A Time-Based Approach to Stochastic Modeling of Intracellular Signaling Events

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1115/DSCC2012-MOVIC2012-8631">http://dx.doi.org/10.1115/DSCC2012-MOVIC2012-8631</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>ASME International</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Mon Apr 08 08:36:49 EDT 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/118928">http://hdl.handle.net/1721.1/118928</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
A TIME-BASED APPROACH TO STOCHASTIC MODELING OF INTRACELLULAR SIGNALING EVENTS

Michaëlle N. Mayalu
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139
mmayalu@mit.edu

H. Harry Asada
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139
asada@mit.edu

ABSTRACT
This paper presents a modeling framework for an intracellular signaling network based on formalisms derived from the fundamental concepts in probability theory. Cellular behavior is mediated by a network of intracellular protein activations that originate at the membrane in response to stimulation of cell surface receptors. Multiple protein signal transductions occur concurrently through diverse pathways triggered by different extracellular cues. Through crosstalk, these pathways intersect at various node proteins. The state of a particular node protein is dependent on the binding order of molecules from various pathways. The probability of a particular binding order is evaluated using state dependent transduction time probabilities associated with each pathway. In this way, the probability of the cell to be in a given internal state is tracked and used to gain insight into the cell’s phenotypic behavior. A simulation example illustrates the approach. Future work will incorporate the proposed method into the development of a feedback control strategy for the development of an in silico control design of endothelial cell migration during angiogenesis.

INTRODUCTION
Accurate control and manipulation of cellular behaviors has serious implications in the study of biological processes and disease. In particular for in-vitro applications such as tissue engineering and drug screening, it is a fundamental requirement to regulate cells’ phenotypic behaviors by accommodating extracellular cues [1]. However, phenotypic changes may be difficult to detect instantaneously and are usually associated with noticeable delay between input cue and output cellular response. Because of this relying on detection of phenotypic behaviors for use in feedback control may lead to instability and decreased controller performance. In order to alleviate these issues, the disciplines of control theory and systems biology are integrated in order to model and control the internal state of the cell in response to extracellular cues. In addition, following this approach allows insight to be gained into the planned mode of phenotypic response before it occurs.

Cellular response is regulated by the transfer of information from the environment to within the cell. This transfer is realized through a complex intracellular network of protein signal transduction pathways. A transduction pathway consists of a cascade of protein activation reactions triggered at the membrane in response to stimulation of cell surface receptors. Multiple protein signal transductions occur concurrently though diverse pathways triggered by different extracellular cues. Through crosstalk, pathways intersect at various node proteins. Signals from various pathways are synthesized at these nodes. The state of a particular node protein is dependent on the order to which molecules from various pathways bind to it [2].

This chronological order of signaling events often causes a significant difference in the succeeding signaling transductions, and thereby the cell exhibits diverse responses to the combination of cues [2]. However, as the size and complexity of a network grows, it becomes more difficult to decipher the effect of various signaling events. Our ultimate aim is to reduce a complex network through use of transduction time probabilities that may be evaluated for each node to determine the state of the system.

To determine the order that multiple pathways bind to a node protein we must consider the signal transduction time of each pathway. Events involved in signal transduction including, receptor activation, protein modification reactions, and protein inter-compartmental transports are intrinsically stochastic and
occur at random times. In a given time span, we define the transduction of a signal as the first instance after receptor activation that the last protein in a cascaded pathway is activated. We define signal transduction time as the time it takes for a signal (or sequence of cascaded activations) to travel from the receptor to the last molecule in the cascade.

Using transduction time probabilities, we derive a method to calculate the probability that the reaction cascades of each pathway occur in a desired chronological order. In this way, we may track the probability of the cell to be in a given internal state and use this probability to gain insight into the cell’s phenotypic behavior. Under the appropriate assumptions, exact distribution can be found to evaluate uncertainties. Although the modeling approach has not yet been experimentally verified, this study may serve to stimulate future experimental work.

BACKGROUND

Basics of Signal Transduction

Cellular behavior is mediated by a network of intracellular protein activations that originate at the membrane in response to stimulation of cell surface receptors. In a simple linear signaling cascade, stimulation of a receptor leads to consecutive activation of several downstream protein kinases. Each protein undergoes a covalent modification cycle in which the protein can transition between an active and inactive state (see Fig 1). Excluding the first protein, activation of the i-th protein (P_i) is triggered by an enzymatic reaction with the previous activated protein (P^*_{i,j}). The activated protein (P^*_{i,j}) is inactivated by a second reaction catalyzed by enzyme protein (E_i). The last activated protein in the pathway may then interact with node molecules intersecting multiple pathways.

With the exception of the first activated protein in the pathway, the modification cycle of the i-th protein can be described by the following enzymatic reactions in table 1. Here, P^*_{i,j}:P_i \rightarrow P^*_{i,j+1}:P_j \rightarrow P^*_{i,j+2}:P_i \rightarrow \cdots \rightarrow P^*_{i,k}:P_k \rightarrow P_{i-1}, and \alpha, \beta, \lambda are reaction constants. Real signal transduction pathways are generally more complex due positive and negative feedback and signaling amplification.

**TABLE 1: ENZYMATIC REACTIONS DESCRIBING THE COVALENT MODIFICATION CYCLE.**

<table>
<thead>
<tr>
<th>Activation</th>
<th>Deactivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>P^<em>_{i,j}+P_i \xrightarrow{\alpha} P^</em><em>{i,j+1}:P_j \rightarrow P^*</em>{i,j+2}:P_i</td>
<td>E_i+P^<em>_{k} \xrightarrow{\beta} E_i: P^</em>_{k+1} \rightarrow P_i+E_i</td>
</tr>
</tbody>
</table>

Effect of Signaling Order on Cellular Response

Cellular response to environmental cues is mediated by intracellular transduction pathways. These pathways consist of complex biochemical networks that have a specific design and structure. Signaling molecules having multiple binding domains often exhibit complex phenotypic behaviors. Depending on which domain binds first the behavior is strikingly different [2].

Figure 2 illustrates a signal transduction network that characterizes the key signaling pathways during endothelial cell migration [3]. As can be seen, pathways originating at specific membrane receptors (VEGFR-2 and \alpha, \beta integrin) interact to form an intricate network. In the figure, both pathways reach the same node molecule (FAK) which has multiple binding domains. The binding of VEGF to VEGFR-2 triggers the recruitment of HSP90 to VEGFR-2, which initiates the activation of RhoA-ROCK and then the phosphorylation of FAK on its Ser732 domain [3]. This changes the configuration of FAK, allowing subsequent the phosphorylation of its Tyr407 domain by Pyk, a molecule in the \alpha, \beta integrin pathway. This specific binding order leads to focal adhesion turnover response by the cell and ultimately endothelial cell migration.

As can be seen, the binding order of VEGFR-2 pathway protein (RhoA-ROCK) and the \alpha, \beta integrin pathway protein...
\[ a_r(\mathbf{Y}(t)) \cdot dt = h_r c_r \cdot dt \]  

where \( a_r(\mathbf{Y}(t)) \) is the propensity function for the \( r \)-th reaction, \( c_r \) is a constant which depends only on the temperature and physical properties of the system, and \( h_r \) is the number of distinct \( R_r \) molecular reactant combinations available in state \( \mathbf{Y}(t) = [X^A(t), X^B(t)] \).

Consider the diagram shown in Fig. 3. The specific reaction time in each pathway (termed occurrence times) are notated as \( t_1, \ldots, t_n \). The time intervals between reactions (when no reaction is occurring) are notated as \( \tau_1, \ldots, \tau_n \) which we have termed delay times. Occurrence times and delay times may be related in the following manner (see Fig. 3):

\[
\begin{align*}
\tau_1 &= t_1 - t_0 \\
\tau_2 &= t_2 - t_1 \\
& \vdots \\
\tau_n &= t_n - t_{n-1}
\end{align*}
\]

Let:

\[
a_r(\mathbf{Y}(t)) = \sum_{A} a_r(X^A(t)) + \sum_{B} a_r(X^B(t))
\]

be the total propensity over all pathways and reactions in state \( \mathbf{Y}(t) = [X^A(t), X^B(t)] \). It has been shown under well-mixed conditions the delay times \( \tau_1, \ldots, \tau_n \) follow an exponential distribution with varying rate parameter \( a_0(\mathbf{Y}(t)) \) that is independent of reaction type given state \( \mathbf{Y}(t) \) [4, 5]. The densities to describe the delay times \( \tau_1, \ldots, \tau_n \):

\[
\begin{align*}
f_{\tau_1}(\tau_1) &= a_0(\mathbf{Y}(t_0)) \exp(-a_0(\mathbf{Y}(t_0)) \cdot \tau_1) \\
f_{\tau_2}(\tau_2) &= a_0(\mathbf{Y}(t_1)) \exp(-a_0(\mathbf{Y}(t_1)) \cdot \tau_2) \\
& \vdots \\
f_{\tau_n}(\tau_n) &= a_0(\mathbf{Y}(t_{n-1})) \exp(-a_0(\mathbf{Y}(t_{n-1})) \cdot \tau_n)
\end{align*}
\]

Assuming the non-overlapping delay intervals are independent, we may use delay time distribution functions to derive occurrence time distributions \( f_{\tau}(t_n) \).

\[
f_{\tau}(t_n) = (f_{\tau_1} \ast f_{\tau_2}) \ast \cdots \ast f_{\tau_n}(t_n) = \int_{0}^{t_n} f_{\tau_1}(t_n - \tau_{n-1}) f_{\tau_{n-1}}(\tau_{n-1}) d\tau_{n-1}
\]

Here we have used induction and the fact that the distribution describing the sum of two independent random variables is the convolution between the distribution functions of the two random variables.
We may re-write (6) using the Laplace transform:

\[
L[f(t)] = F(s) = \int_0^\infty e^{-st} f(t)dt
\]

\[
F_{\tau_i}(s) = F_{\tau_{\text{path}}(s)} \cdot F_{\tau_{\text{path}}(s)} = \prod_{i=1}^n \frac{a_i(Y(t))}{[x + a_i(Y(t))]} \quad \text{(6)}
\]

Here, the convolution reduces to a multiplication in the Laplace domain. Using partial fractions:

\[
F_{\tau_i}(s) = \left(\prod_{i=1}^n a_i(Y(t))\right) \sum_{i=1}^n \frac{A_i}{[x + a_i(Y(t))]} \quad \text{(7)}
\]

where, \(A_i = \frac{1}{a_i(Y(t)) - a_i(Y(t))}\).

Taking the inverse Laplace \(f_{\tau_i}(t) = \) :

\[
L^{-1}\{F_{\tau_i}(s)\} = \left(\prod_{i=1}^n a_i(Y(t))\right) \sum_{i=1}^n \frac{\exp[-a_i(Y(t)) \cdot t_i]}{a_i(Y(t)) - a_i(Y(t))} \quad \text{(8)}
\]

The equation (9) above is valid for the calculation of distribution reaction occurrence times when the total propensity \(a_i(Y(t))\) parameter is distinct and does not repeat over time. In many cases the total propensity of a reaction may repeat over time. In this situation a more general solution may be derived [6] or we may resort to numerical methods such as Markov chains or stochastic simulation algorithms.

PROBLEM STATEMENT AND METHODS

**Problem Statement**

Computationally, we wish to accomplish the following:

1. Calculate the distributions of signal transduction times of multiple pathways.
2. Use these distributions to determine the probability of a specific transduction order.

**Methods**

To calculate the distribution of signal transduction time, we enumerate all valid possible reaction sequences between the initial state \(Y_0 = [X_0, X_0']\) and the first occurrence of the activation reaction of the last cascade protein in time span \((t_0 - t_f)\) (see Fig. 3). By valid we mean reactions that are chemically feasible given the cellular state. Given the complexity of a network, enumeration of these sequences may be computationally difficult. But under specific assumptions and initial conditions and for lower order networks the problem is computationally tractable.

If the possible reaction sequences are enumerated, the corresponding states \(Y(t)\) \((t=1,\ldots,n-1)\) may also be predetermined. Using this knowledge and given that the propensity \(a_i(Y(t))\) is distinct in the set timespan, we may calculate occurrence time distributions for each possible reaction sequence using (9):

\[
f_{\tau_i}(t = t_i \vert S_i = R_i, \ldots, R_n) = \left(\prod_{i=1}^n a_i(Y(t))\right) \sum_{i=1}^n \frac{\exp[-a_i(Y(t_i)) \cdot t_i]}{a_i(Y(t)) - a_i(Y(t))} \quad \text{(9)}
\]

Where \(R_i\) is the first activation reaction of the last cascade protein in time span \((t_0 - t_f)\).

To find the marginal distribution we use a combination of Bayes rule and the Total Probability Theorem:

\[
f_{\tau_i}(t_s) = \sum_{i} f_{\tau_i}(t_s \vert S_i) f_i(S_i) \quad \text{(10)}
\]

Where \(f_i(S_i)\) is the probability that the reactions will occur in order \(S_i = R_i, \ldots, R_n\). We may find the propensity of each reaction given the predetermined states to calculate this probability:

\[
f_i(S_i) = \prod_{i} a_i(Y(t)) \quad \text{(11)}
\]
Here we have used the fact that reactions are independent given the state and therefore the joint probability of a reaction sequence is the product of the separate reactions’ propensities. Given signal transduction probabilities of multiple independent pathways: \( f_1(t_{v_1}), f_2(t_{v_2}), \ldots, f_n(t_{v_n}) \) we may calculate the distribution of a specific transduction order:

\[
f_r(t) = \sum f(t[S|S]) = \prod_{v_i} f(t_{v_i}) dt_{v_i}
\]

SIMULATION EXAMPLE

Let us examine a simplified model of the VEGFR-2 and \( \alpha \beta \) integrin pathways described in Fig. 2. Each pathway contains a one stage cascade made up of the RhoA-ROCK activation cycle (for VEGFR-2 pathway) or the Pyk activation cycle (for the \( \alpha \beta \) integrin pathway). Parameters and initial conditions were obtained from the literature (see Table 2). For reactions where the parameters were unavailable, they were assumed to be in the same range as parameters for similar components.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Parameters</th>
<th>Initial Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>VEGFR2, ( \alpha \beta )</td>
<td>( c_1 = 8.4 \times 10^4 \text{M}^{-1} \cdot \text{s}^{-1} / \text{Vol} \cdot \text{N}_A )</td>
<td>VEGFR2 = 18 molecules ( \mu \text{m}^{-2} )</td>
</tr>
<tr>
<td></td>
<td>( c_2 = 1.3 \times 10^3 \text{cm}^2 \cdot \text{mol} / \text{S.A.} \cdot \text{N}_A )</td>
<td>( \alpha \beta = 36 \text{molecules} \mu \text{m}^{-2} )</td>
</tr>
<tr>
<td>RhoA-ROCK, Pyk</td>
<td>( c_4 = 1 \mu \text{M} \cdot \text{s}^{-1} / \text{Vol} \cdot \text{N}_A )</td>
<td>RhoA−ROCK = 0.1 \mu \text{M}</td>
</tr>
<tr>
<td></td>
<td>( c_3 = 5.7 \text{x}^{-1} / c_4 = 1 \text{s}^{-1} )</td>
<td>Pyk = 0.1 \mu \text{M}</td>
</tr>
</tbody>
</table>

TABLE 2: PARAMETERS AND I.C.’S. NA IS AVAGADROS #, VOL. IS CELL VOLUME, S.A IS CELL SURFACE AREA

The left graph of Figure 4 illustrates the marginal distribution \( f_r(t) \) of the signal transduction (or first occurrence time of RhoA-ROCK activation) in the VEGFR2 pathway for various concentrations of extracellular input VEGF. As can be seen, the transduction time decreases with increased input concentration while the probability increases with increased input concentration. Given 2 reaction cascades occurring independently, we may compute the probability that signal transduction of the VEGFR2 cascade is faster than that of the Integrin cascade. The right graph of figure 4 represents \( \text{Pr}(t_{\text{ROCK}} < t_{\text{Pyk}}) \) for varying input concentrations. As can be seen, the highest probability occurs when VEGFR2 input (VEGF) and integrin input (Fibronectin) are 100pM and 1pM respectively. The lowest probability occurs when Fibronectin is 100pM and VEGF 1pM.

CONCLUSION

A dynamic model of a signaling network was developed based on distributions associated with individual and independent pathways. These distributions where calculated exactly and in addition the probability of a particular order was discussed and defined. The approach was demonstrated with a simulation example. The author’s group is currently working on application of the approach for more complex networks.

ACKNOWLEDGEMENTS

This material is based on work supported in part by the Emerging Frontiers in Research and Innovation Program of the National Science Foundation under grant number NSF EFRI-0735997, and the EBICS Science and Technology Center under grant number NSF STC-0902396. The work is also supported by the Singapore-MIT Alliance for Research and Technology, Bio-Systems and Micromechanics IRG.

REFERENCES