Specificity and completion time distributions of biochemical processes

Brian Munsky,[1] Ilya Nemenman,[1,2] and Golan Bel[1,a)]
[1]*Center for Nonlinear Studies and Computer, Computation and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[2]*Department of Physics, Department of Biology, and Computational and Life Sciences Strategic Initiative, Emory University, Atlanta, Georgia 30322, USA*

(Received 14 September 2009; accepted 24 November 2009; published online 17 December 2009)

In order to produce specific complex structures from a large set of similar biochemical building blocks, many biochemical systems require high sensitivity to small molecular differences. The first and most common model used to explain this high specificity is kinetic proofreading, which has been extended to a variety of systems from detection of DNA mismatch to cell signaling processes. While the specification properties of kinetic proofreading models are well known and were studied in various contexts, very little is known about their temporal behavior. In this work, we study the dynamical properties of discrete stochastic two-branch kinetic proofreading schemes. Using the Laplace transform of the corresponding chemical master equation, we obtain an analytical solution for the completion time distribution. In particular we provide expressions for the specificity as well as the mean and variance of the process completion times. We also show that, for a wide range of parameters, a process distinguishing between two different products can be reduced to a much simpler three-point process. Our results allow for the systematic study of the interplay between specificity and completion times, as well as testing the validity of the kinetic proofreading model in biological systems. © *2009 American Institute of Physics*. [doi:10.1063/1.3274803]

I. INTRODUCTION

The strong bias toward the correct assembly of particular molecular constructs, or specificity, plays a key role in myriad biochemical processes such as DNA assembly, cell signaling, protein folding, and others. A common model accounting for the almost error-free completion of these processes is kinetic proofreading (KPR), which was first suggested to explain the high specificity of protein synthesis.[1] Similar motifs are common in various biological processes where multiple error-prone steps generate error-free results. For example, KPR schemes are common in modeling of DNA synthesis, repair, and replication.[2–4] Similar proofreading ideas appear in other contexts such as protein translation,[1,5] molecular transport,[6] receptor-initiated signaling,[7–12] RNA transcription,[13] and other processes.

Various aspects of the KPR concept have already been studied. Hopfield[1] and Ninio[14] demonstrated the possible increases in specificity due to single-step proofreading. Later explorations of similar proofreading models considered the multistep proofreading process as a "black box" and studied the accuracy achieved by such processes,[15] as well as the energy cost and optimal distribution of the proofreading effort along the proofreading chain.[16] In Ref. 7 the KPR was proposed as a model for the T-cell receptor explaining the high discrimination between foreign antigen and self antigen with only moderately lower affinity. In this context the specificity of a multistep process was studied again, as well as the time delay between initial binding and output signal.

In addition to process specificity, the time required to reach this specificity also plays an important role in biochemical processes. A proofreading strategy must be efficient as well as specific. In different contexts[17–23] it was shown that such completion or first passage times provide a wealth of information about the underlying systems. Extending these results to KPR, suggests that the characterization of the completion time distribution may help researchers to distinguish between different kinetic models and even support or oppose the existence of KPR in specific systems. Surprisingly, the completion time distributions of KPR schemes have not been calculated before.

In this article, we investigate the temporal behavior of different KPR schemes. We derive the chemical master equation [CME (Ref. 24)] and its transform into the Laplace domain, which provides analytical expressions for the directional and nondirectional completion time distribution. In particular, the zeroth, first, and second derivatives of the CMEs Laplace transform provide expressions for the specificity, mean and coefficient of variation of the completion times. In turn, these expressions provide a starting point to examine the tradeoffs between the stationary and temporal behaviors of different KPR schemes. Furthermore, we show that over a wide range of kinetic parameters the complex proofreading process reduces to a three-state process with simple distributions of the transition time between the three states. We also provide a diagram mapping the parameters space into classes of different behavior of the completion time distribution.

This paper is organized as follows. In Sec. II, we introduce the model and provide its CME, as well as the analyti-

a)Electronic mail: golanbel@gmail.com.

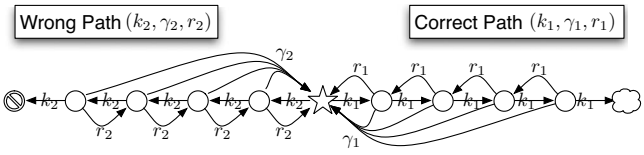

FIG. 1. Schematic description of the two-branch general KPR scheme for error correction. The process begins at the point denoted with a star. From there, it can move one step to the right or left with rates k_1 or k_2, respectively. On the right half of the chain, the process can proceed one step forward with rate k_1, one step backward with rate r_1, or all the way to the origin with rate γ_1. On the left half of the chain, these rates are replaced with k_2, r_2 and γ_2. The leftmost and rightmost sites are absorbing sites: once the process reaches these points, the process is completed. If the process finishes at the rightmost site, it is said to have completed correctly, if it finishes at the leftmost site, the process has completed incorrectly.

cal solution of the CME in the Laplace domain. In Sec. III we show the different behaviors of the completion time distributions and divide the parameters space into regimes corresponding to different typical distributions. We also show the coefficient of variation versus the parameters of the problem and discuss its meaning. In Sec. IV, we summarize our results and their relevance to many of the problems previously studied in the context of KPR.

II. THE MODEL

Here we consider the general model of KPR, which can be represented by the Markov chain in Fig. 1. The initiation state is represented by the star in the center of the chain, and is denoted by $(i,j)=(0,0)$. Depending upon the system, the state $(i,j)=(0,0)$ may have different meanings. In protein assembly as modeled in Ref. 1, this state may correspond to an empty A-site of the mRNA-ribosome complex or other substeps in more realistic models, or in cell signaling the initiation state may correspond to a receptor with no bound ligand.[7] The state just to the right of the star, labeled by $(i,j)=(1,0)$ corresponds to a single step in the correct direction, i.e., the intended tRNA binds to the A-site or the proper ligand binds to the receptor. Conversely, a step to the left is in the wrong direction (wrong tRNA or wrong ligand). In general there may be many wrong directions or additional sub-chains branching from the central initiation point, but for simplicity we consider only the case where there is only one right and one wrong decision. The Markov system can transition one step forward from the initiation point with rate k_1 toward correct completion or with rate k_2 toward incorrect completion. The process may also move one step back away from completion with rate r_1 or r_2, or back to the origin with rate γ_1 or γ_2. The two branches of the chain have L_1 or L_2 nodes correspondingly, the last of which, $(L_1,0)$ or $(0,L_2)$ is an absorbing point (representing the formation of the correct/incorrect product). The CME describing the dynamics of the occupation probabilities is

$$\frac{dp_{0,j}(t)}{dt}=\begin{cases}k_2p_{0,L_2-1}(t) & \text{for } j=L_2\\ -(k_2+\gamma_2+r_2)p_{0,L_2-1}(t)+k_2p_{0,L_2-2}(t) & \text{for } j=L_2-1\\ -(k_2+\gamma_2+r_2)p_{0,j}(t)+k_2p_{0,j-1}(t)+r_2p_{0,j+1}(t) & \text{for } 0<j<L_2-1,\end{cases} \tag{1a}$$

$$\frac{dp_{i,0}(t)}{dt}=\begin{cases}k_1p_{L_1-1,0}(t) & \text{for } i=L_1\\ -(k_1+\gamma_1+r_1)p_{L_1-1,0}(t)+k_1p_{L_1-2,0}(t) & \text{for } i=L_1-1\\ -(k_1+\gamma_1+r_1)p_{i,0}(t)+k_1p_{i-1,0}(t)+r_1p_{i+1,0}(t) & \text{for } 0<i<L_1-1,\end{cases} \tag{1b}$$

and for the initiation point $(i,j)=(0,0)$

$$\frac{dp_{0,0}(t)}{dt}=-(k_1+k_2)p_{0,0}(t)+r_1p_{1,0}(t)+r_2p_{0,1}(t)+\gamma_1\sum_{i=1}^{L_1-1}p_{i,0}(t)+\gamma_2\sum_{j=1}^{L_2-1}p_{0,j}(t). \tag{1c}$$

For any given specific case, this CME may be solved using various methods, such as various projection approaches,[25–29] using stochastic field theory approaches[30,31] or simulated using stochastic simulations.[32–34] Similarly, completions times for a given process could be calculated directly from the CME using projection approaches[35] or analyzed using transition path and transition interface sampling.[36–40] However, in this work we take an analytical approach in an effort to attain explicit expressions for the temporal behavior of the process in terms of the kinetic parameters. Later in Sec. III, these expressions will be used to study the dependence of the specificity and completion time distributions on the system's parameters as the number of intermediate steps, and forward/backward/proofreading rates.

In order to derive explicit expressions for the temporal KPR behavior, we utilize a Laplace transform approach that is similar to approaches previously used to study first passage time distributions for a ladder process[41] and a single-branch KPR process.[42] More specifically, we first simplify the set of differential equation describing the dynamics of the occupation probabilities, by applying the Laplace transform

$$P_{i,j}(s) \equiv \int_0^\infty p_{i,j}(t)e^{-st}dt, \tag{2}$$

where we are using lowercase variables to represent quantities in the time domain and uppercase variables to represent the corresponding quantities in the Laplace domain. Upon application of the Laplace transform, the probabilities are now described by the following algebraic master equation:

$$P_{0,j}(s) = \begin{cases} \dfrac{k_2}{s}P_{0,L_2-1}(s) & \text{for } j = L_2 \\ \dfrac{k_2}{s+k_2+\gamma_2+r_2}P_{0,L_2-2}(s) & \text{for } j = L_2-1 \\ \dfrac{1}{s+k_2+\gamma_2+r_2}(k_2P_{0,j-1}(s)+r_2P_{0,j+1}(s)) & \text{for } 0<j<L_2-1, \end{cases} \tag{3a}$$

$$P_{i,0}(s) = \begin{cases} \dfrac{k_1}{s}P_{L_1-1,0}(s) & \text{for } i = L_1 \\ \dfrac{k_1}{s+k_1+\gamma_1+r_1}P_{L_1-2,0}(s) & \text{for } i = L_1-1 \\ \dfrac{1}{s+k_1+\gamma_1+r_1}(k_1P_{i-1,0}(s)+r_1P_{i+1,0}(s)) & \text{for } 0<i<L_1-1, \end{cases} \tag{3b}$$

and

$$P_{0,0}(s) = \frac{1}{s+k_1+k_2}\left(1+r_1P_{1,0}(s)+r_2P_{0,1}(s)+\gamma_1\sum_{i=1}^{L_1-1}P_{i,0}(s)+\gamma_2\sum_{j=1}^{L_2-1}P_{0,j}(s)\right). \tag{3c}$$

For the above equation we have already imposed the initial condition $p_{i,j}(t=0)=\delta_{i,0}\delta_{j,0}$, where δ is the Kronecker delta. In other words, $p_{0,0}(0)=1$ and $p_{i,j}(0)=0$ for all $(i,j)\neq(0,0)$. The general solution of these equations is explicitly written as

$$P_{i,j}(s) = \begin{cases} A\lambda_1^i + B\lambda_2^i & \text{for } j=0, i\geq 0 \\ A\beta_2^j + B\beta_2^j + C(\beta_1^j-\beta_2^j) & \text{for } i=0, j>0. \end{cases} \tag{4}$$

Here, the space independent parameters $\lambda_{1,2}(s)$ and $\beta_{1,2}(s)$ are obtained from the solution of the quadratic equations

$$\begin{aligned} &\frac{k_1}{s+k_1+\gamma_1+r_1}+\frac{r_1}{s+k_1+\gamma_1+r_1}\lambda^2-\lambda=0, \\ &\frac{k_2}{s+k_2+\gamma_2+r_2}+\frac{r_2}{s+k_2+\gamma_2+r_2}\beta^2-\beta=0, \end{aligned} \tag{5}$$

which come from the expressions for $P_{i,j}(s)$ at the interior points of the two branches. The boundary conditions are satisfied by proper choice of the coefficients $A(s)$, $B(s)$, and $C(s)$. The boundary condition at $(i,j)=(0,0)$ [see Eq. (3c)] is expressed as

$$\begin{aligned} &(s+k_1+k_2)(A+B)-\gamma_2(A+B-C)\sum_{j=1}^{L_2-1}\beta_2^j \\ &\quad = 1+r_1(A\lambda_1+B\lambda_2)+r_2((A+B-C)\beta_2+C\beta_1) \\ &\qquad +\gamma_1\sum_{i=1}^{L_1-1}(A\lambda_1^i+B\lambda_2^i)+\gamma_2C\sum_{j=1}^{L_2-1}\beta_1^j. \end{aligned} \tag{6}$$

The boundary condition at $(i,j)=(L_1-1,0)$ is written as [see Eq. (3b)]

$$A\lambda_1^{L_1-1}+B\lambda_2^{L_1-1}=\frac{k_1}{s+k_1+\gamma_1+r_1}(A\lambda_1^{L_1-2}+B\lambda_2^{L_1-2}), \tag{7}$$

and the boundary condition at $(0,L_2-1)$ is [see Eq. (3a)]

$$(A+B-C)\beta_2^{L_2-1}=\frac{k_2((A+B-C)\beta_2^{L_2-2}+C\beta_1^{L_2-2})}{s+k_2+\gamma_2+r_2}-C\beta_1^{L_2-1}. \tag{8}$$

Using the definitions of $\lambda_{1,2}$ [see Eq. (5)], we can rewrite Eq. (7) as

$$B=-A\frac{\lambda_1^{L_1}}{\lambda_2^{L_1}}. \tag{9}$$

Similarly using the definitions of $\beta_{1,2}$, we rewrite Eq. (8) as

$$C = A\frac{\beta_2^{L_2}(\lambda_2^{L_1} - \lambda_1^{L_1})}{\lambda_2^{L_1}(\beta_2^{L_2} - \beta_1^{L_2})}. \tag{10}$$

Finally, using Eqs. (9) and (10), one can simplify Eq. (6)

$$\frac{1}{A} = \left(1 - \frac{\lambda_1^{L_1}}{\lambda_2^{L_1}}\right)\Bigg(\gamma_2 + k_1 + k_2 + s + \gamma_1 - \gamma_2\frac{\frac{1-\beta_2^{L_2}}{1-\beta_2}\beta_1^{L_2} + \frac{1-\beta_1^{L_2}}{1-\beta_1}\beta_2^{L_2}}{\beta_2^{L_2} - \beta_1^{L_2}} - r_2\frac{\beta_2\beta_1^{L_2} + \beta_1\beta_2^{L_2}}{\beta_2^{L_2} - \beta_1^{L_2}}\Bigg) - r_1\lambda_1\left(1 - \frac{\lambda_1^{L_1-1}}{\lambda_2^{L_1-1}}\right) - \gamma_1\left(\frac{1-\lambda_1^{L_1}}{1-\lambda_1} - \frac{\lambda_1^{L_1}}{\lambda_2^{L_1}}\frac{1-\lambda_2^{L_1}}{1-\lambda_2}\right). \tag{11}$$

Note that in deriving Eqs. (9)–(11), we assumed that the parameters $k_1, k_2, r_1, r_2, \gamma_1, \gamma_2$ are all different from zero.

In order to study the temporal behavior of the KPR model, we compute (i) the probability that the system will reach the correct terminus point and (ii) the distribution of time until the system reaches one of the two possible terminus points. Both of these quantities are found by examining the un-normalized probability density functions (PDFs) for the first passage time to the absorbing sites $(L_1, 0)$ or $(0, L_2)$, which are given by

$$f_1(t) = k_1 p_{L_1-1,0}(t), \qquad f_2(t) = k_2 p_{0,L_2-1}(t). \tag{12}$$

According to Eqs. (12) and (4), the Laplace transform of the first passage time PDF is given by

$$F_1(s) = k_1(A\lambda_1^{L_1-1} + B\lambda_2^{L_1-1}), \qquad F_2(s) = k_2(C\beta_1^{L_2-1} + (A + B - C)\beta_2^{L_2-1}). \tag{13}$$

These expressions now contain a wealth of information about the moments of the escape time distributions. For example, the probability of reaching the correct absorbing site, $(i,j)=(L_1,0)$, is found by evaluating $F_1(s)$ at $s=0$. Furthermore, the mth moment of the arbitrary completion time is

$$T_T^{(m)} = \int_0^\infty t^m(f_1(t) + f_2(t))dt = (-1)^m\left(\frac{d^m F_1(s)}{ds^m} + \frac{d^m F_2(s)}{ds^m}\right)\Bigg|_{s=0}, \tag{14}$$

and the mth normalized moment of the escape time to the correct site $(i,j)=(L_1,0)$ is

$$T_1^{(m)} = \frac{(-1)^m}{F_1(0)}\left(\frac{d^m F_1(s)}{ds^m}\right)\Bigg|_{s=0}. \tag{15}$$

III. RESULTS AND DISCUSSION

The un-normalized Laplace transforms of the two branches, $F_1(s)$ and $F_2(s)$ provide a complete description of the completion process and in particular, we analyze two important quantities: (1) the probability that the process completes via one branch or the other and (2) the distribution of time needed for this completion. In the latter case, we concentrate our attention on the mean and variance of the completion times. For the general two-branch process, it is relatively simple to generate symbolic expressions for the completion probabilities and the moments of the completion times. Where these expressions are simple enough to be informative, we provide their explicit forms for which we use the following notation:

$$l_{1,2} = \lambda_{1,2}|_{s=0}; \quad b_{1,2} = \beta_{1,2}|_{s=0}; \quad \text{and} \quad A_0 = A|_{s=0}. \tag{16}$$

Where the expressions are not sufficiently compact, particularly for the higher moments of the completion time distributions, we use numerical examples to illustrate their dependence on parameters. For these numerical examples, we fix the length of each branch to involve $L_1=L_2=16$ steps. To explore the effect of different time scales in each branch, we consider the case when the forward rates of both branches are equal ($k_1=k_2$) and the case where the forward rate of the correct branch is six times that of the wrong branch ($k_1=6k_2$). In the following subsections we consider the specificity of these processes (Sec. III A), examine the completion time means (Sec. III B) and variances (Sec. III C), and finally show how these processes frequently simplify down to a corresponding three-point process (Sec. III D).

A. "Correct" and "wrong" completion probabilities

In a KPR process, the biochemical process must somehow give preference to completing in the correct way, i.e., adding the correct amino acid to the growing protein chain or initiating intracellular signaling when the correct ligand is bound to the receptor, but not when the incorrect ligand is bound. In our simplified model, this preference corresponds to reaching one absorbing site rather than the other. Here we analyze how changes in the relevant parameters affect this preference. Following the derivations in Sec. II, we can write the correct completion probability as

$$P_C = F_1(0) = k_1 l_1^{L_1-1}(1 - l_1/l_2)A_0, \tag{17}$$

and the wrong completion probability as $P_W=1-P_C$.

For example, one can use these expressions to derive expressions for the directional completion probabilities for the directed KPR (dKPR) scheme ($\gamma_{1,2}>0$ and $r_{1,2}=0$), which are

$$P_{C\text{-dKPR}} = \frac{(k_1/k_2)(1+\psi_2)^{L_2-1}}{(1+\psi_1)^{L_1-1} + (k_1/k_2)(1+\psi_2)^{L_2-1}}, \tag{18}$$

and $P_{W\text{-dKPR}}=1-P_{C\text{-dKPR}}$, where we have used the notation $\psi_{1,2}=\gamma_{1,2}/k_{1,2}$.

Figure 2(a) shows the probability of completing in the first direction as a function of the KPR ratios $\psi_{1,2}$ in the case of equal forward rates ($k_1=k_2=1$). From the figure, it is apparent that a large amount of specificity is achievable for the properly chosen combination of ψ_1 and ψ_2. For example, the system will complete in the correct direction more than 99.99% of the time for any (ψ_1, ψ_2) combination in the lower

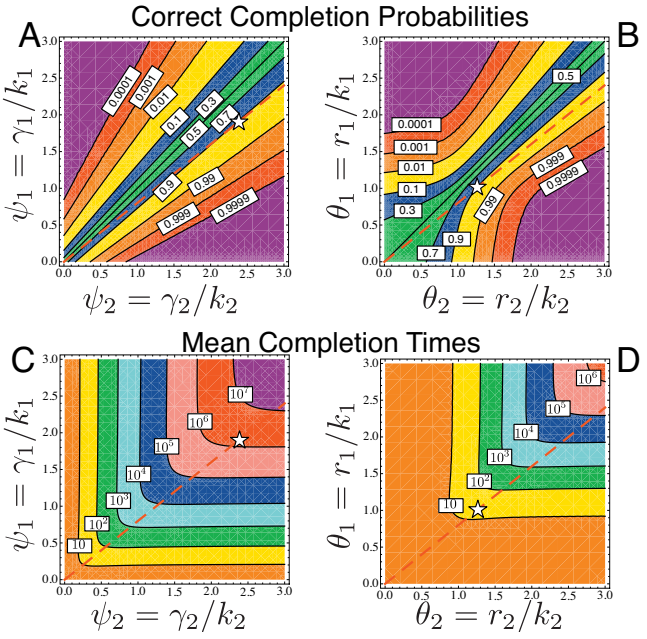

FIG. 2. Proofreading with equal forward rates, $k_1=k_2=1$. Contour plots of [(a) and (b)] the probability of correct completion and [(c) and (d)] the corresponding mean decision time for two different decision processes. [(a) and (c)]. For the dKPR process with varying KPR rates $\psi_1=\gamma_1/k_1$ and $\psi_2=\gamma_2/k_2$ and zero backward rates, $r_{1,2}=0$. [(b) and (d)] For the AM process with varying backward rates $\theta_1=r_1/k_1$ and $\theta_2=r_2/k_2$ and zero proofreading rates, $\gamma_{1,2}=0$. For both plots, the lengths of the branches are $L_1=L_2=16$, and the contour lines denote the probabilities of correct completion (upper panels) or mean completion time in units of $1/k_2$ (lower panels). The red dashed line corresponds to a 20% difference in the proofreading or backward ratios, $\psi_1=0.8\psi_2$ or $\theta_1=0.8\theta_2$, respectively.

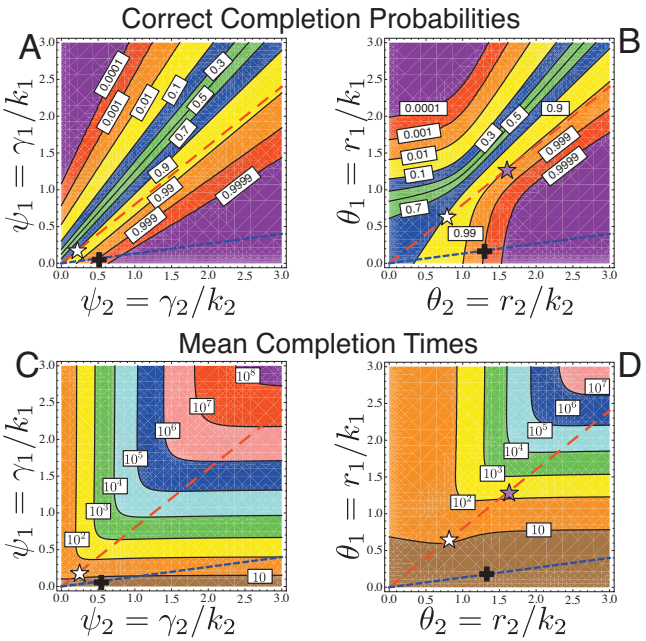

FIG. 3. Proofreading with different forward rates. Same as for Fig. 2, except for different forward rates: $k_1=6$ and $k_2=1$. The red dashed line corresponds to a 20% difference in the proofreading or backward ratios, $\psi_1=0.8\psi_2$ or $\theta_1=0.8\theta_2$, respectively. The blue dashed line corresponds to a 20% difference in the proofreading or backward rates, $\gamma_1=0.8\gamma_2$ or $r_1=0.8r_2$, respectively. The stars correspond to ratios $\psi_1=0.8\psi_2$ or $\theta_1=0.8\theta_2$ and 90% (white) or 99% (purple) specificity, and the cross corresponds to rates $\gamma_1=0.8\gamma_2$ or $r_1=0.8r_2$ and 99.9% specificity.

right corner. Similarly, one can compute the directional probabilities in the case of the absorption mode (AM) (Ref. 19) process [see Fig. 2(b)], where $\gamma_{1,2}=0$, but the backward rates $r_{1,2}$ are allowed to vary. In this case, the contour lines for the completion probabilities are less trivial than for the dKPR case. In particular, the contour lines exhibit a bottleneck near the values of $\theta_{1,2}\equiv r_{1,2}/k_{1,2}=1$, where the specificity can change dramatically despite relatively small changes in the parameter values.

The objective of KPR is to provide large amplification in directional specificity despite small changes in the parameters ψ or θ. To compare how well the dKPR and AM processes achieve this objective, we have drawn red dashed lines in each plot corresponding to $\psi_1=0.8\psi_2$ or $\theta_1=0.8\theta_2$, i.e., there is a 20% difference in the relative proofreading or backward ratios, respectively, between the two branches. Since $k_1=k_2$, this is equivalent to exploring a 20% difference in the actual rates γ and r. As the backward and proofreading rates increase, the specificity also increases for both process, as can be seen by how the dashed lines cross the contour levels. The first observation to note is that both the dKPR and the AM process can attain 90% specificity with 20% difference in rates [see stars in Figs. 2(a) and 2(b)] and values of the parameters which are within the range of the plots.

Figures 3(a) and 3(b) show the completion probabilities for a case where the forward rates are different from one branch to the next. While many qualitative trends of this case are similar to the previous case with equal forward rates, the analysis becomes a little more complicated. First, the different forward rates already provide a certain amount of correction $(k_1/(k_1+k_2)=6/7)$ before any additional effects of proofreading or backward rates. In turn, the proofreading and backward rates can amplify this specificity much higher than in the previous case. With different forward rates, one can consider small relative changes in the ratios (ψ or θ, red dashed lines) or in the absolute rates (γ or r, blue dashed lines). With 20% change in the *ratios* ($\psi_1=0.8\psi_2$ or $\theta_1=0.8\theta_2$), either process can attain a 90% specificity (white stars) but only the AM process is capable of providing 99% specificity (purple star) within the parameter range shown in the figure. When the actual *rates* γ or r are slightly varied from one branch to the other another (blue dashed lines), 99.9% (black cross), greater specificity is achievable with either model. Indeed, a high level of specificity is achievable in either process even when these rates are identical, so long as the forward rates are different (not shown).

B. Average completion times

In addition to forming the correct product, a biochemical process must also complete this construction in a timely manner. For example, the AM and dKPR schemes may make the same amplification of specificity, but one may be able to do so faster than the other. While a detailed analysis of this tradeoff between specificity and efficiency is left for future work, we begin to explore this aspect of the system by examining the mean completion time. Although the expressions for the mean completion times are trivial to generate, they are cumbersome to write in the general case. Therefore, in

the interest of brevity, we provide explicit expressions only for the case of dKPR, for which the mean correct completion time is given by

$$T_{C\text{-dKPR}} = -\frac{\left(\frac{k_1}{k_2}\right)(1+\psi_1)[1-(1+\psi_2)^{L_2}]+\psi_2\left[(1-L_1)(1+\psi_2)+\left(\frac{k_1}{k_2}\right)L_2(1+\psi_1)\right]}{k_1\psi_2(1+\psi_1)\left[(1+\psi_1)^{L_1}(1+\psi_2)+\left(\frac{k_1}{k_2}\right)(1+\psi_1)(1+\psi_2)^{L_2}\right]}$$
$$-\frac{\left(\frac{k_1}{k_2}\right)(1+\psi_2)^{L_2}[1+\psi_1(2+\psi_1)][1-(1+\psi_1)^{L_1}]}{k_1\psi_1(1+\psi_1)^{L_1+1}\left[(1+\psi_1)^{L_1}(1+\psi_2)+\left(\frac{k_1}{k_2}\right)(1+\psi_1)(1+\psi_2)^{L_2}\right]}. \quad (19)$$

The mean wrong completion time $T_{W\text{-dKPR}}$ is given by interchanging the subscripts 1 and 2 in the expression for $T_{C\text{-dKPR}}$, and the average arbitrary completion time is the weighted sum of these two directional completion times

$$T_{\text{dKPR}} = P_{C\text{-dKPR}}T_{C\text{-dKPR}} + P_{W\text{-dKPR}}T_{W\text{-dKPR}}. \quad (20)$$

Figures 2(c) and 2(d) show contour plots for the average completion times of the dKPR and AM processes with equal forward rates, $k_1=k_2$. These plots show that as the backward or proofreading rates increase, the amount of time required to complete the process increases exponentially. While we saw in Figs. 2(a) and 2(b) that both processes were able to provide 90% specificity (for 20% difference in the backward/proofreading rates), the AM process can provide it with a much smaller mean completion time. Similarly, Figs. 3(c) and 3(d) show contour plots of the mean completion times of the dKPR and AM processes with $k_1=1$ and $k_2=6$. We can see again that for a 20% difference in the backward/proofreading rates (blue dashed lines) or their ratios to the corresponding forward rates (the red dashed lines), the AM process can provide the requested specificity for much smaller average completion times.

To better understand the behavior of the mean completion time, we illustrate in Fig. 4 the effects that changes in the parameters $\psi_{1,2}$ have on these mean completion times for the process in which the forward rate on the correct branch is six times the rate on the wrong branch, $k_1=6k_2$. At first glance at Fig. 4(a) or Fig. 3(c), it appears that the behavior of the mean arbitrary completion time is somewhat trivial–as one increases the proofreading rates in both branches, the mean waiting time also increases. However, by zooming in along certain strips of this plot, one finds additional dependencies of the mean waiting times on the parameters. Suppose that one fixes ψ_1 to some nonzero value and then changes ψ_2 [see top edge of Fig. 4(b)]. When ψ_2 is zero, the second branch is biased forward, and the process will quickly complete soon after it enters into that branch. Conversely, when ψ_2 is very large, the process will spend very little time in the second branch, and the process reduces down to the single-branch process as if that second branch

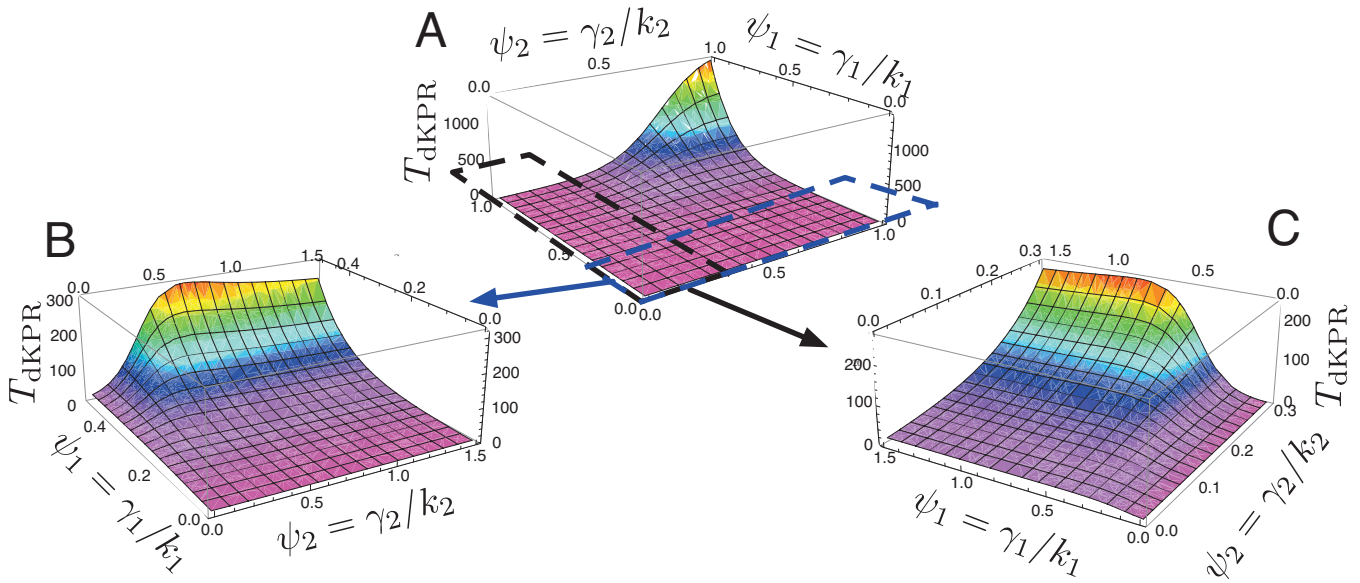

FIG. 4. Plots of the mean arbitrary completion times (units of $1/k_2$) for the dKPR process, with two branches of lengths $L_1=L_2=16$ and forward rates $k_1=6$ and $k_2=1$. Panels b and c show a zoomed in perspective of the mean completion times corresponding to the parameter regions indicated in panel a. Note that we use a relative color scale in each of the subplots.

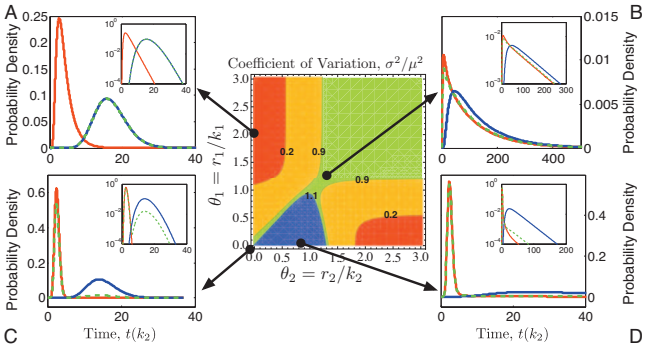

FIG. 5. Contour plot of the coefficient of variation (of the arbitrary completion time) vs $\theta_1=r_1/k_1$ and $\theta_2=r_2/k_2$ and typical completion time distributions. We used the case of zero proofreading rates, $\gamma_{1,2}=0$. We also set $k_1=6$ and $k_2=1$. The different colors correspond to different behaviors of the completion time distributions (see text for more details). The side panels [(a)–(d)] show the distributions of completion times in the correct (red) and incorrect (blue) directions and the arbitrary completion time distribution (green). The inset in each of the panels shows a semilog plot of the distribution to amplify the differences between the lines.

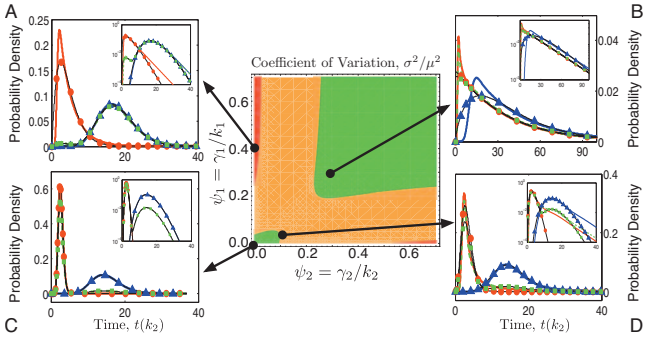

FIG. 6. Contour plot of the coefficient of variation (of the arbitrary completion time) vs $\psi_1=\gamma_1/k_1$ and $\psi_2=\gamma_2/k_2$ and typical completion time distributions. The backward rates are zero, $r_{1,2}=0$, and the forward rates are $k_1=6$ and $k_2=1$. The different colors correspond to different behavior of the completion time distributions (see text for more details, the color scheme is the same as in Fig. 5). The side panels [(a)–(d)] show the distributions of directional and total completion times (same color code as in Fig. 5). The markers correspond to the best fit for a reduced three-state model approximation to the processes.

were not there. However, when ψ_2 is in some middle range, the process will spend significant amounts of time in each of the two branches, thereby increasing the total time until completion. Similar observations can be made for the AM process (not shown), as should be expected from the nontrivial shape of the contours of Fig. 3(d).

C. Variance in completion times

In addition to specificity and the average completion time, a completion process can further be characterized by the shape of its completion time distribution. For some parameters this distribution will have a small variance, and the decision is made in some seemingly deterministic amount of time. For other parameters, the distribution may be much broader (the same behavior was found for single-branch processes, see Ref. 42). The relative broadness of this shape can be described by the squared coefficient of variation ($CV^2=\sigma^2/\mu^2$, where σ^2 is the variance and μ is the mean) of the completion time distribution. The second moments, and therefore the variances, can be derived according to the general relation of Eqs. (14) and (15), but the resulting expressions are too long to provide much valuable insight even in the case of dKPR. Instead, we rely on parametric studies to explore how parameters affect the completion time distribution shapes.

In what follows, we consider the same cases as above and classify the shapes of the resulting completion time distributions. First, we consider the case of zero proofreading rates, $\gamma_{1,2}=0$. Figure 5 shows a contour plot of the coefficient of variation of the arbitrary completion time versus $\theta_1=r_1/k_1$ and $\theta_2=r_2/k_2$ and the side panels show correct (red), wrong (blue) and "arbitrary" (green) completion time distributions for the parameter values $k_1=6k_2$ and $\{(\theta_1,\theta_2)\}=\{(2,1),(1.2,1.2),(0,0),(0,0.88)\}$. The side panels show that forward biased ($\theta<1$) branches have completion time distribution that are well represented by a gamma distribution [see red lines in Figs. 5(c) and 5(d) and blue lines in Figs. 5(a) and 5(c)]. Conversely, backward biased ($\theta>1$) branches exhibit exponential completion time distributions [see red lines in Figs. 5(a) and 5(b) and blue lines in Figs. 5(b) and 5(d)]. In turn, the coefficient of variation and shape of the total completion time distribution is determined by some combination of the two branches. When both branches are biased backward, the total CV^2 is about unity, and the distribution is well approximated by an exponential distribution [see large green area in the upper right corner of Fig. 5 and green line in Fig. 5(b)]. When one branch is strongly biased backward while the other is biased forward, the process is much more likely to finish along the forward biased branch. Hence, the total completion time distribution is well approximated by a single narrow Gamma distribution, and its CV^2 is less than unity [see red areas of Fig. 5 and the green line in Fig. 5(a)]. When both branches are biased forward, the arbitrary completion time distribution has a bimodal shape corresponding to the simple combination of two gamma distributions [see green line in Fig. 5(c)]. Finally, when one branch is strongly biased forward while the other is almost unbiased, we obtain a much less trivial total completion time distribution. In this case, the completion time distribution can be broader than exponential (i.e., $CV^2>1$) as is shown in Fig. 5(d) for the point of maximal CV^2. We now consider the case where there is proofreading ($\gamma_{1,2}>0$) but no backward reactions, $r_{1,2}=0$. Figure 6 shows a contour plot of the coefficient of variation of the arbitrary completion time versus $\psi_1=\gamma_1/k_1$ and $\psi_2=\gamma_2/k_2$ and typical completion time distributions for the parameter values $k_1=6k_2$ and $\{(\psi_1,\psi_2)\}=\{(0.4,0),(0.3,0.3),(0,0),(0.05,0.1)\}$. As above in Fig. 5, we can divide the parameters space into few regions with different shapes for the completion time distribution. For example, the large green area (color online) corresponds to $CV^2\sim 1$ and where the directional and arbitrary completion time distributions are well approximated by exponential distributions [see Fig. 6(b)]. Similarly, for the small red areas where one branch is biased backward and the other forward, the completion time along the backward biased branch is nearly exponential, while the completion time along the forward biased branch is effectively described by a gamma distribution [see Fig. 6(a)].

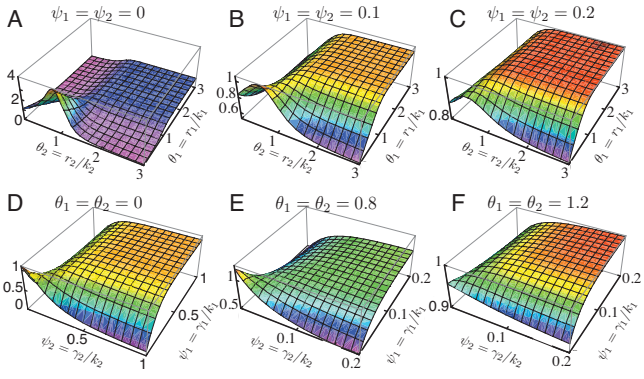

FIG. 7. The coefficient of variation as a function of different parameters in the two-branch gKPR process. [(a)–(c)] CV^2 vs θ_1 and θ_2 for three different fixed values $\psi_{1,2}$. [(d) and (f)] CV^2 vs ψ_1 and ψ_2 for three different fixed values $\theta_{1,2}$. In all cases, when both branches are strongly backward biased $CV^2 \sim 1$ and the completion time distribution is exponential. A relative color scale is used in each of the subplots.

We now turn to the more general case where there is both proofreading and backward reactions ($\gamma_{1,2} > 0, r_{1,2} > 0$). For this case, Fig. 7 shows a three-dimensional plot of the coefficient of variation of the arbitrary completion time versus $\theta_{1,2}$ (upper line) or $\psi_{1,2}$ lower line. These figures emphasize the different effects of changes in θ or ψ. While in all cases strong backward bias on both branches (large $\theta_{1,2}$ or $\psi_{1,2}$) lead to an exponential distribution of the completion time, backward bias has different dependence on the system size and different ranges for θ and ψ.

D. Simplification of the two-branch decision process

In examining the distributions in Figs. 5(a)–5(d), one observes that the completion time distribution of each branch is often similar to a gamma distribution (or an exponential distribution, which is a special case of the gamma distribution). This suggests that one should frequently be able to replace the entire process with a simple three-state chain, as shown in Fig. 8 with the following properties. Each direction (1,2) is assumed to have a non-normalized Gamma distributed completion time with density

$$f_1(t) \approx \tilde{f}_1(t,x_1,y_1) = \alpha t^{x_1-1} y_1^{x_1} \frac{\exp(-y_1 t)}{\Gamma(x_1)},$$

$$f_2(t) \approx \tilde{f}_2(t,x_2,y_2) = (1-\alpha) t^{x_2-1} y_2^{x_2} \frac{\exp(-y_2 t)}{\Gamma(x_2)},$$

where $0 \le \alpha \le 1$ denotes the probability of completion in the first direction. Thus, the total probability density of completing along either branch at time t is approximated by

$$f_T(t) \approx \tilde{f}_T(t) = \tilde{f}_1(t,x_1,y_1) + \tilde{f}_2(t,x_2,y_2).$$

In numerical studies, we have attempted to find parameter sets $\Lambda = \{x_1, y_1, x_2, y_2, \alpha\}$ that best match the direction and time distribution of the full escape process in the one norm sense. In other words, we have found the Λ such that

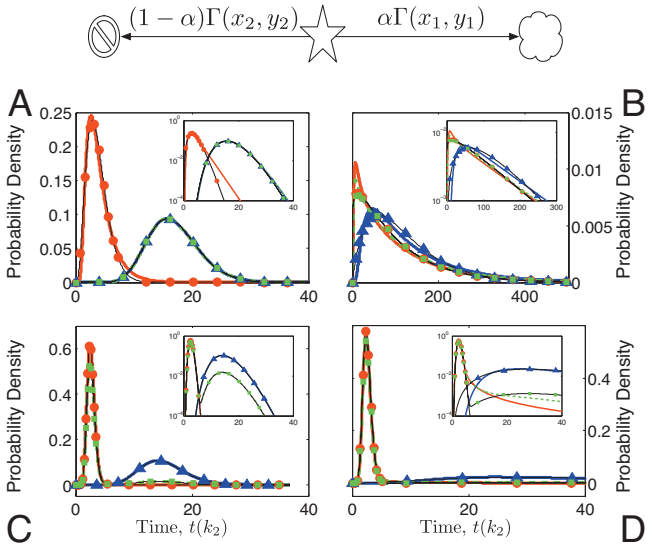

FIG. 8. Three-state model approximation of the original completion time problem. (Top) Schematic description of the three-state model where the conditional escape time in each direction is given by a gamma distribution. [(a)–(d)] Comparison of the escape time distributions using the full original and the reduced three-state model. The parameters used here are the same as those in Figs. 5(a)–5(d).

$$\Lambda = \underset{\{x_1,y_1,x_2,y_2,\alpha\}}{\arg\min} \sum_{n=1}^{2} \int_0^\infty |f_n(t) - \tilde{f}_n(t,\Lambda)|_1 dt. \quad (21)$$

In most cases, we find that this approximation and optimization does an excellent job of capturing the qualitative and quantitative behaviors of the complete process as is shown in Figs. 8(a)–8(d). To further explore the ability of the reduced model to capture the behavior of the full system, we have explored the original parameter space $\{\theta_1, \theta_2\}$ in order to find the regions where this approximation is most valid. From Fig. 9(a), we immediately see that the approximation is valid in all four corners of the contour plot where both θ_1 and θ_2 are either relatively large or relatively small—that is where both branches are biased in one direction or another. However, even in the regions where one or both branches are unbiased ($\theta_1 \approx 1$ or $\theta_2 \approx 1$), we note that the fit is still quite good. Indeed for this system, we can always find a parameter set $\{x_1, y_1, x_2, y_2, \alpha\}$ that captures the full escape time distribution within error [defined by the norm in Eq. (21)] of 0.2. In order to illustrate this approximation success, Fig. 9(b)

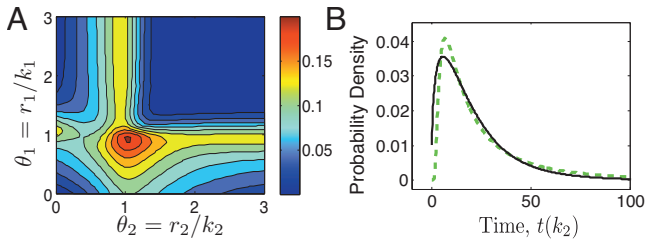

FIG. 9. Numerical comparison of the completion time distributions for the approximate three-state model and the full two-branch process. (a) Contour plots of the approximation error [Eq. (21)] vs the ratios $\theta_{1,2} = r_{1,2}/k_{1,2}$. (b) Illustration of approximate (dashed line) and actual (solid line) completion time distributions (in units of $1/k_2$) for the parameter set ($\theta_1 = 1.03$, $\theta_2 = 0.95$), which corresponds to the largest approximation error of $\Sigma_{n=1}^{2} \int_0^\infty |f_n(t) - \tilde{f}_n(t)|_1 dt = 0.20$.

shows the actual (solid line) and approximate (dashed line) distributions for the case where the fit is the worst ($\theta_1 = 1.03$, $\theta_2 = 0.95$). For every other case, we were able to find a three-state model that did an even better job of matching the full system behavior.

As was the case for the AM process ($\gamma_{1,2}=0$), the dKPR process ($r_{1,2}=0$) is well captured by the same three-state process defined above. To illustrate this, the colored lines in Figs. 6(a)–6(d) correspond to the full system completion time distributions, and the markers correspond to the approximate three-state system.

IV. CONCLUSIONS

In this work we have begun the exploration of the temporal properties of KPR schemes. To accomplish this, we have derived analytical expressions for the Laplace transform of the occupation probabilities from which we obtained the completion time distributions. With this analysis, we have enabled the simple derivation of expressions for the completion time moments. Some of these expressions, such as completion probabilities and the mean waiting times for certain processes are simple enough to be shown explicitly, while others are just as easily derived, but are omitted since their form is too long and not very informative. To enable a better understanding of the interplay of specificity and temporal behaviors, we focused on the first two moments of the completion times, as well as on the completion probabilities (which is actually the zeroth moment). We showed that, for most parameter sets, each of the considered proofreading schemes can be reduced to a three-state process with simple distributions for the waiting times between transitions. The simplified process captures most of the relevant features of KPR schemes, namely, the specificity as well as the magnitude and shape of the completion time distributions. However, the dependence of the simplified behavior on the full system's kinetic parameters is different for the various proofreading schemes, suggesting that some important information about the process is retained despite the simplification.

We have explicitly considered different kinetic schemes including the traditional dKPR scheme where catastrophic reactions force the process to restart, as well as an AM scheme where single-step intermediate reactions can provide the same specificity. Surprisingly, we find that in most cases the simpler AM process outperforms the dKPR process by providing a higher degree of specificity in a shorter amount of time. It is also worth mentioning that the dKPR or general KPR processes violate the detailed balance conditions and therefore are necessarily nonequilibrium processes. The AM process on the other hand may satisfy the detailed balance condition and in this case is an equilibrium process. In this sense, the AM process has the added advantage in that it conserves energy, while the dKPR process must be continually driven with externally applied energy.

High specificity appears in many biological systems and likely results from many different kinetic schemes—suggesting that one needs as much information as possible to distinguish between one such mechanism and the next. Therefore, in addition to using the specificity and mean completion times to compare the different processes, we have also used analyses of the completion time distributions to classify different kinetic schemes and parameter values into separate regimes where these distributions take on different qualitative shapes. By providing this additional information, the temporal analysis and classification tools developed here can more precisely support or oppose hypotheses of particular KPR models for particular biochemical systems. In the future, the next logical step is to apply these tools in order to identify parameters and infer kinetic mechanisms from experimental measurements of completion time distributions.

ACKNOWLEDGMENTS

We thank N. Hengartner, J. Hopfield, and N. Sinitsyn for discussions during early stages of this work. We also thank B. Goldstein, R. Gutenkunst, M. Monine, and especially M. Savageau for helpful comments regarding this work. This work was partially funded by LANL LDRD program.

[1] J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **71**, 4135 (1974).
[2] J. Yan, M. Magnasco, and J. Marko, Nature (London) **401**, 932 (1999).
[3] A. Sancar, K. Unsal-Kacmaz, and S. Linn, Annu. Rev. Biochem. **73**, 39 (2004).
[4] M. Goulian, Z. J. Lucas, and A. Kronberg, J. Biol. Chem. **243**, 627 (1968).
[5] S. C. Blanchard, R. L. Gonzalez Jr., H. D. Kim, S. Chu, and J. D. Puglisi, Nat. Struct. Mol. Biol. **11**, 1008 (2004).
[6] T. Jovanovic-Talisman, J. Tetenbaum-Novatt, A. S. McKenney, A. Zilman, R. Peters, M. P. Rout, and B. T. Chait, Nature (London) **457**, 1023 (2009).
[7] T. Mckeithan, Proc. Natl. Acad. Sci. U.S.A. **92**, 5042 (1995).
[8] J. D. Rabinowitz, C. Beeson, D. S. Lyons, M. M. Davis, and H. M. McConnell, Proc. Natl. Acad. Sci. U.S.A. **93**, 1401 (1996).
[9] C. Rosette, G. Werlen, M. Daniels, P. Holman, S. Alam, P. Travers, N. Gascoigne, E. Palmer, and S. Jameson, Immunity **15**, 59 (2001).
[10] Z.-J. Liu, H. Haleem-Smith, H. Chen, and H. Metzger, Proc. Natl. Acad. Sci. U.S.A. **98**, 7289 (2001).
[11] B. Goldstein, J. Faeder, and W. Hlavacek, Nat. Rev. Immunol. **4**, 445 (2004).
[12] J. R. Faeder, W. S. Hlavacek, I. Reischl, M. L. Blinov, H. Metzger, A. Redondo, C. Wofsy, and B. Goldstein, J. Immunol. **170**, 3769 (2003).
[13] C. F. Springgate and L. A. Loeb, J. Mol. Biol. **97**, 577 (1975).
[14] J. Ninio, Biochimie **57**, 587 (1975).
[15] R. R. Freter and M. Savageau, J. Theor. Biol. **85**, 99 (1980).
[16] M. Savageau, J. Theor. Biol. **93**, 179 (1981).
[17] A. Zilman, J. Pearson, and G. Bel, Phys. Rev. Lett. **103**, 128103 (2009).
[18] M. D'Orsogna and T. Chou, Phys. Rev. Lett. **95**, 170603 (2005).
[19] S. Redner, *A Guide To First-Passage Processes* (Cambridge University Press, Cambridge, 2001).
[20] G. Bel and E. Barkai, Phys. Rev. Lett. **94**, 240602 (2005).
[21] G. Bel and E. Barkai, Phys. Rev. E **73**, 016125 (2006).
[22] C. L. Lee, G. Stell, and J. Wang, J. Chem. Phys. **118**, 959 (2003).
[23] T. Koren, J. Klafter, and M. Magdziarz, Phys. Rev. E **76**, 031129 (2007).
[24] N. van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed. (Elsevier, New York, 2001).
[25] B. Munsky and M. Khammash, J. Chem. Phys. **124**, 044104 (2006).
[26] K. Burrage, M. Hegland, S. Macnamara, and R. Sidje, Proceedings of the A. A. Markov 150th Anniversary Meeting, 2006 (unpublished), p. 21.
[27] B. Munsky and M. Khammash, J. Comput. Phys. **226**, 818 (2007).
[28] S. Peleš, B. Munsky, and M. Khammash, J. Chem. Phys. **125**, 204104 (2006).
[29] B. Munsky and M. Khammash, IEEE Trans. Autom. Control **52**, 201 (2008).
[30] Y. Lan, P. G. Wolynes, and G. A. Papoian, J. Chem. Phys. **125**, 124106 (2006).
[31] N. Sinitsyn, N. Hengartner, and I. Nemenman, Proc. Natl. Acad. Sci. U.S.A. **106**, 10546 (2009).
[32] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).

[33] D. T. Gillespie, J. Chem. Phys. **115**, 1716 (2001).
[34] Y. Cao, D. Gillespie, and L. Petzold, J. Chem. Phys. **122**, 014116 (2005).
[35] B. Munsky and M. Khammash, IET Syst. Biol. **2**, 323 (2008).
[36] C. Dellago, P. Bolhuis, F. Csajka, and D. Chandler, J. Chem. Phys. **108**, 1964 (1998).
[37] A. Faradjian and R. Elber, J. Chem. Phys. **120**, 10880 (2004).
[38] D. Moroni, P. Bolhuis, and T. van Erp, J. Chem. Phys. **120**, 4055 (2004).
[39] T. van Erp and P. Bolhuis, J. Comput. Phys. **205**, 157 (2005).
[40] R. Allen, D. Frenkel, and P. Rein ten Wolde, J. Chem. Phys. **124**, 024102 (2006).
[41] T. Lu, T. Shen, C. Zong, J. Hasty, and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **103**, 16752 (2006).
[42] G. Bel, B. Munsky, and I. Nemenman, "Simplicity of completion time distributions for common complex biochemical processes," Phys. Biol. (in press).