule approach to enzymology. *Single Molecules* **2,** 229–236.
Xing, J., *et al.* (2005). From continuum Fokker-Planck models to discrete kinetic models. *Biophys. J.* **89,** 1551–1563.
Xu, W., *et al.* (2009). Single-molecule kinetic theory of heterogeneous and enzyme catalysis. *J. Phys. Chem. C* **113,** 2393.
Yamaoka, K., *et al.* (1978). Application of Akaike's information criterion (AIC) in the evaluation of linear pharmacokinetic equations. *J. Pharmacokinet Pharmacodyn* **6,** 165–175.
Yildiz, A., *et al.* (2004). Myosin VI steps via a hand-over-hand mechanism with its lever arm undergoing fluctuations when attached to actin. *J. Biol. Chem.* **279,** 37223.
Yildiz, A., *et al.* (2008). Intramolecular strain coordinates kinesin stepping behavior along microtubules. *Cell* **134,** 1030–1041.
Zhou, Y., and Zhuang, X. (2006). Robust reconstruction of the rate constant distribution using the phase function method. *Biophys. J.* **91,** 4045–4053.
Zhou, Y., and Zhuang, X. (2007). Kinetic analysis of sequential multistep reactions. *J. Phys. Chem. B* **111,** 13600–13610.