Three-state kinetic mechanism for scaffold-mediated signal transduction

Jason W. Locasale*
Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

(Received 26 June 2008; revised manuscript received 30 August 2008; published 21 November 2008)

Signaling events in eukaryotic cells are often guided by a scaffolding protein. Scaffold proteins assemble multiple proteins into a spatially localized signaling complex and exert numerous physical effects on signaling pathways. To study these effects, we consider a minimal, three-state kinetic model of scaffold-mediated kinase activation. We first introduce and apply a path summation technique to obtain approximate solutions to a single molecule master equation that governs protein kinase activation. We then consider exact numerical solutions. We comment on when this approximation is appropriate and then use this analysis to illustrate the competition of processes occurring at many time scales that are involved in signal transduction in the presence of a scaffold protein. We find that our minimal model captures how scaffold concentration can influence the times over which signaling is distributed in kinase cascades. For a range of scaffold concentrations, scaffolds allow for signaling to be distributed over multiple decades. The findings are consistent with recent experiments and simulation data. These results provide a framework and offer a mechanism for understanding how scaffold proteins can influence the shape of the waiting time distribution of kinase activation and effectively broaden the times over which protein kinases are activated in the course of cell signaling.

DOI: 10.1103/PhysRevE.78.051921 PACS number(s): 87.16.Xa, 87.18.Mp, 87.18.Vf, 82.20.Uv

INTRODUCTION

Cells detect external signals in the form of stresses, growth factors, DNA damage, hormones, among many others, and integrate them to achieve an appropriate biological response [1]. Biochemical modifications in the form of reversible phosphorylations by enzymes known as kinases are detected by proteins to form networks that are used to integrate these signals [2]. These complex networks are comprised of numerous modular structures that allow for many different biological responses. Signal propagation through these networks is often guided by a scaffold protein [3]. Scaffold proteins assemble multiple kinases (that are activated sequentially in a cascade) in close proximity to form signaling complexes. Scaffold proteins are believed to regulate biochemical signaling pathways in a multitude of ways [3–5].

Experiments have suggested that the scaffold proteins have profound effects on regulating signaling dynamics [6–8]. In particular, a key parameter is believed to be the concentration of scaffold proteins. Recent simulation results [9], which elaborated on these findings, showed that one effect that the concentration of scaffold proteins may have is to control the shape of the waiting time distribution of activation. Recent work has shown that the waiting time distribution is closely related to signal duration (e.g., the time over which an active signaling intermediate persists) [9]. Signal duration is known to be an important determinant in many cell decision making processes [10–12] and therefore, a knowledge of how the concentration of scaffold proteins affect this waiting time distribution is important to understand. The waiting time distribution has been used in multiple theoretical contexts [13,14] to study signaling dynamics and has been measured in different experimental contexts in diverse biology systems [15,16].

In this work, we present a minimal model that aims to understand how changes in signaling dynamics manifested through the first passage time statistics are affected solely by changes in scaffold concentration. The purpose of this study is to first construct and then solve a minimal model that aims to capture these desired effects. Many other factors are undoubtedly important in determining how signaling dynamics are regulated in complex biochemical pathways. These factors include but are not limited to feedback control, allosteric regulation by the scaffold, degradation and internalization rates of the complexes along with many others and have been discussed elsewhere [5,17]. Other complexities such as the multiple phosphorylation sites and the processivity and distributivity of the phosphorylation network [18] also affect the dynamics of signal output. Endocytosis and the time scales associated with protein degradation are also important. Our aim is to investigate concentration effects of scaffolds on regulating signaling dynamics which have shown to be important in experiments and simulations.

We present a coarse grained, minimal model that illustrates how the waiting time distribution of protein kinase activation is modified by the presence of different amounts of scaffold protein. The model involves multiple states in which a single protein kinase, situated at the end of a cascade, resides and corresponding transitions between these states are allowed [19]. We analyze the resulting master equation by first introducing an approximate scheme that involves a weighted path summation over the possible trajectories that an individual kinase can take in the course of its transition from an inactive to an active state [20]. We also consider numerical solutions. We find that, consistent with known simulation results, in certain limits the waiting time distribution of activation sharply decays and is effectively characterized by a single exponential whereas in other re-

*Current address: Department of Systems Biology, Harvard Medical School, Boston, MA 02115; jlocasal@bidmc.harvard.edu

gimes, the waiting time distribution takes on a more complicated form. Our model provides a simple mechanistic description for how scaffold proteins and differences in their concentrations may regulate the waiting time distribution of kinase activation.

TIME SCALES FOR SIGNAL TRANSDUCTION THROUGH SCAFFOLD PROTEINS

Let us first consider physically, the time scales involved in scaffold-mediated cell signaling. The signaling event that we consider consists of the sequential activation of multiple enzymes (kinases) in a cascade. Consider the processes that must occur in order for a kinase at the end of a scaffolded kinase cascade to become activated. In this picture, we follow the trajectory of a single molecule as it interacts with the scaffold and the upstream components of the pathway in which it is involved. Within the biochemical pathway, kinases in solution must encounter its targeted substrates by diffusion; therefore, encounter (or diffusion) times for the kinase in a sequence of a multi-tiered biochemical cascade, are important. These encounter times τ_{ec}, determined by diffusive motion of proteins, behave as $\tau_{ec} \sim (D\rho^{d/2})^{-1}$ in d dimensions, for a concentration, ρ and diffusivity D. Other time scales $(\tau_k, \tau_p, \tau_{\text{on}}, \tau_{\text{off}})$ arise from rates of catalysis and protein-protein interactions such as the binding of a kinase to a scaffold. These times are for activation (τ_k) and deactivation (τ_p) by a kinase and phosphatase (an enzyme that removes a phosphate group) as well as for binding (τ_{on}) and unbinding (τ_{off}) to and from a scaffold.

We investigated the dynamical consequences of a model in which, in the course of activation of a single kinase, the collection of microscopic processes described above interacts with relative scaffold density ζ to give rise to several processes (involving state transitions of single molecule) with eight characteristic time scales $\tau_i; \tau_i \in \{\tau_{1,\ldots,}\tau_8\}$. Scaffold density ζ has been shown in many contexts to be a key variable regulating signal transduction. If too few scaffolds are present, signaling occurs predominantly in solution. If too many scaffolds are present, proteins kinases exist predominantly in complexes that are incompletely assembled. There exists, therefore, an optimal concentration to assemble complete signaling complexes of multiple kinases [3]. Schematics of these different scenarios are shown on the bottom of Fig. 1.

Kinases in solution can upon binding to a scaffold, be assembled into a complex that cannot effectively signal. This is because the complex does not have a complete set of kinases bound to it. Such a kinase then would be trapped in a signaling incompetent state until it either disassociates from the complex or the requisite kinases upstream bind to the complex. Association and disassociation of kinases to and from incompletely assembled complexes are denoted by times τ_1 and τ_2 and are functions of scaffold density, diffusion times, catalysis rates, and binding kinetics.

This scenario also requires additional time scales: the times required for a kinase to bind and disassociate from solution into a signaling incompetent complex (τ_3 and τ_4, respectively) and the times required for a signaling incompetent complex to transition to a signaling competent complex (τ_5 and τ_6, respectively) must be accounted for. Finally activation of the given kinase can occur with times τ_7 and τ_8 that involve diffusion and catalysis. The eight time scales comprising the model are functions of the following microscopic times:

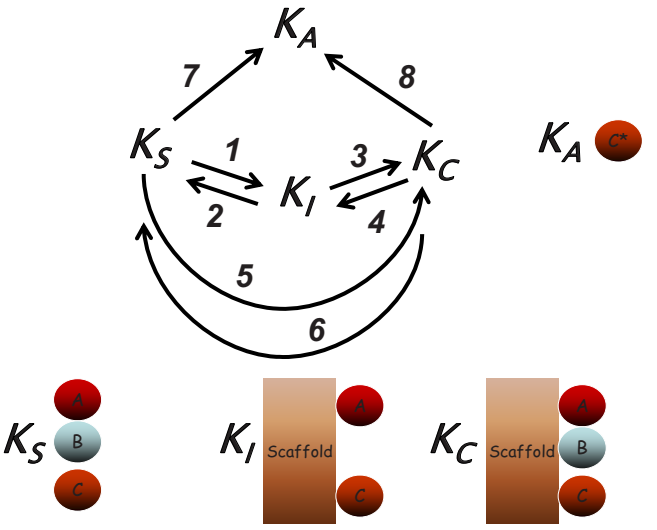

FIG. 1. (Color online) A three-state kinetic mechanism that characterizes scaffold-mediated cell signaling. Important time scales in scaffold-mediated signaling shown in a graph of a multistate kinetic model whose dynamics are governed by eight transitions. The final kinase molecule in a signaling cascade (C) can transition between three states denoted with subscripts: in solution (K_S), bound to a signaling competent complex (K_C), bound to a signaling incompetent complex (K_I), and eventually reaches an activated absorbing state, (K_A). The relative amount of each state and the rates of transitioning are functions of the density of scaffold proteins.

$$\tau_1 = f(\tau_{ec}, \tau_{\text{on}}, \tau_{\text{off}}, \tau_k, \tau_p, \zeta),$$

$$\tau_2 = f(\tau_{\text{off}}),$$

$$\tau_3 = f(\tau_{ec}, \tau_{\text{on}}, \tau_{\text{off}}, \tau_k, \tau_p, \zeta),$$

$$\tau_4 = f(\tau_{ec}, \tau_{\text{on}}, \tau_{\text{off}}, \tau_k, \tau_p, \zeta),$$

$$\tau_5 = f(\tau_{ec}, \tau_{\text{on}}, \tau_{\text{off}}\tau_k, \tau_p, \zeta),$$

$$\tau_6 = f(\tau_{\text{off}}),$$

$$\tau_7 = f(\tau_{ec}, \tau_k, \tau_p),$$

$$\tau_8 = f(\tau_{ec}, \tau_k, \tau_p).$$

We studied the dynamics of competing processes occurring at these eight phenomenological time scales.

MARKOV MODEL ILLUSTRATING THE COMPETITION BETWEEN MANY PROCESSES

We considered this kinetic model in which a kinase, at the end of a biochemical cascade, K_i can transition between four

states denoted with four subscripts: In solution (S), bound to a signaling competent complex (C), bound to a signaling incompetent complex (I), and activated (A). Any bound kinase that is a part of an incomplete complex is said to be in state I. Figure 1 shows a graph of the stochastic transitions to neighboring states that involve random waiting times that correspond to a set of eight random variables. The waiting time for a kinase to transition to a neighboring state is then Poisson distributed with rate constants, $k_i; k_i \in \{k_1, k_2, \ldots, k_8\}$. These rate constants are the inverse of the time scales previously mentioned. Thus for the ith process, the waiting time distribution, $F(\tau_i)$, is the probability density for the first passage time (FPT) distribution and takes the form $F(\tau_i)=k_i e^{-k_i \tau_i}$. Ultimately, the quantity of interest is the first-passage time distribution $F(t)$ (or its integrated value) for a kinase to transition to its activated state which we denote by $F(t)$. $F(t)$ is the time derivative of the cumulative probability distribution (CDF), $P(\tau_A < t) \equiv \int_0^t F(\tau\prime)d\tau\prime$; so that $\frac{d}{dt}P(\tau_A < t)=F(t)$ where τ_A is the random waiting time for activation of a protein kinase. The survival probability $S(t)$ is related to the FPT and the CDF in the following way, $S(t)=\int_t^\infty F(t\prime)dt\prime = 1-\int_0^t F(t\prime)dt\prime = 1-P(\tau_A < t)$. The kinetic equation, with an absorbing boundary condition at arrival at state K_A, is written as follows:

$$\frac{d}{dt}\vec{P} = Q \cdot \vec{P}, \tag{1}$$

where $\vec{P}(t)=[P_{K_A}(t)P_{K_S}(t)P_{K_C}(t)P_{K_I}(t)]^T$ and

$$Q = \begin{pmatrix} 0 & k_7 & k_8 & 0 \\ 0 & -(k_1+k_5+k_7) & k_6 & k_2 \\ 0 & k_5 & -(k_4+k_6+k_8) & k_3 \\ 0 & k_1 & k_4 & -(k_2+k_3) \end{pmatrix} \tag{2}$$

with the initial condition, $\vec{P}(0)=[0\ C_S\ C_C\ C_I]^T$ where C_S, C_C, and C_I are the probabilities that a given kinase initially resides in the solution, in a signaling-competent complex, or an incompletely assembled complex; and, for normalization $C_S+C_I+C_C=1$. In principle, an exact solution to the equation can be obtained by finding the eigenvalues and eigenvectors of Q. However, this calculation requires a solution to a cubic equation and is too complicated to extract any significant physical information—the solution to Eqs. (1) and (2) contains well over 100 terms. Therefore, we first employed an approximate method that in our view clearly shows the dependence of the relevant parameters on the behavior of the signaling dynamics. The method can also be applied to other kinetic schemes. We also considered numerical solutions.

PATH SUMMATION OF THE MASTER EQUATION

Formally, we can compute $P(\tau_A < t)$ by considering a weighted sum over all paths that lead to the absorbing state, K_A,

$$P(\tau_A < t) = \sum_{\text{steps},i=1}^{\infty} \sum_{\text{states},j=1}^{3} C_j \sum_{\text{branches}} \omega\left[\left(\sum_{\text{jumps},l=1}^{i} \tau_l\right) < t\right] \times \prod_{l=1}^{i} \prod_{l'} \omega(\tau_l < \tau_{l'}). \tag{3}$$

The first summation decomposes $P(\tau_A < t)$ into separate contributions for each set of paths that contain equivalent numbers of steps required to reach the absorbing state, K_A; i.e., for $i=1$, all paths requiring one jump are considered, for $i=2$, all paths requiring two jumps are considered, etc. Since there can be more than one path containing i steps leading to K_A, the next summation considers the weighted probability that, *a priori*, a kinase is in one of three states: In solution (S), bound to a signaling incompetent complex (I), or bound to a signaling competent complex (C); i.e., $j \in \{S, I, C\}$ and for normalization, $C_S+C_I+C_C=1$. The next summation occurs over each branch. A branch is defined here as a unique, way in which a path of fixed i and j can be traversed. We account for the probability that a specific path, i, with j steps on a branch is taken by computing the probability $\omega[(\Sigma_{\text{jumps},l=1}^{i}\tau_l) < t]$ that, in time t an enzyme transitions through a given sequence of jumps. We then avoid over counting by taking the union of this probability with the joint probability, $\Pi_{l=1}^{i}\Pi_{l\prime}\omega(\tau_l < \tau_{l\prime})$, that no transitions are made in the lth step along the path to any state, l', not along the considered path; τ_l is the waiting time to transition along the lth step of the path and $\tau_{l\prime}$ is the waiting time for a transition at the lth step to a position m that is not along the selected path. This term, $\Pi_{l=1}^{i}\Pi_{l\prime}\omega(\tau_l < \tau_{l\prime})$, ensures that each transition $l \to l+1$ takes place before any transitions to a point not along the given path.

Such a path summation is difficult to compute exactly but conveniently lends itself to approximate evaluations. We first denote the contribution of each path requiring i steps, a_i so that . Then, the contribution of a_i to the overall cumulative distribution, $P(\tau_A < t)$, decreases monotonically with increasing i. So, $a_i \geq a_{i+1}$ and thus

$$\sum_{j=1}^{3} C_j \sum_{\text{branches}} \omega\left[\left(\sum_{\text{jumps},l=1}^{i} \tau_l\right) < t\right] \prod_{l=1}^{i} \prod_{l'} \omega(\tau_l < \tau_{l'}) \geq \sum_{j=1}^{3} C_j \sum_{\text{branches}} \omega\left[\left(\sum_{\text{jumps},l=1}^{i+1} \tau_l\right) < t\right] \prod_{l=1}^{i+1} \prod_{l'} \omega(\tau_l < \tau_{l'}). \tag{4}$$

Thus, the sum can be truncated at all paths requiring i steps with an error that is bounded by $O(a_{i+1})$. As the number of steps, i, increases, the total contribution of each path becomes smaller by a factor involving the ratio of the total contribution to $P(\tau_A < t)$ for the path containing $i+1$ and i steps; i.e.,

$$\frac{a_{i+1}}{a_i} = \frac{\sum_{j=1}^{3} C_j \sum_{\text{branches}} \omega\left[\left(\sum_{\text{jumps},l=1}^{i+1} \tau_l\right) < t\right] \prod_{l=1}^{i+1} \prod_{l'} \omega(\tau_l < \tau_{l'})}{\sum_{j=1}^{3} C_j \sum_{\text{branches}} \omega\left[\left(\sum_{\text{jumps},l=1}^{i} \tau_l\right) < t\right] \prod_{l=1}^{i} \prod_{l'} \omega(\tau_l < \tau_{l'})}. \tag{5}$$

We can simplify this formula by making use of two identities that hold for continuous time Markov chains. For two independent random variables, τ_i and τ_j, that are exponentially distributed with time constants, k_i and k_j, the probability of τ_i being less than $\tau_j \omega(\tau_i < \tau_j)$ is $\omega(\tau_i < \tau_j) = \frac{k_i}{k_i + k_j}$. Also, for a sum of n exponentially distributed random variables $\tau_i; \tau_i \in \{\tau_1, \tau_2, \ldots, \tau_n\}$ with time constants, $k_i; k_i \in \{k_1, k_2, \ldots, k_n\}$, the probability of the sum of n independent random variables being less than t is a convolution of those variables, $\omega[(\Sigma_i^n \tau_i) < t] = \omega(\tau_1 < t) \otimes \cdots \otimes \omega(\tau_n < t)$ where the $\otimes$ symbol denotes a convolution, and has the following form:

$$\omega\left[\left(\sum_{i}^{n} \tau_i\right) < t\right] = 1 - \sum_{i=1}^{n} \left(\prod_{j \neq i} \frac{k_j}{k_j - k_i}\right) e^{-k_i t}.$$

By substituting these two expressions where appropriate,

$$\frac{a_{i+1}}{a_i} = \frac{\sum_{j=1}^{3} C_j \sum_{\text{branches}} \sum_{n=1}^{i+1} \left(1 - \prod_{j \neq n} \frac{k_j}{k_j - k_n} e^{-k_n t}\right) \prod_{l=1}^{i+1} \prod_{m} \frac{k_l}{k_l + k_m}}{\sum_{j=1}^{3} C_j \sum_{\text{branches}} \sum_{n=1}^{i} \left(1 - \prod_{j \neq n} \frac{k_j}{k_j - k_n} e^{-k_n t}\right) \prod_{l=1}^{i} \prod_{m} \frac{k_l}{k_l + k_m}}. \tag{6}$$

This formula can be rearranged by factoring out the $i+1$ term inside the summation over the different "branches"

$$\frac{a_{i+1}}{a_i} = \frac{\sum_{j=1}^{3} C_j \sum_{\text{branches}} \left\{\left[\left(\sum_{n=1}^{i} 1 - \beta_n \gamma_n\right) + \alpha\right] \kappa_{i+1}\right\}}{\sum_{j=1}^{3} C_j \sum_{\text{branches}} \sum_{n=1}^{i} (1 - \gamma_n) \kappa_i}, \tag{7}$$

where $\alpha = \Pi_{j \neq i+1} \frac{k_j}{k_j - k_{i+1}} e^{-k_{i+1} t}$, $\beta_n = \frac{k_{i+1}}{k_{i+1} - k_n}$, $\gamma_n = \Pi_{j \neq i+1} \frac{k_j}{k_j - k_n} e^{-k_n t}$, and $\kappa_i = \Pi_{l=1}^{i} \Pi_m \frac{k_l}{k_l + k_{lm}}$. The error, E, that is introduced by truncating the sum at a given number of steps is

$$E = O\left(\frac{\sum_{j=1}^{3} C_j \sum_{\text{branches}} \left\{\left[\left(\sum_{n=1}^{i} 1 - \beta_n \gamma_n\right) + \alpha\right] \kappa_{i+1}\right\}}{\sum_{j=1}^{3} C_j \sum_{\text{branches}} \sum_{n=1}^{i} (1 - \gamma_n) \kappa_i}\right). \tag{8}$$

From the formula in Eq. (6), we see that many conditions allow for $\frac{a_{i+1}}{a_i} \ll 1$, in which case, the summation quickly decays and can be truncated at i steps. Moreover, any significant difference in time scales for processes in successive steps results in such a decrease.

Now we consider the application of this formalism to the model. Summing Eq. (3) from $j=1$ to $j=3$ gives us

$$P(\tau_A < t) = P(\tau_{A,\text{solution}} < t) + P(\tau_{A,\text{incomplete}} < t) + P(\tau_{A,\text{complete}} < t). \tag{9}$$

Equation (9) gives us the cumulative density as a composition of many terms contributing from initial states; from solution, incomplete complexes, and complete complexes where the sum is carried out up to $i=3$ steps,

$$\begin{aligned} P(\tau_{A,\text{solution}} < t) = C_s\{&[P(\tau_7 < t)P(\tau_7 < \tau_1)P(\tau_7 < \tau_5)] + [P(\tau_5 + \tau_8 < t)P(\tau_5 < \tau_7)P(\tau_5 < \tau_1)P(\tau_8 < \tau_4)P(\tau_8 < \tau_6)] \\ &+ [P(\tau_1 + \tau_2 + \tau_7 < t)P(\tau_1 < \tau_5)P(\tau_1 < \tau_7)P(\tau_2 < \tau_3)P(\tau_7 < \tau_1)P(\tau_7 < \tau_5)] \\ &+ [P(\tau_1 + \tau_3 + \tau_8 < t)P(\tau_1 < \tau_5)P(\tau_1 < \tau_7)P(\tau_3 < \tau_2)P(\tau_8 < \tau_4)P(\tau_8 < \tau_6)]\}, \end{aligned}$$

$$\begin{aligned} P(\tau_{A,\text{incomplete}} < t) = C_I\{&[P(\tau_2 + \tau_7 < t)P(\tau_2 < \tau_3)P(\tau_7 < \tau_1)P(\tau_7 < \tau_5)] + [P(\tau_3 + \tau_8 < t)P(\tau_3 < \tau_2)P(\tau_8 < \tau_4)P(\tau_8 < \tau_6)] \\ &+ [P(\tau_2 + \tau_5 + \tau_8 < t)P(\tau_2 < \tau_3)P(\tau_5 < \tau_1)P(\tau_5 < \tau_7)P(\tau_8 < \tau_4)P(\tau_8 < \tau_6)] \\ &+ [P(\tau_3 + \tau_6 + \tau_7 < t)P(\tau_3 < \tau_2)P(\tau_6 < \tau_4)P(\tau_6 < \tau_8)P(\tau_7 < \tau_1)P(\tau_7 < \tau_5)]\}, \end{aligned}$$

$$\begin{aligned} P(\tau_{A,\text{complete}} < t) = C_C\{&P(\tau_8 < t)P(\tau_8 < \tau_4)P(\tau_8 < \tau_6) + P(\tau_6 + \tau_7 < t)P(\tau_6 < \tau_4)P(\tau_6 < \tau_8)P(\tau_7 < \tau_1)P(\tau_7 < \tau_5) \\ &+ [P(\tau_3 + \tau_4 + \tau_8 < t)P(\tau_4 < \tau_6)P(\tau_4 < \tau_8)P(\tau_3 < \tau_2)P(\tau_8 < \tau_4)P(\tau_8 < \tau_6)] \\ &+ [P(\tau_2 + \tau_4 + \tau_7 < t)P(\tau_4 < \tau_6)P(\tau_4 < \tau_8)P(\tau_2 < \tau_3)P(\tau_7 < \tau_1)P(\tau_7 < \tau_5)]\}. \end{aligned} \tag{10}$$

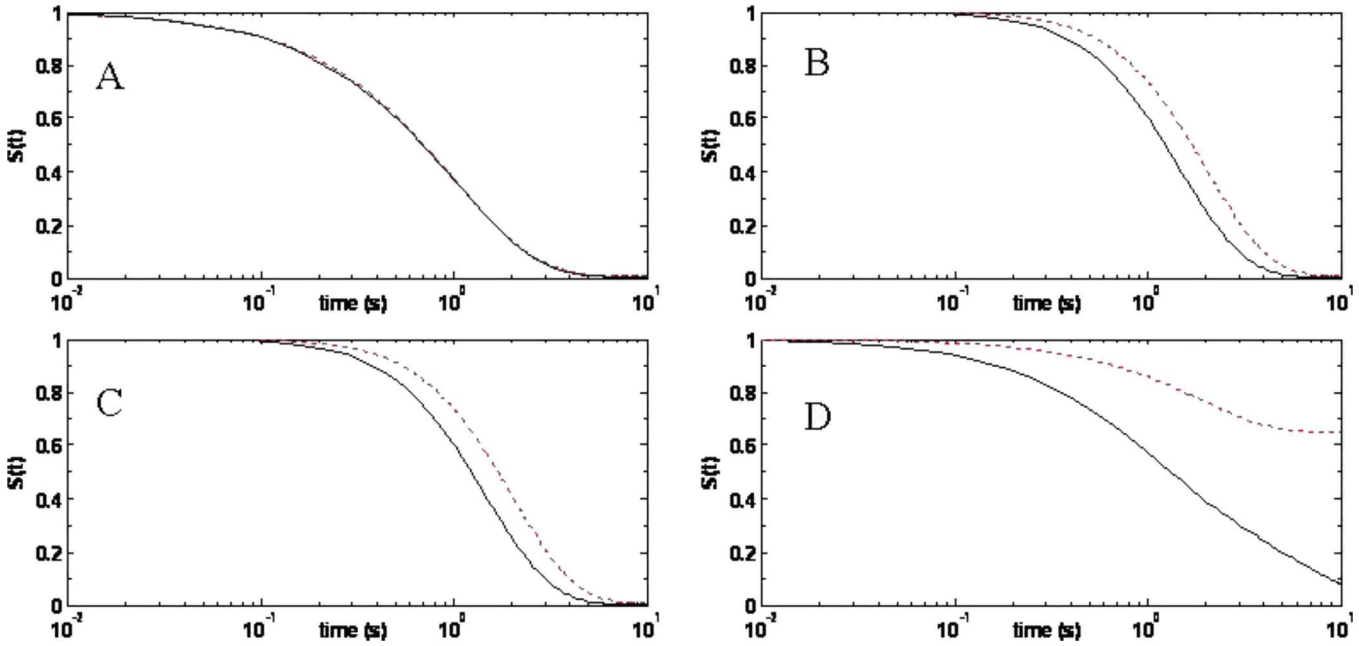

FIG. 2. (Color online) Comparison of the path summation approximation and the exact numerical solution of the three-state model. Exact numerical solutions of Eq. (1) are compared to Eq. (9) which constitutes a truncation of the path summation in Eq. (4) at $i=3$ steps. Exact solutions are shown in solid lines and the approximate solutions in dotted lines. Parameters used are (a) $k_1=0.001$, $k_2=0.002$, $k_3=0.003$, $k_4=0.004$, $k_5=0.005$, $k_6=0.006$, $k_7=k_8=1.0$, $C_S=0.5$, $C_C=0.5$, $C_I=0.0$. (b) $k_1=0.001$, $k_2=1.002$, $k_3=1.003$, $k_4=0.004$, $k_5=0.005$, $k_6=0.006$, $k_7=k_8=1.0$, $C_S=0.0$, $C_C=0.0$, $C_I=1.0$. (c) $k_1=0.101$, $k_2=0.102$, $k_3=0.103$, $k_4=0.104$, $k_5=0.105$, $k_6=0.106$, $k_7=k_8=1.0$, $C_S=0.5$, $C_C=0.5$, $C_I=0.0$. (d) $k_1=1.001$, $k_2=1.002$, $k_3=1.003$, $k_4=1.004$, $k_5=1.005$, $k_6=1.006$, $k_7=k_8=1.0$, $C_S=0.333$, $C_C=0.333$, $C_I=0.333$.

Figure 2 considers Eq. (9) as compared to the exact numerical solution of Eq. (1). In each case, the survival probability $S(t)=1-P(\tau_A<t)$ is plotted. In Fig. 2(a), processes involving activation, τ_7 and τ_8, dominate the kinetic pathway and the approximate solution is quantitatively accurate at all times. In Figs. 2(b) and 2(c), τ_7 and τ_8 are still dominant time scales but other transitions within the pathway also contribute significantly; as a result, Eq. (9) still captures much of the qualitative behavior such as the shapes and overall time to relaxation but quantitative deviations in Eq. (9) and Eq. (2) are readily apparent. Last in Fig. 2(d), each state transition contributes to the arrival at the absorbing state. In this scenario, Eq. (9) provides a poor description of the dynamics in Eq. (2). Combined, the plots in Fig. 2 illustrate the regimes of validity of the use of such a path summation technique in comparison with the exact numerical solution of the kinetic model in Eqs. (1) and (2).

We note that Eq. (9) is not particularly useful in itself. It does however provide a starting point from which simple analytical expressions can be obtained in different limiting cases. In the case of the model of scaffold-mediated signaling, Eq. (9) can be simplified in different regimes of scaffold density.

APPLICATION TO THE DEPENDENCE ON SCAFFOLD DENSITY

Now, consider the case in which scaffold proteins are present in negligible amounts. In this case, a separation of time scales is apparent with one dominant time scale controlling kinase activation. For low scaffold concentrations ($C_S \sim 1$), most kinases are present in solution and the transition to either scaffold-bound state is very slow, i.e., $k_7 \gg k_5$ and $k_7 \gg k_1$, the summation can therefore, with small error, be cut off at one step giving (with h.o. denoting higher order terms)

$$P(\tau_A < t) = 1 - \left(\frac{k_7}{k_7 + k_1}\right)\left(\frac{k_7}{k_7 + k_5}\right)e^{-k_7 t} + \text{h.o.} \approx 1 - e^{-k_7 t}. \tag{11}$$

If signaling is only allowed to take part on a scaffold (i.e., $k_7=0$) then by the same argument,

$$P(\tau_A < t) \approx 1 - e^{-k_5 t}. \tag{12}$$

Now consider the situation where scaffold concentration is very high (i.e., $C_I \approx 1$; in this case, we consider paths starting from K_I that end in the active state. Starting at K_I, there are two branches that lead to K_A, $K_I \xrightarrow{2} K_S \xrightarrow{7} K_A$ and $K_I \xrightarrow{3} K_C \xrightarrow{8} K_A$. However, the dominant time scale in this scenario is k_3^{-1}. Keeping only the branch that involves process 3, since $k_4 \ll k_3$ and $k_8 \ll k_3$ we have

$$P(\tau_A < t) = P(\tau_3 + \tau_8 < t)P(\tau_3 < \tau_4)P(\tau_8 < \tau_4)P(\tau_8 < \tau_6)$$
$$= 1 - \left(\frac{k_3}{k_3 + k_4}\right)\left(\frac{k_8}{k_8 + k_4}\right)\left(\frac{k_8}{k_8 + k_6}\right)$$
$$\times\left(\frac{k_8}{k_8 - k_3}e^{-k_3 t} + \frac{k_3}{k_3 - k_8}e^{-k_8 t}\right) + \text{h.o.} \approx 1$$
$$- \left(\frac{k_3 k_8^2}{k_4(k_8 + k_4)(k_8 + k_6)}\right)e^{-k_3 t}. \quad (13)$$

In the intermediate regime of scaffold density, no such separation of time scales is apparent; also, there is significant probability that a kinase resides in any of the three initial states. Typical equilibrium disassociation constants,K_d, for association to a scaffold are $\sim 1\ \mu$M (1 kT/molecule ~ 0.6 kCal/mol) and this corresponds to a binding energy $F = -k_b T \ln(K_d)$ of 12–15 kT [21]. Given this binding affinity and typical kinase concentrations that result in an excess number of signaling residing in solution than potentially bound to a scaffold [22], $\sim 85\%$ ($P_{\text{scaf}} \approx 0.85$) of proteins are bound to a scaffold. This implies that $(P_{\text{scaf}})^n$, for $n=3$ binding sites, $\sim 72\%$ of the bound signaling proteins exist in fully assembled signaling-competent complexes. In this situation, all pathways are important and the summation in Eq. (3) requires more terms than the truncation at $i=3$ that is contained in Eq. (9). This expression gives us, when decomposed into a contribution from solution, incomplete complexes, and complete complexes, a superposition of many exponential terms. In this case, one can see from the formula that the cumulative distribution is a composition of many characteristic time scales that govern signaling dynamics. We show this by solving Eq. (2) numerically in Fig. 3.

Figure 3 illustrates the behavior of the parametrized concentration dependence on the shapes of the integrated waiting time distributions. The shapes of these curves provide information on how kinase activation is distributed over time and provide a useful measure for the characterization of signaling dynamics and the time scale dependence of kinase activation in the model. From the physical processes occurring in our model, there are multiple ways in which parameters are affected by changes in scaffold concentration. One effect is the alteration of C_I and C_S, the number of proteins that exist in complexes that are incomplete and signaling competent, respectively. The rates of transitions between states corresponding to processes occurring with times τ_1, τ_3, and τ_5 are each affected by the relative scaffold concentrations and thus C_I and C_S.

We first consider a parametrization in which the rates of the transitions to each scaffold containing state are proportional to concentration. In this situation, the rates in the model are weakly (linearly) coupled to the relative scaffold concentration. This parametrization describes the scenario in which the density of scaffolds relative to the concentration enzymes is small and the system is dilute. For this scheme, $k_1 = k_1^0 C_I$, $k_3 = k_3^0 C_S$, and $k_5 = k_5^0 C_S$. With this parametrization, Fig. 3(a) contains plots of $S(t)$, obtained from the numerical solution of Eq. (2). For unbinding transitions, a typical off rate [21] ($\sim 0.1\ \text{s}^{-1}$) is used. Also, the rate of activation of a kinase when completely assembled onton a scaffold is taken

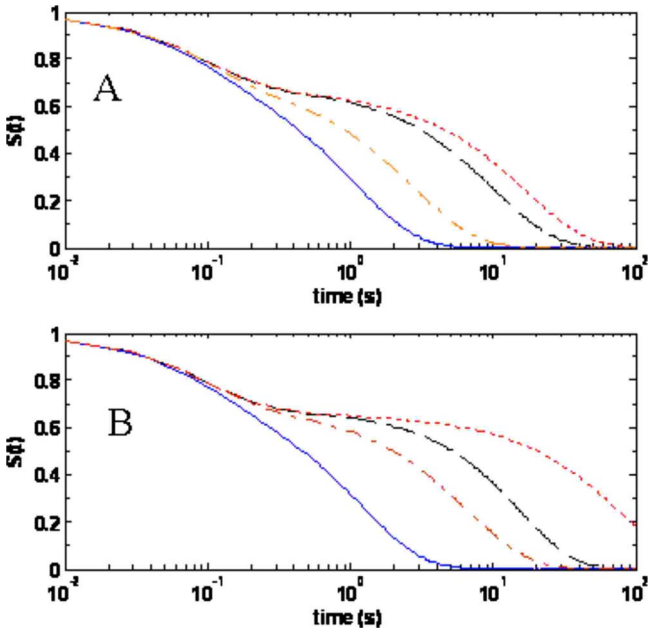

FIG. 3. (Color online) Variations in scaffold concentration. Numerical solutions of Eq. (1) are considered with base parameters: $k_1^0 = 1.001$, $k_2 = 0.102$, $k_3^0 = 1.003$, $k_4 = 0.104$, $k_5^0 = 1.005$, $k_6 = 0.106$, $k_7 = 0.1 k_8 = 10.0$. C_C, C_I, and $C_S = 1 - C_C - C_I$ are varied. (a) Linear ($k_i = k_i^0 C_j$) and (b) quadratic [$k_i = k_i^0 (C_j)^2$], (for $i = 1,3,5$ and $j = S, I$), parametrizations of the effects of the rates by changes in scaffold concentration are used. Four cases are considered: $C_C = 0.05$, $C_I = 0.05$, $C_S = 0.90$ (dashed lines); $C_C = 0.90$, $C_I = 0.05$, $C_S = 0.05$ (solid lines); $C_C = 0.05$, $C_I = 0.90$, $C_S = 0.05$ (dotted lines); $C_C = 0.333$, $C_I = 0.333$, $C_S = 0.333$ (dashed-dotted lines).

to be ~ 100 times greater than that in solution. This effect follows from the elimination of the encounter time between the enzyme and the substrate.

In Fig. 3(a), first consider when kinase activation is primarily occurring on the scaffold ($C_C = 0.05$, $C_I = 0.05$, $C_S = 0.90$, solid line). In this situation, $S(t)$ decays over a time scale that characterizes signal propagation. This time scale results from the time it takes for a kinase in solution to become active. Additional contributions from kinases on a scaffold and those in solution serve to broaden the shape of the distribution away from a single exponential expected from Eq. (10).

Next, consider the case when a significant population of each state is present ($C_C = 0.333$, $C_I = 0.333$, $C_S = 0.333$, dashed-dotted line). In this case, kinases are evenly distributed over each of the inactive states and the decay of $S(t)$ and thus signal propagation extends across multiple decades. The smooth decay of $S(t)$ is interpreted as originating from the many processes that contribute to signal propagation.

In the other cases considered, proteins are confined to incomplete signaling complexes ($C_C = 0.05$, $C_I = 0.90$, $C_S = 0.05$, dotted line) or exist predominantly in solution ($C_C = 0.05$, $C_I = 0.05$, $C_S = 0.90$, dashed line). In these situations, kinase activation occurs predominantly through a two-stage mechanism. The first stage involves the population of kinases that either reside in or those that transition to the fully, assembled (K_C) state and are quickly activated. The second

stage of activation involves the population of kinases that are slowly activated in the solution state.

Alternatively, it is conceivable that the rates, k_1, k_3, and k_5, depend on scaffold concentration in a more complicated manner. Therefore, we also considered a quadratic dependence of the rates on the relative scaffold concentrations. We chose a quadratic dependence because it represents the next simplest way in which the rate constants can depend on scaffold concentration. In this situation, the relative concentration of scaffolds is denser so that nonlinear concentration effects become more important. In Fig. 3(b), we considered the rates to depend on scaffold concentration in a nonlinear, quadratic manner; i.e., $k_1=k_1^0(C_I)^2$, $k_3=k_3^0(C_S)^2$, and $k_5=k_5^0(C_S)^2$. The primary effect as seen in the plots is to broaden the shape of the decay curves of kinase activation in the later stage.

We showed that a multistate kinetic model with Markovian dynamics can give rise to complicated multiexponential kinetics in some range of scaffold concentrations. It is likely, however, that transitions between these states in actuality have more complicated transitions. The minimal model presented, therefore, illustrates the simplest mechanism that illustrates how a waiting time distribution for signal transduction can be affected by scaffold concentration.

SUMMARY

In summary, we illustrated in a simple model, how scaffold concentration affects the competition of the many processes that govern scaffold-mediated signal transduction. The shape of such a distribution is believed to be important in detecting the time scale dependence of biochemical signaling [23,24]. We applied an approximate technique, involving a weighted path summation, along with exact numerical solutions to show how the shape of waiting time distribution of kinase activation and thus the nature of the signal output can depend on scaffold density within the framework of a minimal kinetic model. Since such models occur in many different areas of biology and chemistry, it may be interesting to investigate the behaviors of other models [25,26] in the context of this weighted path summation technique.

Ultimately, the mimimal model makes experimentally testable predictions. Advances in imaging techniques that have been used to study signal propagation allow for the monitoring of kinase activation in more detail than is traditionally allowed for in bulk measurements such as those that involve antibody-based immunoprecipitation techniques. For example, resonance energy transfer techniques with donor and acceptor fluorophores have been used to study signal propagation in live cells [27,28] and may be able to monitor waiting time distributions in detail and also connect how these distributions relate experimentally to other readouts of signal output.

ACKNOWLEDGMENTS

This work is supported in part from NIH Grant No. PO1AI071195-01. I thank Arup Chakraborty for his support. I am also grateful to Fei Liang and Roger York for helpful discussions and comments on this work.

[1] T. Pawson, Cell **116**, 191 (2004).
[2] T. Pawson and J. D. Scott, Trends Biochem. Sci. **30**, 286 (2005).
[3] W. R. Burack and A. S. Shaw, Curr. Opin. Cell Biol. **12**, 211 (2000).
[4] D. Bray, Annu. Rev. Biophys. Biomol. Struct. **27**, 59 (1998).
[5] J. W. Locasale, A. S. Shaw, and A. K. Chakraborty, Proc. Natl. Acad. Sci. U.S.A. **104**, 13307 (2007).
[6] R. L. Kortum *et al.*, Mol. Cell. Biol. **25**, 7592 (2005).
[7] R. L. Kortum and R. E. Lewis, Mol. Cell. Biol. **24**, 4407 (2004).
[8] C. J. Bashor *et al.*, Science **319**, 1539 (2008).
[9] J. W. Locasale and A. K. Chakraborty, PLOS Comput. Biol. **4**, e1000099 (2008).
[10] R. Heinrich, B. G. Neel, and T. A. Rapoport, Mol. Cell **9**, 957 (2002).
[11] L. O. Murphy *et al.*, Nat. Cell Biol. **4**, 556 (2002).
[12] S. D. M. Santos, P. J. Verveer, and P. I. H. Bastiaens, Nat. Cell Biol. **9**, 324 (2007).
[13] T. Lu *et al.*, Proc. Natl. Acad. Sci. U.S.A. **103**, 16752 (2006).
[14] Y. H. Lan and G. A. Papoian, Phys. Rev. Lett. **98**, 228301 (2007).
[15] H. P. Lu, L. Y. Xun, and X. S. Xie, Science **282**, 1877 (1998).
[16] X. S. Xie *et al.*, Annu. Rev. Biophys. Biomol. Struct. **37**, 417 (2008).
[17] N. Dard and M. Peter, BioEssays **28**, 146 (2006).
[18] J. Gunawardena, Biophys. J. **93**, 3828 (2007).
[19] S. C. Kou *et al.*, J. Phys. Chem. B **109**, 19068 (2005).
[20] S. X. Sun, Phys. Rev. Lett. **96**, 210602 (2006).
[21] R. P. Bhattacharyya *et al.*, Science **311**, 822 (2006).
[22] J. E. Ferrell, Trends Biochem. Sci. **21**, 460 (1996).
[23] M. Behar, H. G. Dohlman, and T. C. Elston, Proc. Natl. Acad. Sci. U.S.A. **104**, 16146 (2007).
[24] J. W. Locasale, e-print arXiv:0802.2683.
[25] O. Flomenbom and R. J. Silbey, Proc. Natl. Acad. Sci. U.S.A. **103**, 10907 (2006).
[26] Xiaochuan Xue, Linchen Gong, Fei Liu, and Zhong-can Ou-Yang, Phys. Rev. E **77** 050903(R) (2008).
[27] W. R. Burack and A. S. Shaw, J. Biol. Chem. **280**, 3832 (2005).
[28] B. D. Slaughter, J. W. Schwartz, and R. Li, Proc. Natl. Acad. Sci. U.S.A. **104**, 20320 (2007).