Stochastic Modeling of Intracellular Signaling Dynamics for the Purpose of Regulating Endothelial Cell Migration

by

Michaëlle Ntala Mayalu

B.S., Mechanical Engineering, Massachusetts Institute of Technology, 2010

Submitted to the Department of Mechanical Engineering in Partial Fulfillment of Requirements for the Degree of

Master of Science in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2012

Signature of Author.....
Department of Mechanical Engineering
May 11, 2012

Certified by.....
H. Harry Asada
Ford Professor Mechanical Engineering
Thesis Supervisor

Accepted by.....
David E. Hardt
Chairman, Department Committee on Graduate Students

Stochastic Modeling of Intracellular Signaling Dynamics for the Purpose of Regulating Endothelial Cell Migration

by

Michaëlle Ntala Mayalu

Submitted to the Department of Mechanical Engineering
on May 11, 2012 in partial fulfillment of the requirements of the Degree of
Masters of Science in Mechanical Engineering

Abstract

Effective control of cellular behaviors has serious implications in the study of biological processes and disease. 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, a new approach to regulating cell behaviors through control of intracellular signaling events is presented.

Many cell behaviors are mediated by a network of intracellular protein activations that originate at the membrane in response to stimulation of cell surface receptors. Multiple protein signaling transductions occur concurrently through diverse pathways triggered by different extracellular cues. Cell behavior differs, depending on the chronological order of multiple signaling events.

This thesis develops several modeling frameworks for an intracellular signaling network specific to endothelial cell migration in angiogenesis. Unlike previous works, the models developed in this thesis exploit the effect of signaling order on extracellular response. Our approach examines the transduction time associated with each pathway of a cascaded signaling network. Transduction times of multiple pathways are compared, and the probability that the multiple signaling events occur in a desired chronological order is evaluated.

We begin our development with an input-output time-delay model derived from simulated data that is used to predict the optimal extracellular input intensity for a desired response. We then present a stochastic "pseudo-discrete" model of the signal transduction time. We conclude by presenting several control strategies for control of intracellular signaling events.

Thesis Supervisor: H. Harry Asada

Title: Ford Professor Mechanical Engineering

Contents

List of Figures

List of Tables

1. Introduction

1.1 Thesis Objectives

Using signaling dynamics for the purpose of controlling cellular response requires identification and manipulation of intracellular signals. Consequently, an appropriate model is needed to describe the biological system. The purpose of this thesis is to model internal protein signaling dynamics within the cell and use the model to predict extracellular response to various external inputs and cues. The ultimate aim is to verify model predictions with experimental data and create a model structure that is usable for control. Some desired model characteristics include a "condensed" modeling framework with few parameters that capture key biological mechanisms. In addition, we wish to create a model that incorporates heterogeneity among cells and the inherent stochasticity of intermolecular interactions.

1.2 The Angiogenic Process

1.2.1. Definition

In order to model the intercellular signaling process that leads to endothelial cell migration during angiogenesis, it is important to understand the key mechanisms in the angiogenic process. Angiogenesis is a cell migration dependent process involving the movement of endothelial cells to form blood vessels. It is present in physiological situations such as tissue repair and skin renewal as well as in pathological situations such as cancer, arthritis, and vascular disease. Migration of endothelial cells is driven by a growth factor gradient such as VEGF, which has been known to induce angiogenesis [1].

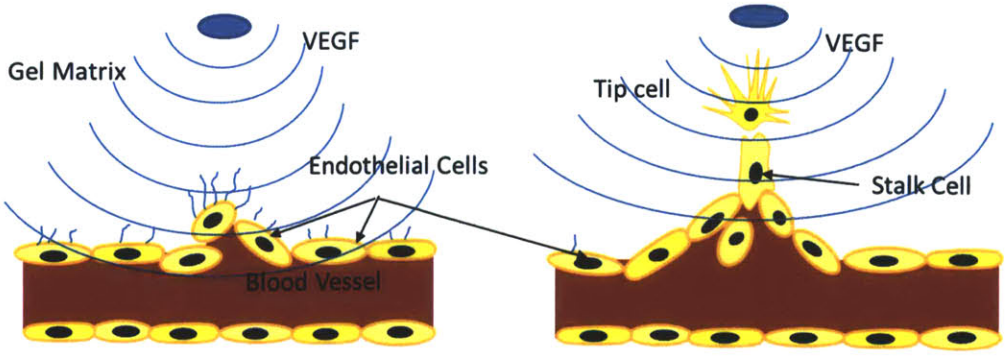

Figure 1-1: The Angiogenic Process. Endothelial cells migrate to form blood vessels in response to a biochemical growth factor (VEGF).

In cancer, tumors secrete VEGF in order to initiate blood vessel formation. The blood vessels grow towards the concentration of VEGF and supply nutrients [2].

1.2.2. Angiogenic Growth Assays

It has been shown that the angiogenic process may be manipulated in a 3D in-vitro microfluidic environment by varying environmental factors and external cues such as growth factor concentration [3]. As described in [3], angiogenic response can be regulated using a microfluidic device consisting of 2 channels separated by a central region containing extracellular gel matrix. By varying the growth factor concentration in the channels, a concentration gradient is established across the gel region to stimulate cellular responses (see Figure 1-2).

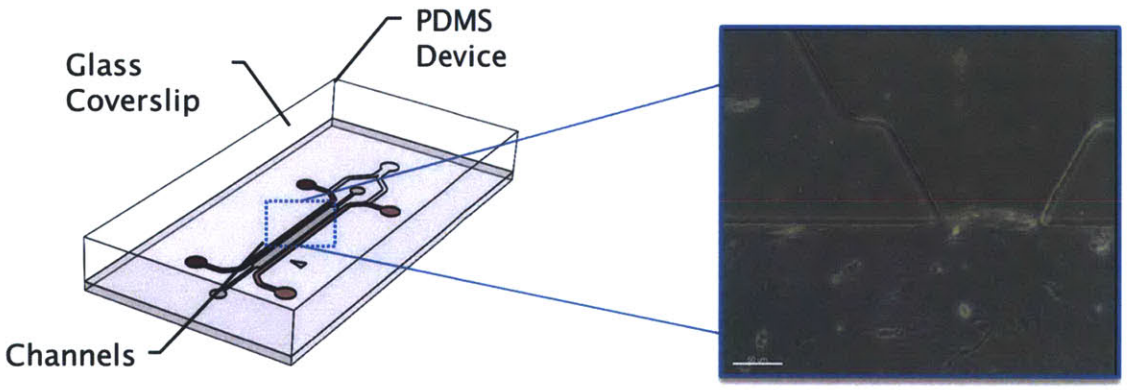

Figure 1-2: Microfluidic device used to monitor endothelial cell migration. Source: Wood, L. B. *Quantitative modeling and control of nascent spout geometry in in-vitro angiogenesis.* Unpublished Doctoral Thesis, Massachusetts Institute of Technology, 2012, Cambridge, Ma.

Using this technology we may implement a control system in hardware consisting of a confocal microscope to measure overall vessel morphology. This may then be feedback and compared to a reference or desired morphology. A control law may then be designed to calculate the optimal growth factor concentrations for the microfluidic device to drive the output vessel. morphology to the desired vessel morphology(see figure 1-3).

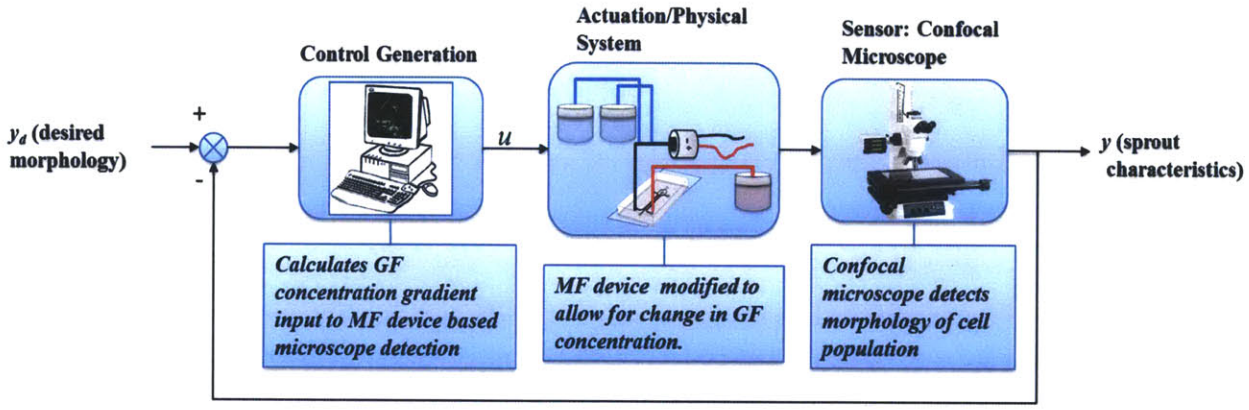

Figure 1-3: Diagram describing the control of angiogenesis using cell population characteristics for feedback

However, morphological changes are 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 morphological characteristics alone may lead to instability and decreased performance in feedback control. In order to alleviate these issues, intracellular signaling dynamics are observed and modeled in response to extracellular cues.

1.3 Signal Transduction Process

1.3.1. Definition

Cellular response is regulated by the transfer of information from the environment to within the cell. This transfer is realized through a complex network of protein signaling transduction pathways. Multiple protein signaling transductions occur concurrently through diverse pathways triggered by different extracellular cues. Signals from various pathways are synthesized to determine the overall cellular response.

Signal transduction is initiated when a ligand binds to a receptor present on the cell membrane. The receptor then underdoes a change (conformational or polymerization with other trans-membrane proteins) to become active. The active receptor then activates other proteins present inside the cell which in turn activates other pr[4](Bauer, Jackson, Jiang, & Rohlf, 2010)oteins downstream. This leads to a cascade which travels to the nucleus where a specific function is performed.

Fig. 1-4 illustrates a signal transduction network that characterizes some of the key signaling pathways during angiogenesis[4]. Arrows between nodes signify activation and hammerheads signify inhibition. As can be seen, pathways originating at specific membrane receptors interact to form an intricate network. In addition, the network indicates that varying activation/inhibition

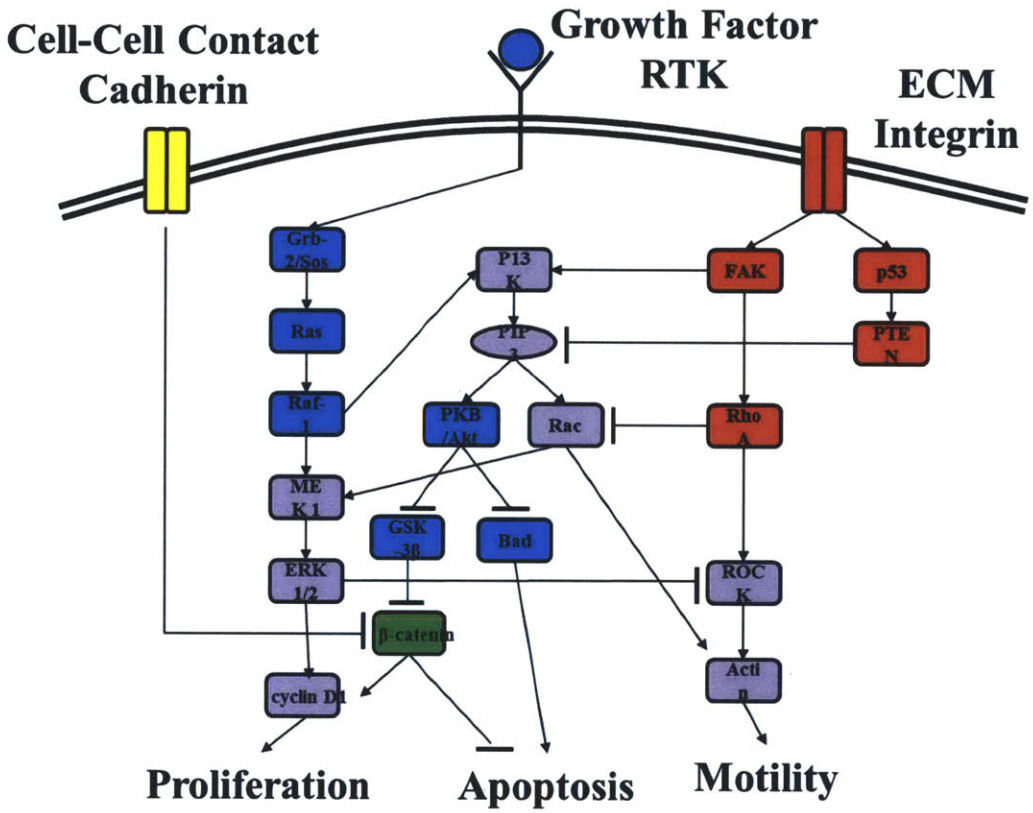

Figure 1-4: Reduced signal transduction network specific to angiogenesis. Illustrates the complex interplay between growth factor, integrin, and cadherin stimulated pathways to decide the cell's phenotypic response. Colors associate downstream molecules with the initiating receptor: blue is the GF-RTK pathway, red the ITG pathway, yellow is the cadherin pathway, and purple and green signify molecule influenced by multiple pathways. An arrow signifies activation and a hammerhead indicates inhibition. Source: Bauer, A. L., Jackson, T. L., Jiang, Y., & Rohlf, T. (2010). Receptor cross-talk in angiogenesis: Mapping environmental cues to cell phenotype using a stochastic, boolean signaling network model. *Journal of Theoretical Biology, 264*(3), 838-846.

sequences lead to a different overall phenotypic response.

1.3.2. Activation mechanisms

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 1-5). 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-1}). The activated protein (P^*_i) is inactivated by a second reaction catalyzed by enzyme protein (E_i). The last activated protein in the pathway may then interact molecules from different pathways.

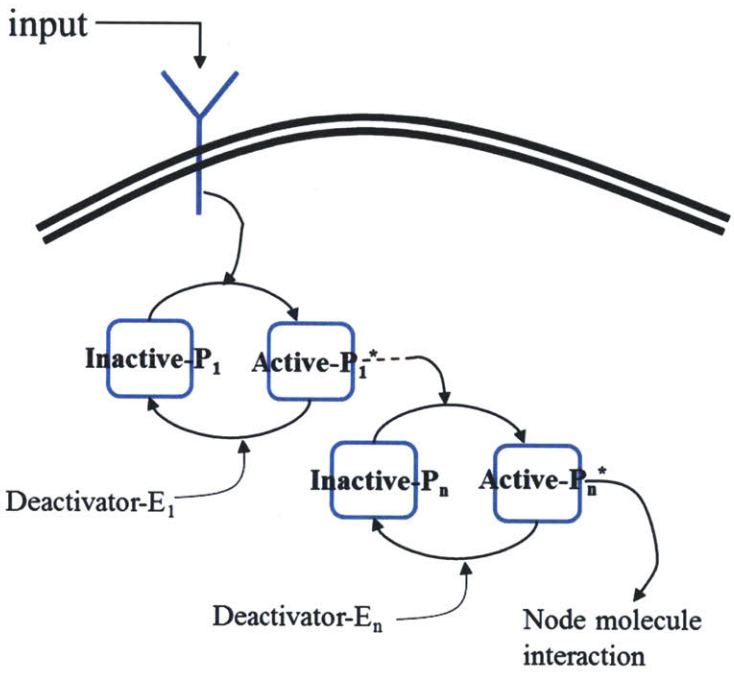

Figure 1-5: Simple Network showing linear covalent modification cycles for each a given pathway

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 1-1. Here, $P^*_{i-1}{:}P_i$, and $E_i{:}P^*_i$ represent the intermediate complexes and α, β, λ are reaction constants. Real signal transduction pathways are generally more complex due positive and negative feedback and signaling amplification.

Table 1-1: Enzymatic reactions describing the covalent modification cycle.

Reaction	Chemical equation
Activation	$P^*_{i-1} + P_i \underset{\beta i}{\overset{\alpha_i}{\rightleftharpoons}} P^*_{i-1} : P_i \xrightarrow{\lambda_k} P^*_i + P^*_{i-1}$
Deactivation	$E_i + P^*_i \underset{\beta i}{\overset{\bar{\alpha}_i}{\rightleftharpoons}} E_i : P^*_i \xrightarrow{\bar{\lambda} i} P_i + E_i$

1.3.3. Measurement techniques

Recent breakthroughs in imaging are providing us with the tools needed for building a signaling dynamics model. Among others, Fluorescence Resonance Energy Transfer (FRET) imaging technology allows activation of targeted signaling molecules to be observed in a live cell in real time. Fluorescence Resonance Energy Transfer or FRET is a distance-dependent interaction between two fluorescent molecules or fluorophores in which emission of the first fluorophores (donor) is used to excite the 2nd fluorophores (acceptor).The efficiency of FRET is dependent on the inverse sixth power of the separation, making it useful technique in the study of biological occurrences that produce changes in molecular proximity. One experimental approach engineers both fluorophores into a single molecule which undergoes a conformational change upon activation or interaction with a specific ligand. The fluorophores are integrated into the protein so that the conformational change brings about a change in proximity or orientation of the fluorophores that alters FRET efficiency (see figure 1-6).

Using FRET, a unique signal can be detected that reveals the location and concentration of an activated species [5]. Limitations to FRET measurement are that only one bio-sensor can be used at a time. Furthermore, the number of experimental data sets is limited. Recently, a biosensor for the key signaling molecule Src has been developed and experimental data has been obtained [6]. The molecular dynamics resemble an impulse response to a second order or higher system.

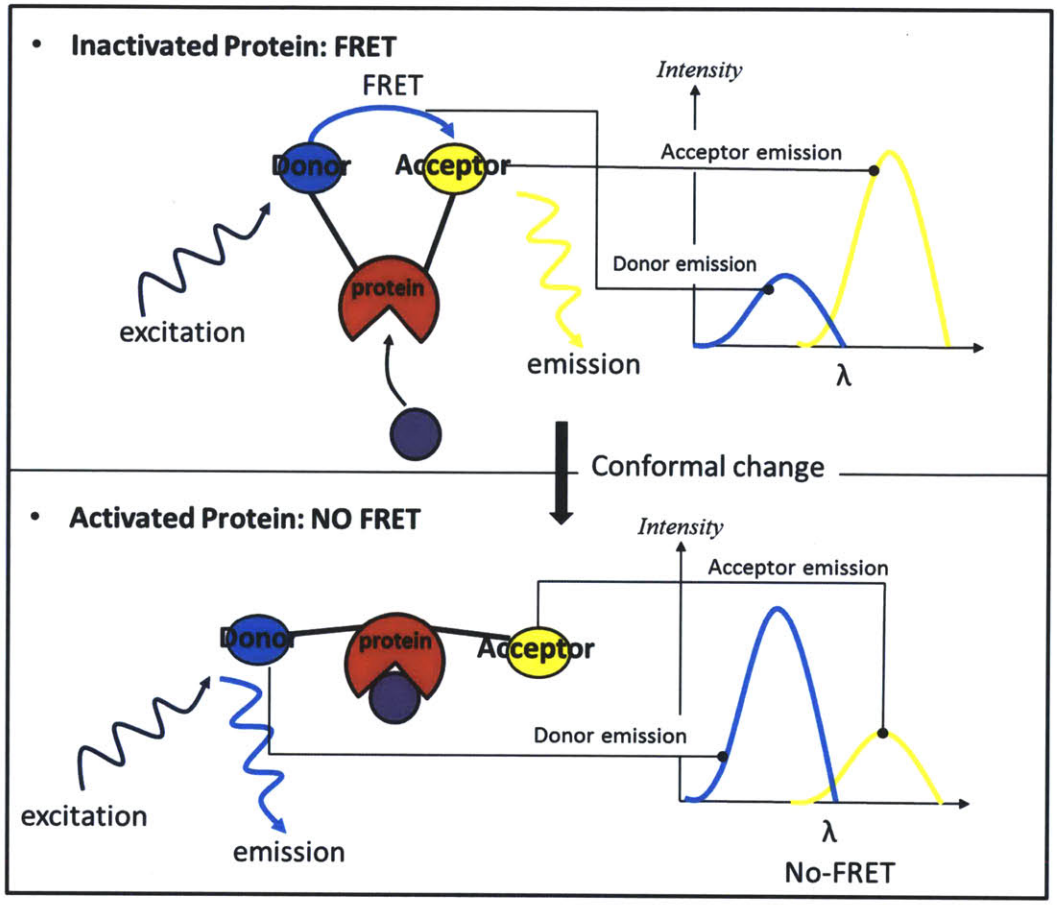

Figure 1-6:Diagram describing the Fluorescence Resonance Energy Transfer Process

Using this technique, has demonstrated that varying external cues may cause different signaling hierarchies, which lead to different modes of response [7]. This suggests a correlation between the sequence of protein activations and mode of phenotypic response.

Since controlling the migratory state of endothelial cells in angiogenesis may be difficult

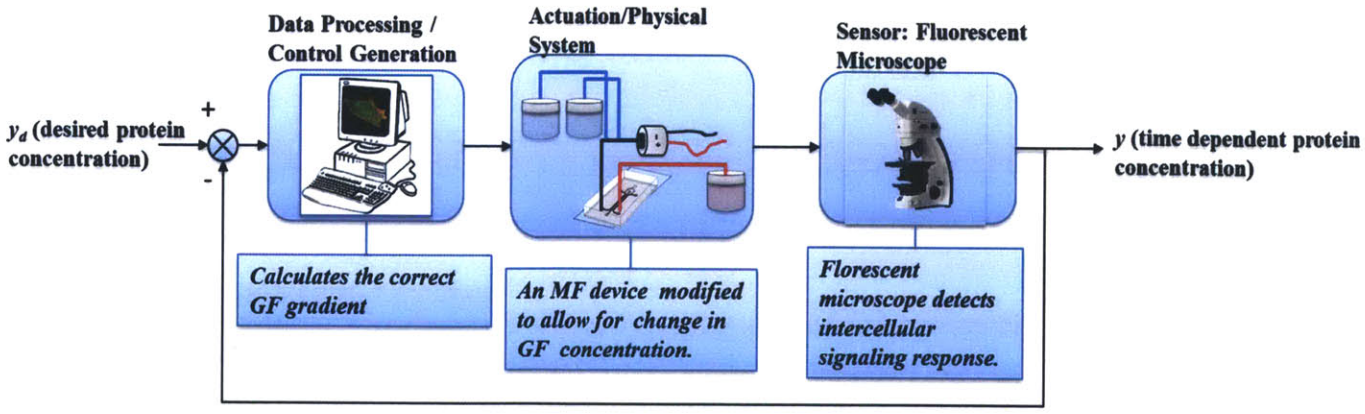

Figure 1-7: Diagram describing the control of angiogenesis using signaling dynamics for feedback

based on its output morphology alone, future control designs may incorporate the use of FRET biosensors to measure activation of proteins in migration specific pathways. The measurements obtained through FRET may then be used for feedback control to the microfluidic device [8].

1.4 Effect of Signaling Order on Cellular Response

Signaling molecules having multiple binding domains often exhibit complex phenotypic behaviors. Depending on which domain binds first the behavior is strikingly different. Figure 1-8 illustrates a signal transduction network that characterizes the key signaling pathways during endothelial cell migration [1]. As can be seen, pathways originating at specific membrane receptors (VEGFR-2 and 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 [1].

This changes the configuration of FAK, allowing subsequent the phosphorylation of its Tyr407 domain by Pyk, a molecule in the 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 integrin pathway protein (Pyk) is essential to the function of FAK since the phosphorylation of its Ser732 domain by RhoA-ROCK triggers a conformational change that allows phosphorylation of its Tyr407 domain by Pyk. If Pyk attempts to bind before FAK has been phosphorylated by RhoA-ROCK the cell will not exhibit the same phenotypic behavior.

Dependence of the node molecule on the order of signaling events is ubiquitous occurrence in the cell and thereby the cell exhibits diverse responses.

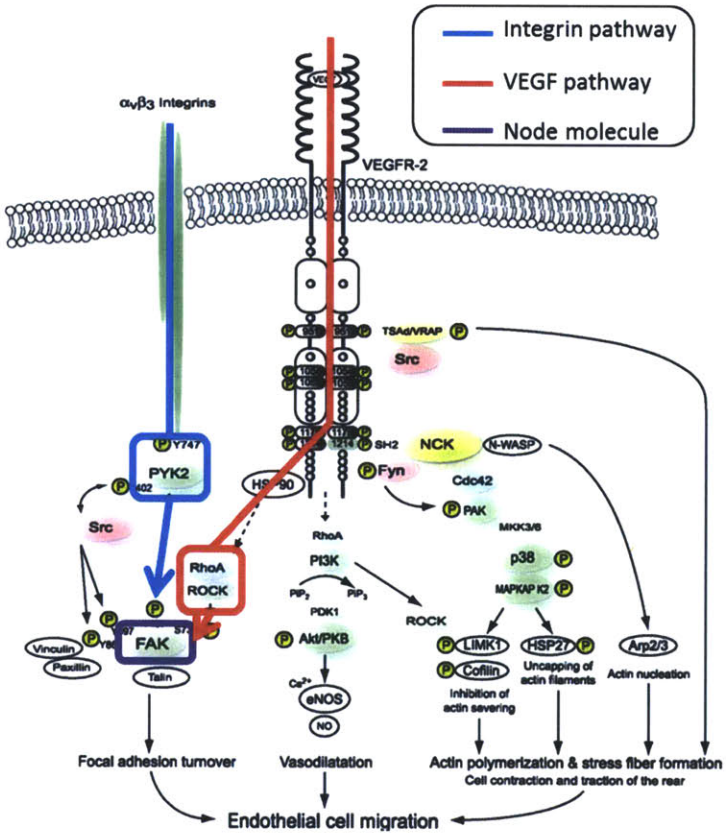

Figure 1-8 signal transduction network specific to Endothelial Cell Migration Source: Lamalice, L., Le Boeuf, F., & Huot, J. (2007). Endothelial cell migration during angiogenesis. *Circulation Research, 100*(6), 782-794.

1.5 Review of Intracellular Signaling Pathway Modeling

Detailed computational models have been developed for various signal transductions. These include modeling of the Mitogen-activated protein kinase (MAPK) cascade in response to growth factor stimuli [9],[10],[11],[12]. These computational models simulate dynamic responses of signaling activities for known signaling pathways. However, in many cases, quantitative information about the kinetic properties in the system is unknown. Therefore, experimental data is used along with the model to obtain more accurate representations.

One way to model the transient response for each signaling pathway is to create a continuous system model. For this modeling framework the state of the system at any particular instant is regarded as a vector containing the concentration of each molecule. Continuous state changes are defined by a sum of rates describing the increase and/or decrease of the molecular

concentrations. The changes in concentration are assumed to occur by a continuous and deterministic process. Numerical integration algorithms are used to generate simulated data from the derived equations [13].

Due to the heterogeneity of cells, the response of the cell will vary given the same environmental conditions. Furthermore, events involved in signal transduction including, receptor activation, protein modification reactions, and protein inter-compartmental transports are intrinsically stochastic and occur at random times. The deterministic approach to modeling fails to capture heterogeneity and variability among cells. The multiple sources of stochasticity and heterogeneity in biological systems have important consequences to understanding system behavior [14]. In order to account for these uncertainties we turn to modeling based on probability theory. Stochastic modeling uses probability theory to account for the inherent uncertainty in biochemical network.

Signaling dynamics may be modeled stochastically using the Markov jump process. This is a well-understood model from the theory of stochastic processes. In this modeling framework, the change in the system occurs discretely after random time period, with the change and the time both depending on the previous state.

As the complexity of signaling pathways increases, both deterministic and stochastic computational models are too complex to use for control design. In order to obtain a compact model usable for control design, we take a different approach, focusing on signal transduction time delays.

2. Computational Modeling and Analysis of Signal Transduction Time

2.1 Introduction

The approach described in this chapter regulates the chronological order of signaling events by controlling the extracellular cues in microfluidic in-vitro environment. Transient response curves are generated using deterministic reaction rate equations. To account for variability and heterogeneity of cells, initial conditions and parameter values are varied to create a family of response for each pathway, given a specific input. An optimal value of input cue is then determined for use in feedback control of the process. Future work will incorporate the proposed method into the development of a feedback control strategy for endothelial cell migration during angiogenesis.

2.2 Model Definition

We model each pathway as a simple linear signaling cascade in which receptor stimulation initiates successive activation of several downstream protein kinases (see figure 2-1). As described in the previous section, each pathway protein undergoes a modification cycle the reactions of which are summarized in table 2-1.

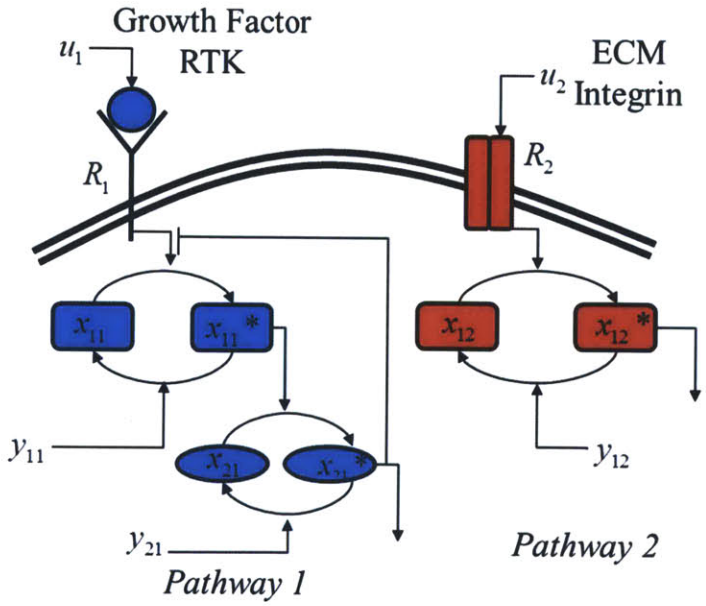

Figure 2-1: Activation mechanisms for each pathway characterized by a simple linear signal transduction cascade.

Ligand-receptor activation mechanisms differ depending on receptor type. For receptor tyrosine kinases (RTK), ligand-receptor binding causes dimerization of the receptor. The dimerized receptor then undergoes trans-phosphorylation, a process in which each receptor in the dimer complex activates the other. The activated dimer may then activate signaling molecules localized near the membrane such as Src, which plays an important role in cell migration such as cytoskeletal organization [15].

For integrins (ITG), the binding of extracellular matrix proteins causes activation [16]. In its active state the integrin the cytoplasmic "tails" may interact with other cytosolic molecules and induce activation of proteins such as FAK which is involved in focal adhesion assembly and disassembly [17].

The reactions can be modeled by a set of differential equations where the rate of concentration of each state is written as the sum of the reaction rates that produce the given state minus the sum of the consuming rates. The signaling cascade reactions and equations are summarized in the table 2-1.

In summary the entire dynamic signaling model for a given pathway may be represented as follows:

$$\begin{aligned} \dot{\vec{x}}(t) &= f(\vec{x}) + b(\vec{x})u(t) \\ \vec{y}(t) &= h(\vec{x}) \end{aligned} \tag{2.1}$$

where the state variable $\vec{x}(t)$ represents a vector of state variables that describe the concentration of individual species at a given time t, u represents input concentration which may be varied independently and $y(t)$ represents the desired state to be monitored.

Table 2-1: Enzymatic reactions describing the covalent modification cycle. Here are activation reaction constants and are inactivation reaction constants

Reaction	Chemical Equation	Ordinary Differential Equation
RTK Activation	$R_1 + u_1 \underset{b_1}{\overset{a_1}{\rightleftharpoons}} R_1 : u_1$ $R_1 : u_1 + R_1 \underset{d_1}{\overset{c_1}{\rightleftharpoons}} D_1 \xrightarrow{k_1} D_1{}^*$ $D_1{}^* + x_{11} \underset{\beta_{11}}{\overset{\alpha_{11}}{\rightleftharpoons}} D_1{}^* : x_{11}$ $D_1{}^* : x_{11} \xrightarrow{\lambda_{11}} D_1{}^* + x^*{}_{11}$	$\frac{dR_1}{dt} = b_1 \cdot (R_1 : u_1) - a_1 \cdot R_1 \cdot u_1 - d_1 \cdot D_1$ $\frac{dD_1}{dt} = c_1 \cdot R_1 \cdot (R_1 : u_1) - (d_1 + k_1) \cdot D_1$ $\frac{dD_1^*}{dt} = k_1 \cdot D_1 - \alpha_{11} \cdot (D_1^* : x_{11}) + (\beta_{11} + \lambda_{11}) \cdot (D_1^* : x_{11})$ $\frac{dx_{11}{}^*}{dt} = \lambda_{11} \cdot (D^* : x_{11}) + \bar{\beta}_{11} \cdot (z_{11} : x_{11}{}^*) - \bar{\alpha}_{11} \cdot z_{11} \cdot x_{11}{}^*$
ITG Activation	$R_2 + u_2 \underset{b_2}{\overset{a_2}{\rightleftharpoons}} R_2 : u_2$ $R_2 : u_2 + x_{12} \underset{\beta_{12}}{\overset{\alpha_{12}}{\rightleftharpoons}} R_2 : u_2 : x_{12}$ $R_2 : u_2 : x_{12} \xrightarrow{\lambda_{12}} R_2 : u_2 + x^*{}_{12}$	$\frac{dR_2}{dt} = b_2 \cdot (R_2 : u_2) - a_2 \cdot R_2 \cdot u_2$ $\frac{d(R_2 : u_2)}{dt} = (\beta_{12} + \lambda_{12}) \cdot (R_2 : u_2 : x_{12}) - \alpha_{12} \cdot (R_2 : u_2) \cdot x_{12}$ $\frac{dx_{12}{}^*}{dt} = \lambda_{12} \cdot (R_2 : u_2 : x_{12}) + \bar{\beta}_{12} \cdot (z_{12} : x_{12}{}^*) - \bar{\alpha}_{12} \cdot z_{12} \cdot x_{12}{}^*$
Activation	$x^*{}_{pk-1} + x_{pk} \underset{\beta_{pk}}{\overset{\alpha_{pk}}{\rightleftharpoons}} x^*{}_{pk-1} : x_{pk}$ $x^*{}_{pk-1} : x_{pk} \xrightarrow{\lambda_{pk}} x^*{}_{pk} + x^*{}_{pk-1}$ $x^*{}_{pk} + y_k \underset{\bar{\beta}_{pk}}{\overset{\bar{\alpha}_{pk}}{\rightleftharpoons}} x^*{}_{pk} : y_{pk}$ $x^*{}_{pk} : y_{pk} \xrightarrow{\bar{\lambda}_{pk}} x_{pk} + y_{pk-1}$ $p(\textit{pathway number})$ $k(\textit{pathway protein number}) \neq 1$	$\frac{dx_{pk}}{dt} = \beta_{pk} \cdot (x^*{}_{pk-1} : x_{pk}) - \alpha_{pk} \cdot x^*{}_{pk-1} \cdot x_{pk} + \bar{\lambda}_{pk} \cdot (x^*{}_{pk} : y_{pk})$ $\frac{dx^*{}_{pk}}{dt} = \bar{\beta}_{pk} \cdot (y_{pk} : x^*{}_{pk}) - \bar{\alpha}_{pk} \cdot y_{pk} \cdot x^*{}_{pk} + \lambda_{pk} \cdot (x^*{}_{pk-1} : x_{pk})$
Auxiliary Equations	Mass conservation: $R_{1tot} = R_1 + R_1 : u_1 + D_1 + D_1{}^* + D_1{}^* : x_{11}$ $R_{2tot} = R_2 + R_2 : u_2 + R_2 : u_2 : x_{12}$ $x_{pktot} = x^*{}_{pk-1} : x_{pk} + x^*{}_{pk} + x_{pk}$ $y_{pktot} = y_{pk} + x^*{}_{pk} : y_{pk}$	

2.3 Problem Statement and Methods

2.3.1. Problem Statement

Computationally, we wish to accomplish the following:

1. Generate multiple response curves based on different initial condition for each pathway given the input intensity.
2. Use the simulated data to create a time delay distributions of signal transduction times.
3. Use distributions to determine the optimal input intensity that leads to maximum probability of a specified signal transduction order.

2.3.2. Methods

Transient response data for each pathway may be simulated by creating a continuous system model defined by a set of state variables. The continuous state changes are modeled by a sum of rates describing the increase and decrease of the molecular concentrations as discussed in the previous section. Numerical integration algorithms are used to generate simulated data from the derived equations.

To account for the heterogeneity among cells, the simulation should be repeated for perturbed parameter values and initial conditions in a realistic range. The time delay can be determined from simulated data as the time from zero to fifty percent of the steady state value of the time response. The distribution of the activation time delay can be created by plotting a histogram (see Fig. 2-2).

Consider pathway 1 in the representative network shown in Fig. 2-2. Let τ_1 be the time delay from the onset of input cue at the receptor to the activation and subsequent binding of signaling molecule and u_1 be the input cue concentration. The probability of τ_1 when input cue u_1 is applied is expressed as:

$$f_{X1}(\tau_1 \mid u_1)d\tau = \Pr\left(\tau_1 \le X_1 \le \tau_1 + d\tau \mid u_1\right) \tag{2.2}$$

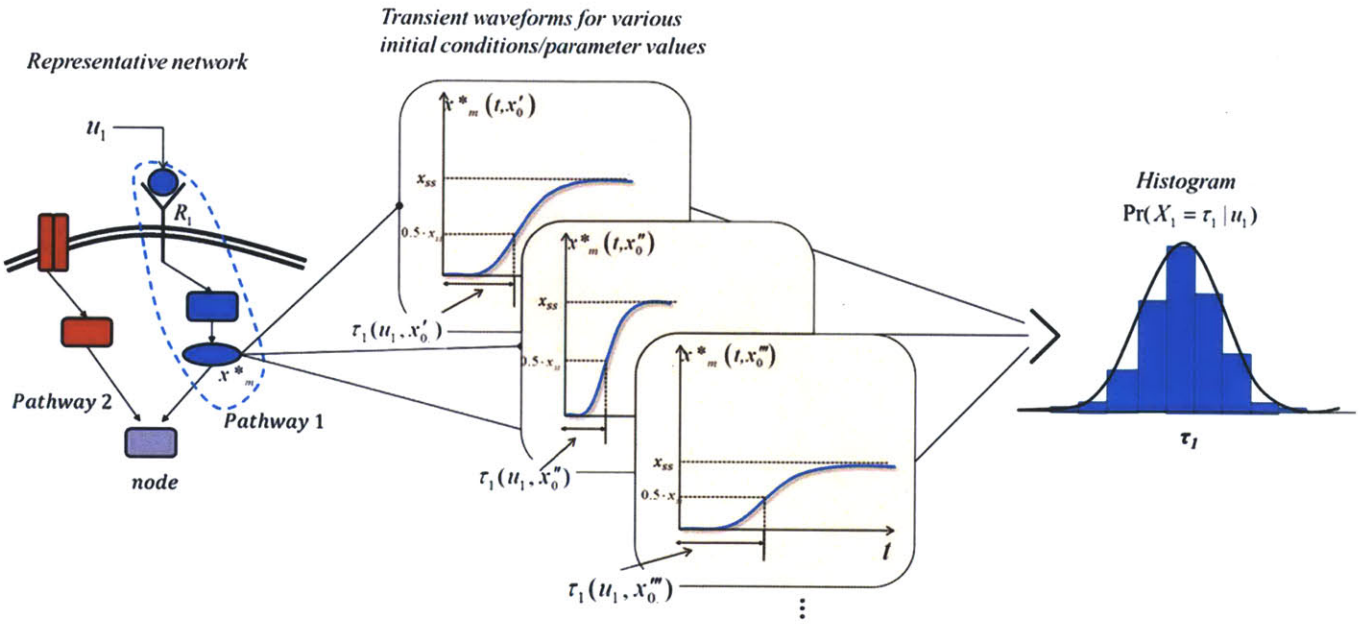

Figure 2-2: Diagram illustrating the approach to calculating time delay distribution from a representative pathway. Transient responses for the activated form of the last molecule in each pathway (xm* (t) where m is the number of molecules in a given pathway) are simulated from chemical kinetic differential equations under various initial conditions and parameter values (x0', x0'', x0''',...). Time delay is calculated from the response curve as the time from zero to fifty percent of the steady state value (0.1·xss). The time delay distribution can be created by plotting a histogram. By varying the input cue concentration (u1), a family of distributions may be found for a given pathway.

By varying the input cue concentration (u_1), a family of distributions may be found for a given pathway. In general, for multiple pathways ($i=1,\ldots,n$) the probability for τ_i given input cue u_i is expressed:

$$f_{Xi}(\tau_i \mid u_i)d\tau = \Pr\left(\tau_i \le X_i \le \tau_i + d\tau \mid u_i\right), \quad i = 1,\ldots,n \tag{2.3}$$

Once probability densities for individual pathways are identified, the probability of a specific chronological order of signaling events $\tau_1 \ge \tau_2 \ge \cdots \ge \tau_n$ for pathways ($1,..,n$) can be computed as:

$$g(\tau_1 \ge \tau_2 \ge \cdots \ge \tau_n \mid u_1,\ldots,u_n) = \iiint_{\tau_1 \ge \tau_2 \ge \cdots \ge \tau_n} f_{X1}(\tau_i \mid u_1) f_{X2}(\tau_2 \mid u_2) \cdots f_{Xn}(\tau_n \mid u_n) d\tau_1 \ldots d\tau_n \tag{2.4}$$

Assuming that $\tau_1,\ldots,\tau_n$ are statistically independent. Using Bayes' theorem, the concentrations of input cues for generating a desired chronological order of signaling events can be obtained from:

$$\begin{aligned} &g(u_1,\cdots,u_n \mid \tau_1 \geq \tau_2 \geq \cdots \geq \tau_n) \\ &= \eta g(\tau_1 \geq \tau_2 \geq \cdots \geq \tau_n \mid u_1,\cdots,u_n) P(u_1,\cdots,u_n) \end{aligned} \tag{2.5}$$

where η is a scaling factor.

2.4 Numerical Example

2.4.1. Data Simulation

We would like to test our approach for multiple pathways with varying input cue intensities. Consider the simple network consisting of two pathways as discussed previously (see Fig. 2-1). Pathway 1 is activated through GF-RTK stimulation and consists of a two stage linear signaling cascade. We have modified pathway 1 slightly to incorporate negative feedback (i.e. the activated form of the last kinase initiates activation of the first kinase by binding to it). This type of negative feedback mechanism is common in signaling pathways. Pathway 2 is activated through ECM-ITG stimulation and consists of 1 stage linear signaling cascade.

Response curves for each pathway were generated by numerical integration ODE model described by (2.1). Integrations were repeated for uniformly random parameter values and initial conditions for a given input concentration to simulate heterogeneity of the pathway. Using this simulated data, a family of time delay distributions given various inputs could be created. Fig. 2-5A, shows a histogram of τ_1 (delay associated with pathway 1) given $u_1 = 1pM$.

By examining the distributions as a function of the given input ($f_{X1}(\tau_1 \mid u_1)$), we detect that the delay decreases with increased input concentration (see Fig. 2-5B). Similar relationships have been observed in other more complex computational models. In particular, it was observed that with increased ligand concentration, the speed of activation increases, decreasing the delay before activation.

2.4.2. Analysis

Based on trends observed through simulation our goal is to find the optimal input values (u_1, u_2) that will ensure the sequence of molecular activations $(\tau_1 < \tau_2)$. In other words we wish to find the optimal value of u_1, u_2 that maximizes the conditional probability: $\Pr(\tau_1 < \tau_2 \mid u_1, u_2)$.

Let τ_1 be the random variable associated with the time delay of the last kinase in pathway 1 (x_{21}^*) in response to growth factor stimuli u_1 (see Fig. 2-1). Let τ_1 be the random variable associated with the time delay of the last kinase in pathway 2 (x_{21}^*) in response to a different external stimuli u_1.

Re-writing (2.4):

$$g(\tau_1 < \tau_2 \mid u_1, u_2) = \iint_{\tau_1 < \tau_2} f_{X1}(\tau_1 \mid u_1) \cdot f_{X2}(\tau_2 \mid u_2) d\tau_1 d\tau_2 \tag{2.6}$$

Using the distributions generated for each pathway we can find the desired conditional distribution for $f_{Xi}(\tau_i \mid u_i)$ as shown in Fig. 7. As is expected, the highest probability of $\tau_1 < \tau_2$ occurs when input u_1 is the greatest and input u_2 is the least and the lowest probability occurs when u_2 is the greatest and u_1 is the least.

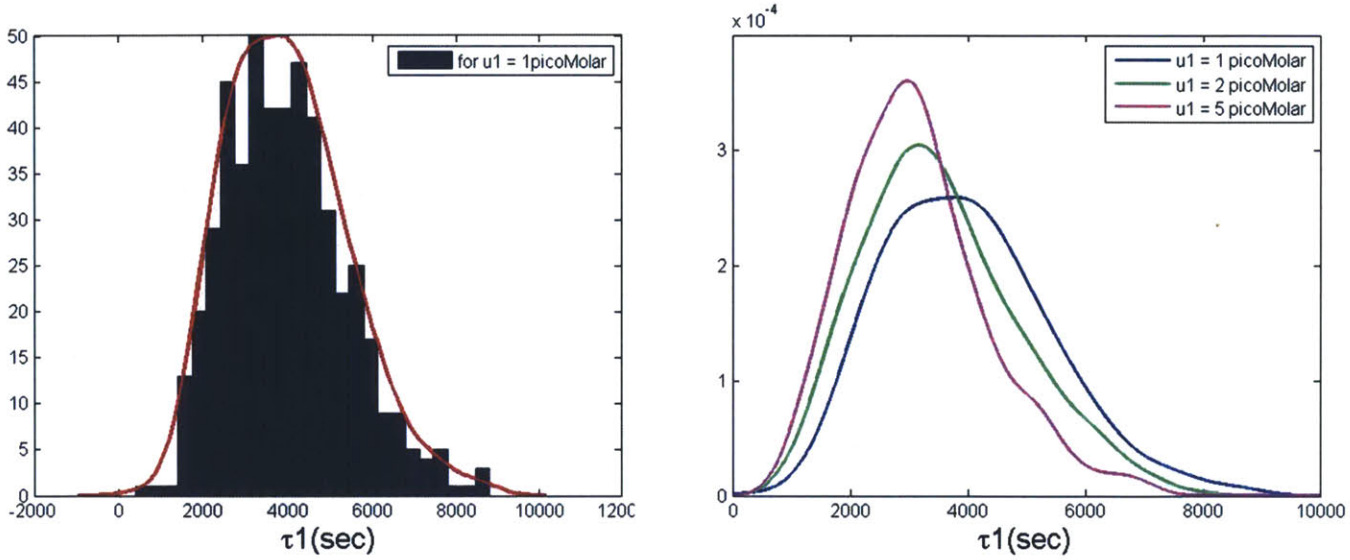

Figure 2-3: A. Histogram of τ1 (delay associated with pathway 1) given u1=1picoMolar. B. Time delay distributions for τ1 given various inputs .

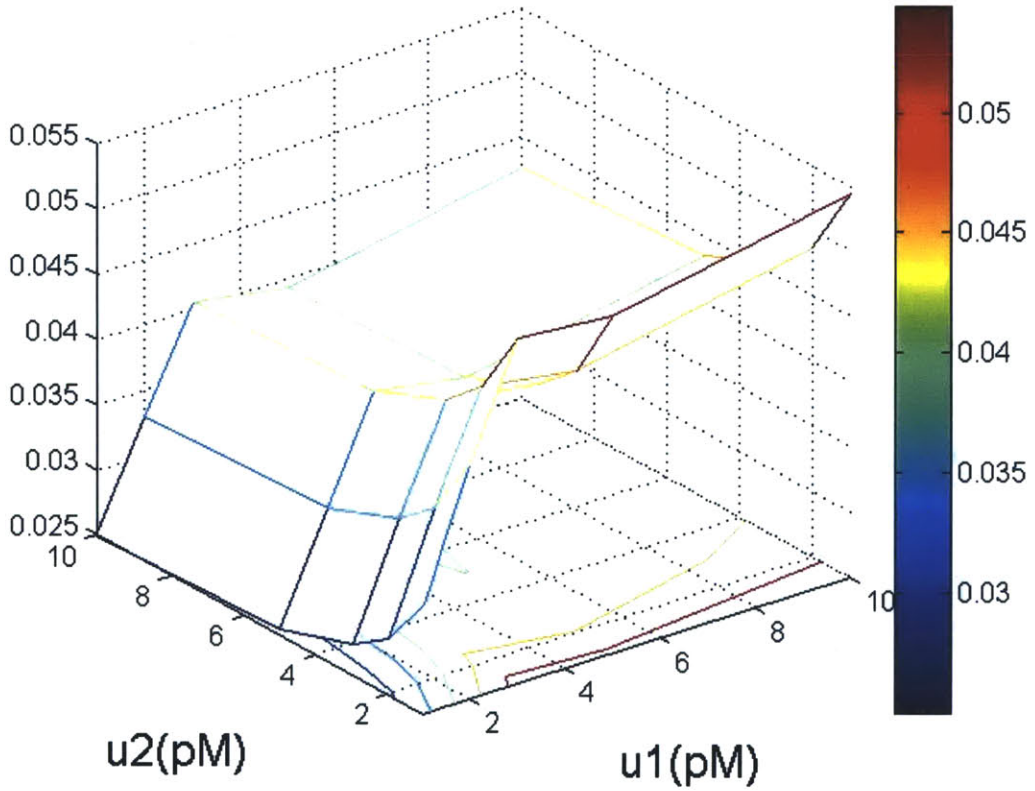

Figure 2-3: 3D plot of probability distribution for varying inputs. The maximum is the optimal for inputs u1 and u2.

2.5 Control Applications

Based on the stochastic time delay model presented in the previous section we can construct various control systems. Consider a population of agents, each of which is an independent unit that can receive extracellular cues, transduce signaling molecules, and produce a certain response. Agents include not only individual cells but also compartments of a single cell, such as focal adhesions, lamellipodia, and fillipodia.

The control objective is to guide the population towards a desired state, such as a desired proportion of multiple phenotypic states. For example, consider two distinct phenotypic states, 1 and 2, created with two signaling pathways and two extracellular cues, as in the case of the Growth Factor – Integrin Induced Signaling discussed in the previous section. State 1 emerges when $\tau_1 \geq \tau_2$, and vice versa.

Suppose that we want to drive the entire population towards State 1. The optimal input cues are given by:

$$\mathbf{u}^o = \arg\max_{\mathbf{u}\in\mathbf{U}} g(\tau_1 > \tau_2 \mid u_1, u_2) \tag{2.7}$$

If the desired goal is to drive 70 % of the population towards State 1 and 30% to State 2, the optimal input cues are given by:

$$\mathbf{u}^o = \arg\min_{\mathbf{u}\in\mathbf{U}} \left[\{ g(\tau_1 > \tau_2 \mid u_1, u_2) - 0.7 \}^2 \right] \tag{2.8}$$

If the solution is not unique, one can take the one having the minimum squared input magnitude:

$$\mathbf{u}^o = \arg\min_{\mathbf{u}\in\mathbf{U}} \left[(u_1^2 + u_2^2) \mid g(\tau_1 > \tau_2 \mid u_1, u_2) = 0.7 \right] \tag{2.9}$$

2.6 Conclusion

In this chapter, a dynamic model of a signaling network was developed based on time delays associated with individual pathways in relation to the extracellular cue. As an example, the time delay associated with 2 signaling pathways was examined in relation to the extracellular cue. Time delays of multiple pathways were compared, and the probability that the multiple signaling events occur in a desired chronological order was evaluated in relation to input cues. From this analysis we conclude that the model developed allows for optimal input intensity to be predicted for feedback control of the process. However there are several caveats. Firstly, although we account for heterogeneity of cells, inherent stochasticity of protein interactions remain un-modeled. In addition, the simulated waveforms approximate discrete state changes occurring randomly in continuous time. The next chapter attempts develop a model that this "pseudo-discrete" aspect of the system using standard probability theory to calculate signal transductions.

3. Stochastic Modeling of Signal Transduction Time

3.1 Introduction

This chapter presents a modeling framework for an intracellular signaling network based on formalisms derived from the fundamental concepts in probability theory. 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 numerical example illustrates the approach.

3.2 Model Definition

We model the cell as a well-mixed biochemical reactor with uniformly distributed molecules. In reality species are segregated into different spatial domains and are not well-mixed. Even though extensions to modeling spatial inhomogeneity or variations in temperature are possible, [18], for simplicity we will consider a well-mixed uniformly distributed assumption.

Consider the network shown in Fig. 3-1. Pathways A and B are linear cascades that intersect at a common node molecule. We assume that molecules from pathway A do not interact with molecules from pathway B except at the node molecule (i.e the pathways are independent). The time dependent number of each molecular species in all pathways is given by vector: $\Upsilon(t) = [X^A(t), X^B(t)]$.

$$\begin{aligned} X^A(t) &= [n^A_{P_1}, n^A_{P_1*}, n^A_{P_2}, n^A_{P_2*}, \ldots] \\ X^B(t) &= [n^B_{P_1}, n^B_{P_1*}, n^B_{P_2}, n^B_{P_2*}, \ldots] \end{aligned} \tag{3.1}$$

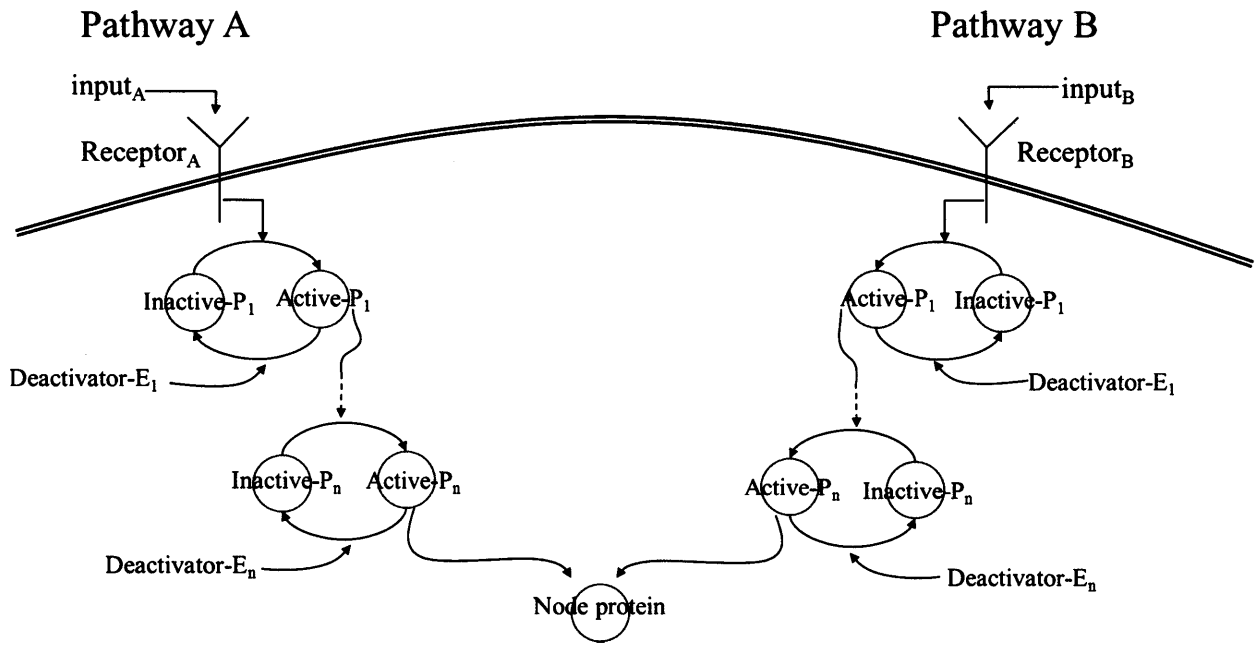

Figure 3-1: Simple network showing linear covalent modification cycles for each given pathway

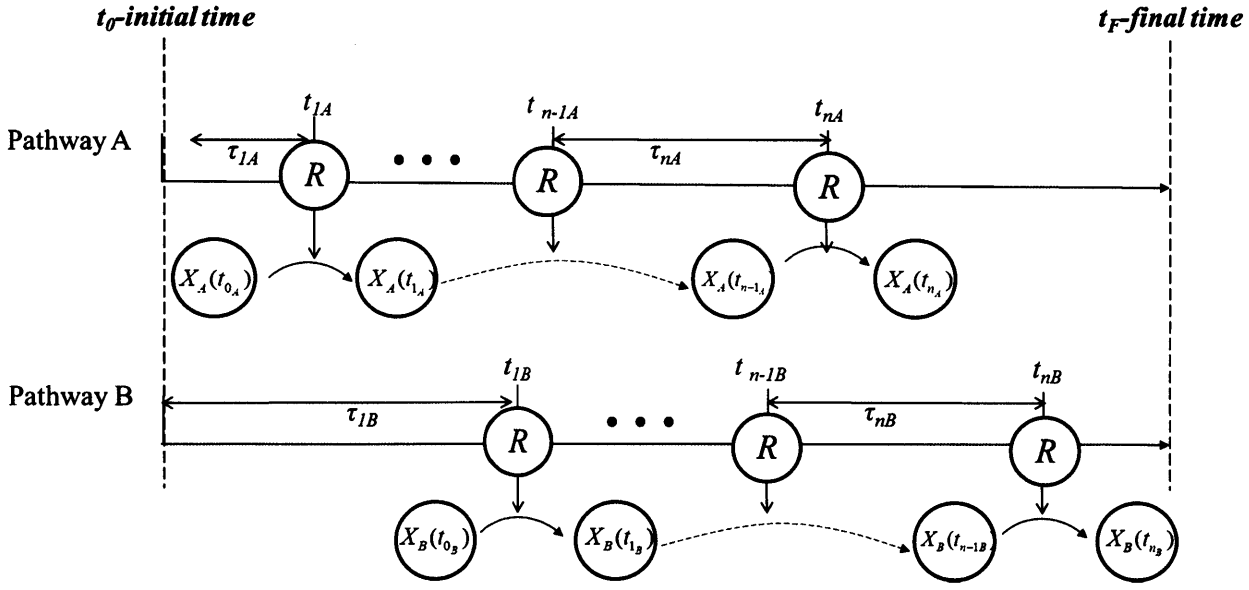

Figure 3-1: Representative Diagram of parallel reaction occurrences in pathways

Each reaction (R_ν) will occur with probability:

$$a_\nu(\Upsilon(t))\cdot dt = h_\nu c_\nu \cdot dt \tag{3.2}$$

where $a_\nu(\Upsilon(t))$ is the propensity function for the $\nu - th$ reaction, c_ν is a constant which depends only on the temperature and physical properties of the system, and h_ν is the number of distinct R_ν molecular reactant combinations available in state $X^A(t), X^B(t)$.

Consider the diagram shown in fig. 3-1. The specific reaction time in each pathway (termed occurrence times) are notated as $t_1,...,t_n$. The time intervals between reactions (when no reaction is occurring) are notated as $\tau_1,...,\tau_n$ which we have termed delay times. Occurrence times and delay times may be related in the following manner (see fig. 3-1):

$$\begin{cases} \tau_1 = t_1 - t_0 \\ \tau_2 = t_2 - t_1 \\ \vdots \\ \tau_n = t_n - t_{n-1} \end{cases} ; t_n = t_0 + \tau_1 + \tau_2 + \cdots + \tau_n \tag{3.3}$$

Let:

$$a_0(\Upsilon(t)) = \sum_{\nu^A=1}^{m^A} a_{\nu^A}(X^A(t)) + \sum_{\nu^B=1}^{m^B} a_{\nu^B}(X^B(t)) \tag{3.4}$$

be the total propensity over all pathways and reactions in state $\Upsilon(t) = X^A(t), X^B(t)$. It has been shown under well-mixed conditions the delay times follow an exponential distribution with varying rate parameter $a_0(\Upsilon(t))$ that is independent of reaction type given state $\Upsilon(t)$ [19, 20]. The densities to describe the delay times $\tau_1,\cdots,\tau_n$:

$$\begin{cases} f_{T_1}(\tau_1) = a_0(\Upsilon(t_0))\exp(-a_0(\Upsilon(t_0))\cdot\tau_1) \\ f_{T_2}(\tau_2) = a_0(\Upsilon(t_1))\exp(-a_0(\Upsilon(t_1))\cdot\tau_2) \\ \vdots \\ f_{T_n}(\tau_n) = a_0(\Upsilon(t_{n-1}))\exp(-a_0(\Upsilon(t_{n-1}))\cdot\tau_n) \end{cases} , \tau_1,...,\tau_n \geq 0 \tag{3.5}$$

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

$$f_{\bar{T}_n}(t_n) = (f_{T_n} * f_{\bar{T}_{n-1}})(t_n) = \int_0^{t_n} f_{Tn}(t_n - t_{n-1}) \cdot f_{\bar{T}_{n-1}}(t_{n-1}) dt_{n-1} \tag{3.6}$$

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: $L\{f(t)\} \equiv F(s) = \int_0^\infty e^{-st} f(t) dt$

$$F_{\bar{T}_n}(s) = F_{T_n}(s) \cdot F_{\bar{T}_{n-1}}(s) = \prod_{i=1}^{n} F_{T_n}(s) = \frac{\prod_{i=0}^{n-1} a_0(\Upsilon(t_i))}{\prod_{i=0}^{n-1} [s + a_0(\Upsilon(t_i))]} \tag{3.7}$$

The convolution have reduces to a multiplication in the Laplace domain. Using partial fractions:

$$F_{\bar{T}_n}(s) = \left(\prod_{i=0}^{n-1} a_0(\Upsilon(t_i)) \right) \sum_{i=0}^{n-1} \frac{A_i}{[s + a_0(\Upsilon(t_i))]} \tag{3.8}$$

$$where,\ A_i = \frac{1}{\prod_{j \neq i} a_0(\Upsilon(t_j)) - a_0(\Upsilon(t_i))}$$

Taking the inverse Laplace:

$$f_{\bar{T}_n}(t_n) = L^{-1}\{F_{\bar{T}_n}(s)\} = \left(\prod_{i=0}^{n-1} a_0(\Upsilon(t_i)) \right) \sum_{i=0}^{n-1} \frac{\exp[-a_0(\Upsilon(t_i)) \cdot t_n]}{\prod_{j \neq i} a_0(\Upsilon(t_j)) - a_0(\Upsilon(t_i))} \tag{3.9}$$

The equation (3.9) above is valid for the calculation of distribution reaction occurrence times when the total propensity $a_0(\Upsilon(t))$ parameter is distinct and does not repeat over time. In many cases the total propensity of a reaction may repeat over time.

We would like to find a more general expression if possible. Let us group *delay-time* distributions with same rate parameters:

$$t_n = t_0 + \tau_1 + \cdots + \tau_i + \cdots + \tau_j + \cdots + \tau_k + \cdots + \tau_l + \cdots + \tau_m + \cdots + \tau_p + \cdots + \tau_n$$
$$t_n = \underbrace{\tau_i + \cdots + \tau_j + \ldots + \tau_n}_{\substack{V_1 \\ same\ a_0 = A_1}} + \cdots + \underbrace{\tau_p + \cdots + \tau_k + \cdots}_{\substack{V_i \\ same\ a_0 = A_i}} + \cdots + \underbrace{\tau_l + \cdots + \tau_m}_{\substack{V_r \\ same\ a_0 = A_r}} \quad (3.10)$$

We rename the equal propensity parameters $(A_1, \ldots, A_r)$ where $r(1<r<n)$ is the number of distinct rate parameters.

Next we define a new set of random variables $V_i\,(i = 1, \ldots, r)$ which is the sum of the delay times with equal rate parameters (A_i) : $V_i = \tau_l + \tau_m + \ldots + \tau_{N_i}$. Where N_i is the number of delay times with same parameter A_i.

The distribution of V_i is the convolution of exponential *delay time* distributions with constant rate parameter. This distribution is the well-known Erlang distribution:

$$f_V(V_i) = \frac{A_i^{N_i} V_i^{N_i - 1}}{(N_i - 1)!} \exp(-A_i V_i);\ V_i = \tau_l + \tau_m + \ldots + \tau_{N_i} \quad (3.11)$$

We now have a set of r random variables each with Erlang distributions described by the above equation:

$$t_n = V_1 + \cdots + V_r;\ f_V(V_i) = \frac{A_i^{N_i} V_i^{N_i - 1}}{(N_i - 1)!} \exp(-A_i V_i) \quad (3.12)$$

To obtain the closed form solution, we now have to propagate the uncertainty of the sum of Erlang distributed independent random variables V_i. Finally we obtain:

$$f_T(t_n) = \sum_{i=1}^{n} A_i^{N_i} \exp[-t_n A_i] \sum_{j=1}^{N_i} \frac{(-1)^{N_i - j}}{(N_i - j)!} t_n^{\,j-1} \times \sum_{\substack{m_1 + \ldots + m_n = N_i - j \\ m_i = 0}} \prod_{\substack{l=1 \\ l \neq i}} \binom{N_l + m_l - 1}{m_l} \frac{A_l^{N_l}}{(A_l - A_i)^{N_l + m_n}} \quad (3.13)$$

3.3 Problem Statement and Methods

3.2.1. 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.

3.2.2. Methods

To calculate the distribution of signal transduction time, we enumerate all *valid* possible reaction sequences between the initial state $\Upsilon_0 = \left[X_0{}^A(t), X_0{}^B(t)\right]$ and the first occurrence of the activation reaction of the last cascade protein in time span $(t_0 - t_f)$ (see fig. 3-2). 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 $\Upsilon(t_i)$ $(i = 1,...,n-1)$ may also be predetermined. Using this knowledge and given that the propensity $a_0(\Upsilon(t))$ is distinct in the set timespan, we may calculate occurrence time distributions for each possible reaction sequence using (3.9):

$$f_{\bar{T}_n}(t = t_n \mid S_i = R_1,...,R_n) \tag{3.14}$$

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

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

$$f_{\bar{T}}(t_n) = \sum_i f_{\bar{T}_n}(t_n \mid S_i) f_{\mathbf{S}}(S_i) \tag{3.15}$$

Where $f_S(S_i)$ is the probability that the reactions will occur in order $S_i = R_1, ..., R_n$. We may find the propensity of each reaction given the predetermined states to calculate this probability:

$$f_S(S_i) = \prod_i^n a_i(\Upsilon(t_i)) \tag{3.16}$$

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 pathways: $f_{T_n}(t_{n_A}), f_{T_n}(t_{n_B}), ... f_{T_n}(t_{n_Z})$ we may calculate the distribution of a specific transduction order:

$$f_{T_n}(t_{n_A} < t_{n_B} < \cdots < t_{n_Z}) = \iint_{t_{n_A} < t_{n_B} < \ldots t_{n_Z}} f_{T_n}(t_{n_A}) \cdot f_{T_n}(t_{n_B}) \cdots f_{T_n}(t_{n_Z}) dt_{n_A} dt_{n_B} .. dt_{n_Z} \tag{3.17}$$

3.4 Numerical Example

Let us examine a simplified model of the VEGFR-2 and $\alpha_v\beta_3$ integrin pathways described in Fig. 1-8. 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 $\alpha_v\beta_3$ integrin pathway). Each cascade is capable of 3 reactions involved in protein activation. Let R_1 be the forward binding activation between protein and receptor, R_2 is the reverse binding reaction and R_3 is the catalysis reaction in which the activated form of the protein is produced:

$$\text{Protien} + \text{Activated-Receptor} \underset{R_2}{\overset{R_1}{\rightleftarrows}} \text{Complex}$$

$$\text{Complex} \xrightarrow{R_3} \text{Activated-Receptor} + \text{Activated-Protien}$$

Sequences leading to activation of the RhoA-ROCK or Pyk protein are those that end in R_3 for each pathway. Therefore we enumerate all sequences in which the only occurrence of R_3 is the last reaction in the sequence. We assume binary initial conditions were the activated form of

the protein is initially 0. All forward and catalysis reaction constants are $(c_1, c_3 = 1)$ and reverse reactions $(c_2 = .1)$

Figure 3.4B illustrates the distribution of the signal transduction (or the first occurrence time of $R_3 : f_{\bar{T}}(t = t_n \mid S_i)$) for several sequences. The red curve represents the sequence is $S = (R_1, R_3)$ which is the most probable and direct sequence to occur. As the number of reactions in the sequence increase we see a decrease in the probability that activation will occur and shift in the distribution curve. Figure 3.5 represents the marginal distribution $f_{\bar{T}_n}(t = t_n)$. Given 2 reaction cascades occurring independently, we may compute the probability that the signaling transduction of the first cascade is faster that the second. Calculating, we obtain $\Pr(t^{\text{RhoA-ROCK}} < t^{\text{Pyk}}) \sim .49$ which is expected since the parameters are the same for both pathways.

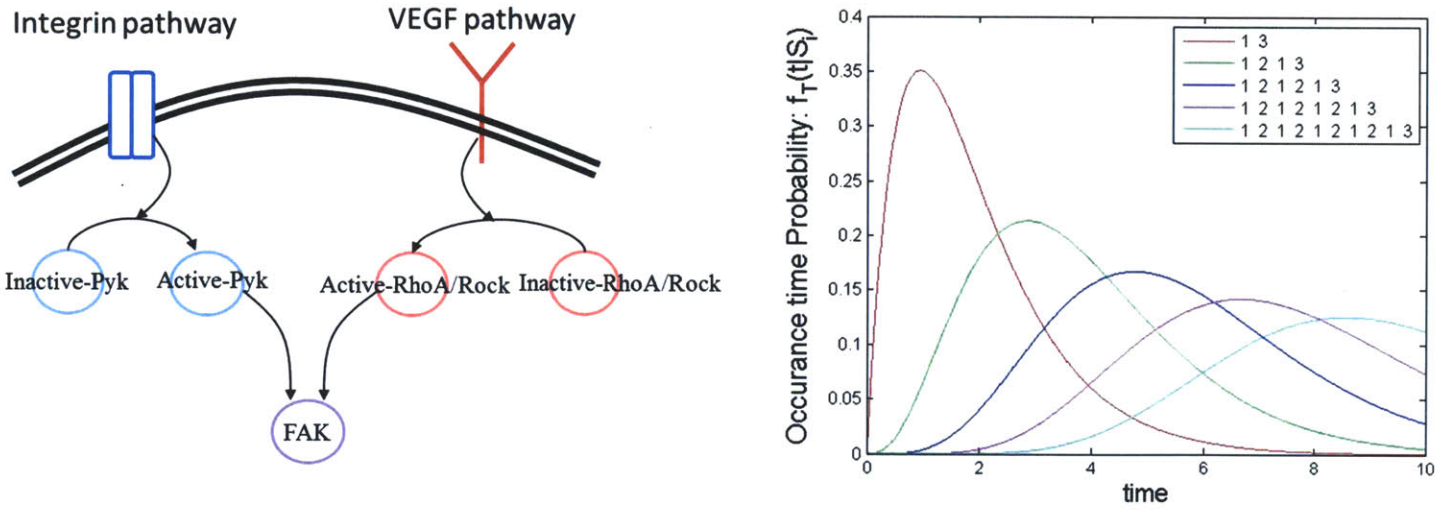

Figure 3-2:A. Diagram of simplified VEGFR-and $\alpha_v\beta_3$ integrin pathway. B. Occurrence time probability for variyning sequnces Si. $f_{\bar{T}}(t = t_n \mid S_i)$

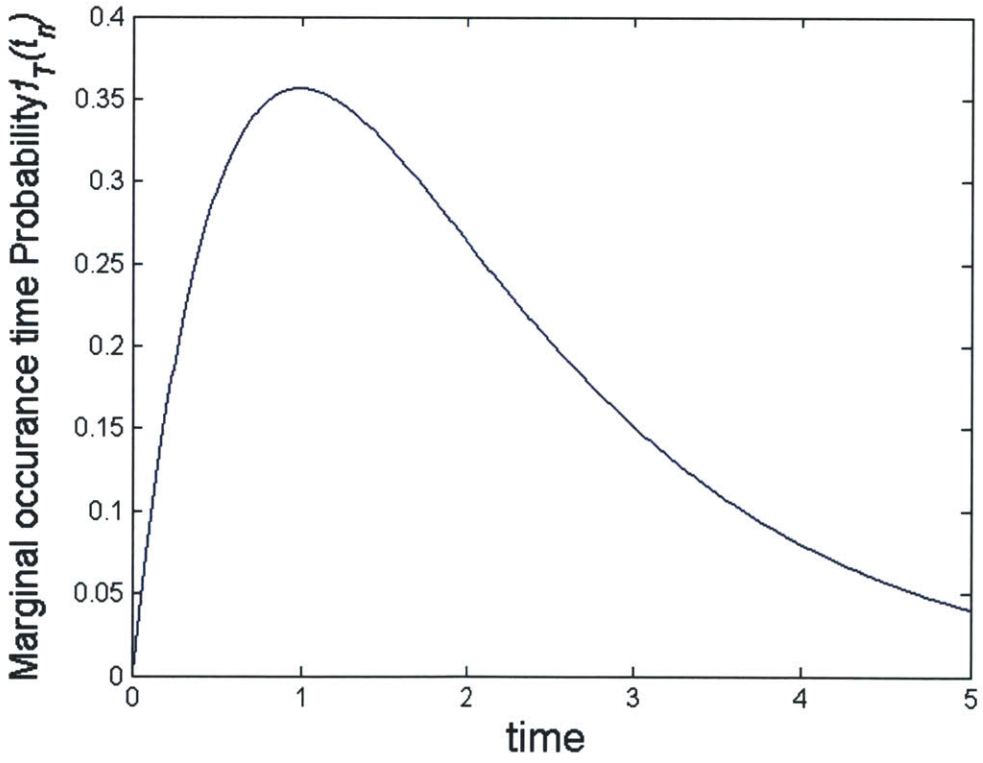

Figure 3-3: Overall occurrence time probability $f_{\bar{T}_n}(t = t_n)$

3.5 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 simple numeric example.

4. Conclusion

4.1 Contribution of this Work

This thesis has developed a several modeling frameworks for regulating intracellular signaling dynamics for the purpose of controlling endothelial cell migration in angiogenesis. We began by developing an input-output time-delay model derived from simulated data that was used to predict the optimal extracellular input intensity for a desired response. We concluded that the simulated waveforms approximated discrete state changes occurring randomly in continuous time. Therefore we developed a "pseudo discrete "modeling framework based on standard probability theory. Using this approach an exact solution could be found for the signal transduction time.

Unlike previous works, the models developed in this thesis exploit the effect of signaling order on extracellular response. Future work should include comparing models with actual data as well as developing a control law to fit the model. The following sections summarize several control strategies that may be explored.

4.2 Future Directions in Control

A main setback of microfluidic device technology is the delay associated with development of the correct concentration gradient in the device. This delay is dependent on the diffusion characteristics of the growth factor medium, the properties of the gel matrix as well as the geometric constraints of the device. To increase performance, a more immediate feedback mechanism may be required to manipulate protein activity at precise time and places. Recent advances in genetic engineering allow protein activity to be controlled reversibly and repeatedly using light [21]. In particular, it has been shown that photo-activatable Rac1 could be activated using light to cell motility and control the direction of cell movement [21].

4.2.1 System Definition

Consider a protein kinase signal transduction cascade as shown in figure 4-1 Each kinase may transition to its activated state with a probability (p_i) proportional to the amount of inactivated kinase present (P_i) as well as the amount of previous activated kinase (P^*_{i-1}). The probability is also governed by the forward/backward and catalytic reaction constants associated with activation ($\alpha_i, \lambda_i, \bar{\beta}_i$) (See table 1-1). The probability (q_i) for the kinase to transition into its inactivated state is governed by the amount of activated kinase (P^*_i) and in-activator enzyme (E_i) present as well as the backward/forward rates and catalytic rates associated with deactivation ($\bar{\alpha}_i, \bar{\lambda}_i, \beta_i$).

4.2.2 Problem Statement

We wish to alter intracellular environment in such a way that the state transition probabilities of the last protein in the cascaded pathway (p_3, q_3) may be controlled. Using photo-activatable proteins, we may locally change the reaction rates ($\bar{\alpha}_3, \bar{\lambda}_3, \beta_3, \alpha_3, \lambda_3, \bar{\beta}_3$) related to these probabilities. In this way, the molecular concentration of the activated protein P_3^* may be driven to a desired amount. We assume that the light source used to control activity of P_3^* has no effect on the activity of the upstream proteins. We examine the convergence of activated protein P_3^* with and without micro-fluidic device feedback.

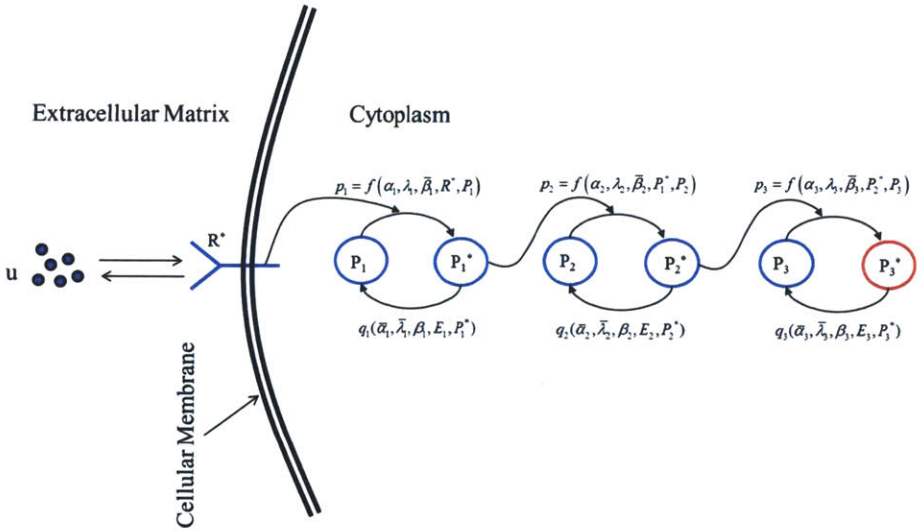

Figure 4-1: Protein cascade state transition diagram

4.2.3 Feedback Control of Single Cell

Let us consider n_t^P as a discrete random variable capable of 3 distinct states. n_t^P takes the value +1 when the protein transitions from its inactive to active state, -1 when the protein transitions from active to inactive state, and zeros when the protein remains in its current state. Figure 4-2 shows the probability mass function of the random variable n_t^P.

$$n_t^P = \begin{cases} +1 & inactive \rightarrow active \\ 0 & remain\ in\ current\ state \\ -1 & active\ to\ inactive \end{cases}$$

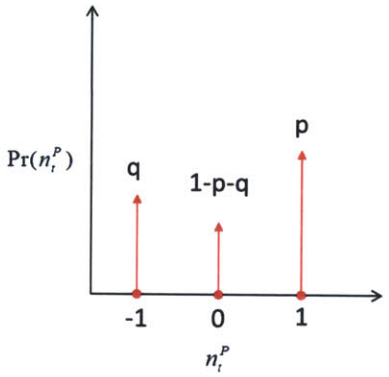

Figure 4-2: Probability Mass function of random variable n_t^P

Then we may define the number of activated protein at time t as:

$$X_t^{P*} = X_{t-1}^{P*} + n_t^P \tag{4.1}$$

Using the transition activation/inactivation transition probabilities, the expected amount of activated protein in the next time step can be written as:

$$E[X_t^{K^*} \mid X_{t-1}^{K^*}] = X_{t-1}^{K^*} + E[n_t^K] = p - q + X_{t-1}^{K^*} \tag{4.2}$$

The variance can be written as:

$$Var[X_t^{K^*} \mid X_{t-1}^{K^*}] = X_{t-1}^{K^*} + E[n_t^K] = p + q - (p-q)^2 \tag{4.3}$$

The goal is to bring the number of molecules $P_3^*(X_t^{P_3^*})$ to reach the desired value: $y = X_t^{K_3^*} \rightarrow y_d$. We define the error between the desired output and the desire values as:

$$e_t = y_d - y_t \tag{4.4}$$

We consider unilateral control such that q is approximately zero while $e_t > 0$ and p is approximately zero while $e_t < 0$.

$$p = \begin{cases} 0, & e_t < 0 \\ p(e_t), & e_t > 0 \end{cases}$$
$$q = \begin{cases} q(e_t), & e_t < 0 \\ 0, & e_t > 0 \end{cases} \tag{4.5}$$

This may be a valid assumption if the a reaction rates associated with activation will be increased to several orders of magnitude higher than the deactivation rates while $e_t > 0$ which actively decreases the probability of a deactivation reaction occurring. The reverse scenario is true for $e_t < 0$.

To ensure a stable control law, we use Lyapunov-like stability theory[22]. Let us define $V^s(e_t)$ as the stochastic Lyapunov function. V^s is a nonnegative , super-martingale function satisfying $V^s(e_t = 0) = 0$ and $V^s(e_t) > 0,\ e_t \neq 0$. Define $Q_m = \left[e_t : V^s(e_t) < m\right],\ m < \infty$. If a nonnegative, real, scalar valued function $k(e_t)$ exists such that the difference between $V^s(e_t)$ and the conditional mean $E\left[V^s(e_{t+1} \mid e_t)\right]$ at time $t+1$ is bounded as:

$$E\left[V^s(e_{t+1} \mid e_t)\right] - V^s(e_t) \triangleq -k(e_t) \leq 0 \tag{4.6}$$

in Q_m ,then e_t converges to

$$e_t \to P_m = Q_m \cap \{e : k(e) = 0\} \tag{4.7}$$

with probability no less than $1 - V^s(e_0)/m$. If m is sufficiently large, then the probability of convergence is 1.

Let us choose the Lyaponov function as:

$$V_t^s = e_t^2 \tag{4.8}$$

For asymptotic stability we require that $\Delta V_t^s = E[V_{t+1}^s(e_{t+1} \mid e_t)] - V_t^s(e_t) \le 0$. Substituting expressions and simplifying:

$$\begin{aligned} \Delta V_t^s &= Var(e_{t+1} \mid e_t) + E[e_{t+1} \mid e_t]^2 - e_t^2 \le 0 \\ &= p(1-2e_t) + q(1+2e_t) \le 0 \end{aligned} \tag{4.9}$$

As discussed previously we assume unilateral control. Therefore for $e_t > 0$ ($q = 0$) reverts to:

$$p(1-2e_t) \le 0 \tag{4.10}$$

Therefore the control law that will guarantee stability:

$$p = \begin{cases} 0 < p < 1 & e_t > 0.5 \\ p = 0 & e_t < 0.5 \end{cases} \tag{4.11}$$

As can be seen, for $e_t > 0.5$ any probability will ensure stability regardless of error. The value that maximizes ΔV_t^s is therefore $p = 1$.

4.2.4 Feedback Control of Homogeneous Cell Populations

Up till now we have only considered the response of a single cell to various inputs. In reality the angiogenic process consists of a vast number of cells responding to the given stimuli. It is currently infeasible to measure multiple cells at a time using FRET techniques. However, other florescent methods to determine protein expression may be an useful alternative. Assuming that a method to measure protein levels is available we may redefine the output as an aggregate output given.

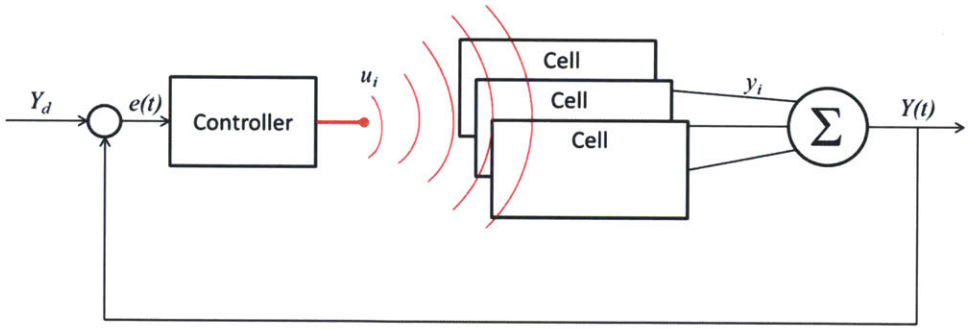

Figure 4-3: Population: Diagram Depicting Proposed Protein Feedback of Cell Population

Using similar techniques in the previous section it has been shown in [23, 24] that the aggregate output (i.e. the average output protein concentration over multiple cells $Y_t = \frac{1}{N}\sum_i^N y_t^i$) can be controlled using the following stability criterion.

$$\Delta V_t^s = \frac{N_t^R}{N}\left[p - p^2\right] + \left[e_t - \frac{N_t^R}{N}p\right] - e_t^2 \le 0 \tag{4.12}$$

Where N is the number of cells N_t^R $e_t = Y_d - Y_t$.

4.2.5 Feedback Control of Heterogeneous Cell Populations

In many cases we cannot assume homogeneity of the cell because each cell has different initial condition and parameters. In this case the analytical expression for of the variance is not available. However if we use the central limit theorem to approximate the aggregate output as a normal distribution:

$$Z_N = \frac{1}{\sqrt{N}}\left[\sum_{i=1}^{N}(y_i - m_i)\right] \sim \mathbb{N}(0, Q), \text{ for } N \to \infty \tag{4.13}$$

A. MATLAB Codes

A.1 MATLAB SymBiology Code: Numerical Integration of Reaction Rate Equations + histogram generation

```
%%10/10/2011
% Created by Michaelle N Mayalu
% Toy Model for the intracellular signalling pathway: 2 stages with
feedback:using matlab SimBiology toolbox
%% Create Model
clear all
Toy2 = sbmlimport('Toy_model');
NA = 6.022e23; %%avagadro's number(molecules/mole)
Cell_Volume = 10e-12; %Liters
Matrix_Volume = 10e-6; %Liters
mconU =1e-12;
n_U = mconU*NA*Matrix_Volume;
tf = 15000 ;
%% Simulate for varied initial conditions p10,p20 and y10,y20 following a
u_variants = addvariant(Toy2, 'u_variants');
variants = addvariant(Toy2, 'variants');
addcontent(u_variants, {'species', 'U', 'InitialAmount', n_U});
u_v = [1.1,3,10];
input1 = {};
input2 = {};
for p = 1:length(u_v)
    rmcontent(u_variants,1);
    conU = u_v(p)*1e-12;
    nconU = conU*NA*Matrix_Volume;
    addcontent(u_variants, {'species', 'U', 'InitialAmount', nconU});
    %%uniform distribution
    n = 100; %% number of runs
    cs1 = getconfigset(Toy2);
    set(cs1, 'StopTime', tf); %set stop time
    set(cs1, 'SolverType', 'sundials')
    addcontent(variants, {{'species', 'p1', 'InitialAmount', 1e6},{'species',
'p2', 'InitialAmount', 1e6}...
        ,{'species', 'y1', 'InitialAmount', 1e4},{'species', 'y2',
'InitialAmount', 1e4}});
    ensemble1 = [];
    ensemble2 = [];
    rp1 = []; rp2 = [];
    ry1 = []; ry2 = [];
    for a = [1:n]
        rmcontent(variants,1);
        rmcontent(variants,1);
        rmcontent(variants,1);
        rmcontent(variants,1);
        rp1(a) = randi([1e3,1e8]); rp2(a) = randi([1e3,1e8]);
        ry1(a) = randi([1e3,1e5]); ry2(a) = randi([1e3,1e5]);
        addcontent(variants, {{'species', 'p1', 'InitialAmount',
rp1(a)},{'species', 'p2', 'InitialAmount', rp2(a)}...
            ,{'species', 'y1', 'InitialAmount', ry1(a)},{'species', 'y2',
'InitialAmount', ry2(a)}});
        [t_ode, x_ode, names] = sbiosimulate(Toy2,cs1,[variants,u_variants]);
        time = [0:100:tf];
        p1i = interp1(t_ode,x_ode(:,16),time);
```

```
        p2i = interp1(t_ode,x_ode(:,19),time);
        ensemble1 = [ensemble1,p1i'];
        ensemble2 = [ensemble2,p2i'];
    end
    input1{p} = {input1,ensemble1};
    input2{p}= {input2,ensemble2};
end

% Find the delay ascociated with responses
%delay p1

tau1_u =[];
tau2_u =[];
for p=1:length(u_v)
    for a = [1:n]
        p1 = input1{p}{2}(:,a);
        p1_lim = .5*p1(length(time));
        a_M = 1;
        while  abs(p1(a_M)) <= p1_lim
                a_M = a_M+1;
        end
        delta_tau = time(a_M) - time(1);
        i_delay_p1 = a_M;
        tau_p1(a) = time(a_M);
    end
    tau1_u =[tau1_u;tau_p1];
end

%delay p2
for p = 1:length(u_v)
    for a = [1:n]
        p2 = input2{p}{2}(:,a);
        p2_lim = .5*p2(length(time));
        a_M = 1;
        while  abs(p2(a_M)) <= p2_lim
                a_M = a_M+1;
        end
        delta_tau = time(a_M) - time(1);
        i_delay_p2 = a_M;
        tau_p2(a) = time(a_M);
    end
    tau2_u =[tau2_u;tau_p2];
end
for p = 1:length(u_v)
    tau_u1{p} = [-1000:mean(tau1_u(p,:))+1000];
    tau_u2{p} = [-1000:mean(tau2_u(p,:))+1000];
    f1{p} = ksdensity(tau1_u(p,:),tau_u1{p},'width',200);
    h1 = max(f1{p});
    b1 = max(hist(tau1_u(p,:)));
    f2{p} = ksdensity(tau2_u(p,:), tau_u2{p},'width',250);
    h2 = max(f2{p});
    b2 = max(hist(tau2_u(p,:)));
end

figure(3)
    plot(tau_u1{1},f1{1},'r-',tau_u1{2},f1{2},'m-',tau_u1{3},f1{3},'c-')
```

```
    title('Distribution of tau1 for p1 given u1 = 1pM')
    xlabel('tau1 seconds')
    legend('1','3','5','10')

    figure(4)
    plot(tau_u2{1},f2{1},'r-',tau_u2{2},f2{2},'m-',tau_u2{3},f2{3},'c-')
    title('Distribution of tau1 for p2 given u1 = 1pM')
    xlabel('tau1 seconds')
    legend('1','3','5','10')
```

A.2 MATLAB Code: Calculation of signal transduction time probabilities

```
%%4/20/2012
%%Created by Michaelle Mayalu

%% Calculation of signal transdution time probabilites
%%assumptions:
%%2) intial activated states zero
%%3) initial inactivated states nonzero
clear all
%% 1. compute valid sequences of leading to R--> X(state where X1* = 1)
% %% define parameters
c = 1; %% number of cascades
n = 3*c; %%number of reations and states
Rtype = [1:n]; %% vector containg reaction types
tf = 100; %sec
% S = perms(Rtype); %%matrix of  all possible reaction sequences
%% sequence generation
X = [];
X0 = [10 1  0 0 1 0]'; %%initial state
X(:,1) = X0;

c = [1 .1 1 1 .1 .01];
a = [];
% a(1,1) = c(1)*X(1,1)*X(2,1);
% a(2,1) = c(2)*X(3,1);
% a(3,1) = c(3)*X(3,1);
% a(4,1) = c(4)*X(4,1)*X(5,1);
% a(5,1) = c(5)*X(6,1);
% a(6,1) = c(6)*X(6,1);
% t_max = 1/max(a); %sec
n_total= 20; %ceil(tf/t_max)

v12 = ones(1,n_total);
 if mod(n_total,2) == 0
     last = [2:2:n_total];
     v12(2:2:n_total-2) = ones(size([2:2:n_total-2]))*2;
 else
     last = [2:2:n_total-1];
     v12(2:2:n_total-3) = ones(size([2:2:n_total-3]))*2;
 end
```

```
 for l = last
     v = zeros(1, length(l));
     v(l-1:l)= [1, 3];
     v(1:l-2) = v12(1:l-2);
     S{l}= v;
 end
S{last};

% if tf/t_max < 1;
%     display('final time(tf) is too small')
%     break
% end
% Rp = 0;
% n_total= 1; %ceil(tf/t_max)
%  for i = [1:n_total] ; %%time i
%      for j = [1:6] %%reaction j
%      if a(j,i) > 0 &&  Rp ~= j,
%         Rp = [Rp,j] %%save this reaction
%      end
%      end
%  end
%
%  for k = 1:length(Rp)
%  a(find(S(:,1) == Rp(i))
%  end


%% State Space Matrix
%X(1) - activated receptor, X(2) - inactivated protien, X(3)-activated
receptor:inactivated protien ,
%X(4)- activated protien, X(5)- phosphotase, X(6) - a
%% enumerate states;
A = zeros(6,6);
A(1,:) = [-1 -1 1 0 0 0];
A(2,:) = [1 1 -1 0 0 0 ];
A(3,:) = [1 0 -1 1 0 0];
A(4,:) = [0 0 0 -1 -1 1];
A(5,:) = [ 0 0 0 1 1 -1];
A(6,:) = [ 0 1 0 0 1 -1];
u = zeros(size(X0));
for i = last
    X(:,1) = X0;
    x = S{i};
    w = 1;
    for R = x;
        u = zeros(size(X0));
        u(R) = 1;
        X(:,w+1) = X(:,w) + A'*u;
        w = w +1;
    end
    X_X{i} =  X;
end


%% propenities for each state:
for i = last
```

```
    h = X_X{i};
    for t = 1:length(h(1,:))
        if X(1,t)>= 0 && X(2,t) >= 0
        a(1,t) = c(1)*X(1,t)*X(2,t);
        end
        if X(3,t)>= 0
        a(2,t) = c(2)*X(3,t);
        a(3,t) = c(3)*X(3,t);
        end
        if X(4,t)>= 0 && X(5,t) >= 0
        a(4,t) = c(4)*X(4,t)*X(5,t);
        end
        if X(6,t)>= 0
        a(5,t) = c(5)*X(6,t);
        a(6,t) = c(6)*X(6,t);
        end
    end
    alphas{i} = a;
end

%%
% sym t_o
% for i = last
%     C{i} = prod(sum(alphas{i}))
% end
% a_0 = [];
% res = [];
% for i = last
%     a_o = sum(alphas{i})
%     for t = 1:length(a_o)
% %         res(t) = 1/ a_o(
%
% end

for i  = last
    a_0{i} = sum(alphas{i});
    Q{i} = zeros(length(a_0{i}),length(a_0{i}));
    for y = 1:length(a_0{i})-1
        Q{i}(y,y) = -a_0{i}(y);
        Q{i}(y+1,y) = a_0{i}(y);
    end
    time = 0:.01:80;


end
% l{2} = string(g);
% l{4} = 'b';
% l{6} = 'r';
% l{8} = 'm';
% l{10} = 'c';
for i = last
     w = 1;
    for t = time
        E{i} = expm(Q{i}*t);
        p(w) = E{i}(length(a_0{i})-1,1);
        w = w+1;
```

```
    end
%     figure(1)
    P{i} = p;
%     plot(time, p/sum(p))
%     hold on
end
% xlim([0,500])
% ylim([0,.5])
% legend(num2str(S{2}),num2str(S{4}), num2str(S{6}),num2str(S{8}),
num2str(S{10}))
% xlabel('time', 'FontSize', 12)
% ylabel('\fontsize{12} f_T(t|S_i)','Interpreter','tex')
% % text(pi,0,' \leftarrow sin(\pi)','FontSize',18)

% for i = last
%     tk =0:length(P{i})-1;
%     plot(tk, P{i})
% end


alphas{i};

joint = 0;


for i = last
    prosalphas =
prod([alphas{i}(1,1:2:length(S{i})),alphas{i}(2,2:2:length(S{i})-1)]);
    joint = joint + P{i}*prosalphas;
end

tau1 = time;
tau2 = time;
yt1 = P{2};
yt2 = P{2};
i_k = [];
 pt1gt2 = [];
p =[];
          a = 1;
          while a <length(tau1)
              k = tau2(a);
              i_k = find(tau1>k);
              joint(i_k);
              pt1 = (joint(a))*joint(i_k);
              pt2(a) = sum(pt1);
              a = a+1;
          end
          p=sum(pt2)/(sum(joint))^2;

joint30 = joint/sum(joint);
figure(2)
plot(time, joint/sum(joint))
xlim([0,30])
% ylim([0,.35])
xlabel('time', 'FontSize', 16)
```

```
ylabel('Marginal occurance time Probability: \itf_T(t_n)', 'FontSize', 16)
hold on
%
save('JointP30', 'joint30')
```

Bibliography

[1] L. Lamalice, F. Le Boeuf and J. Huot, "Endothelial cell migration during angiogenesis," *Circ. Res.*, vol. 100, pp. 782-794, 2007.

[2] M. Kowanetz and N. Ferrara, "Vascular endothelial growth factor signaling pathways: therapeutic perspective," *Clinical Cancer Research,* vol. 12, pp. 5018-5022, 2006.

[3] L. B. Wood, "Quantitative Modeling and Control of Nascent Spout Geometry in In-Vitro Angiogenesis," .

[4] A. L. Bauer, T. L. Jackson, Y. Jiang and T. Rohlf, "Receptor cross-talk in angiogenesis: Mapping environmental cues to cell phenotype using a stochastic, Boolean signaling network model," *J. Theor. Biol.*, vol. 264, pp. 838-846, 2010.

[5] F. Gaits and K. Hahn, "Shedding light on cell signaling: interpretation of FRET biosensors," *Sci. STKE,* vol. 2003, pp. PE3, Jan 14, 2003.

[6] M. Ouyang, J. Sun, S. Chien and Y. Wang, "Determination of hierarchical relationship of Src and Rac at subcellular locations with FRET biosensors," *Proceedings of the National Academy of Sciences,* vol. 105, pp. 14353, 2008.

[7] J. Seong, M. Ouyang, T. Kim, J. Sun, P. Wen, S. Lu, Y. Zhuo, N. M. Llewellyn, D. D. Schlaepfer, J. Guan, S. Chien and Y. Wang, "Detection of focal adhesion kinase activation at membrane microdomains by fluorescence resonance energy transfer " *Nature Communications,* vol. 2, pp. 406, 2011.

[8] H. H. Asada, Y. X. Wang and M. N. Mayalu, "Molecular signaling observer and predictor: A framework for closed-loop control of cell behaviors having long time delay," in *American Control Conference (ACC), 2011,* 2011, pp. 2314-2319.

[9] S. Park, H. H. Jung, Y. H. Park, J. S. Ahn and Y. H. Im, "ERK/MAPK pathways play critical roles in EGFR ligands-induced MMP1 expression," *Biochem. Biophys. Res. Commun.*, 2011.

[10] M. Kobayashi, M. Nishita, T. Mishima, K. Ohashi and K. Mizuno, "MAPKAPK-2-mediated LIM-kinase activation is critical for VEGF-induced actin remodeling and cell migration," *EMBO J.*, vol. 25, pp. 713-726, 2006.

[11] A. R. Asthagiri and D. A. Lauffenburger, "A Computational Study of Feedback Effects on Signal Dynamics in a Mitogen-Activated Protein Kinase (MAPK) Pathway Model," *Biotechnol. Prog.*, vol. 17, pp. 227-239, 2001.

[12] J. S. Krueger, V. G. Keshamouni, N. Atanaskova and K. B. Reddy, "Temporal and quantitative regulation of mitogen-activated protein kinase (MAPK) modulates cell motility and invasion." *Oncogene,* vol. 20, pp. 4209, 2001.

[13] R. Heinrich, B. G. Neel and T. A. Rapoport, "Mathematical models of protein kinase signal transduction," *Mol. Cell,* vol. 9, pp. 957-970, 2002.

[14] D. J. Wilkinson, "Stochastic modelling for quantitative description of heterogeneous biological systems," *Nature Reviews Genetics,* vol. 10, pp. 122-133, 2009.

[15] A. J. Ridley, M. A. Schwartz, K. Burridge, R. A. Firtel, M. H. Ginsberg, G. Borisy, J. T. Parsons and A. R. Horwitz, "Cell migration: integrating signals from front to back," *Science,* vol. 302, pp. 1704, 2003.

[16] H. Lodish and J. E. Darnell, *Molecular Cell Biology.* Wiley Online Library, 2000.

[17] C. V. Carman and T. A. Springer, "Integrin avidity regulation: are changes in affinity and conformation underemphasized? " *Curr. Opin. Cell Biol.,* vol. 15, pp. 547-556, Oct, 2003.

[18] S. S. Andrews and D. Bray, "Stochastic simulation of chemical reactions with spatial resolution and single molecule detail," *Physical Biology,* vol. 1, pp. 137, 2004.

[19] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *J. Phys. Chem.,* vol. 81, pp. 2340-2361, 1977.

[20] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of Computational Physics,* vol. 22, pp. 403-434, 2005.

[21] Y. I. Wu, D. Frey, O. I. Lungu, A. Jaehrig, I. Schlichting, B. Kuhlman and K. M. Hahn, "A genetically encoded photoactivatable Rac controls the motility of living cells," *Nature,* vol. 461, pp. 104-108, 2009.

[22] H. Kushner, "Stochastic stability," *Stability of Stochastic Dynamical Systems,* pp. 97-124, 1972.

[23] L. B. Wood and H. H. Asada, "Cellular Stochastic Control of the Collective Output of a Class of Distributed Hysteretic Systems," *Journal of Dynamic Systems, Measurement, and Control,* vol. 133, pp. 061011, 2011.

[24] L. B. Wood, A. Das, R. D. Kamm and H. H. Asada, "A stochastic broadcast feedback approach to regulating cell population morphology for microfluidic angiogenesis platforms," *Biomedical Engineering,* IEEE *Transactions on,* vol. 56, pp. 2299-2303, 2009.

[25] H. Ai, K. L. Hazelwood, M. W. Davidson and R. E. Campbell, "Fluorescent protein FRET pairs for ratiometric imaging of dual biosensors," Nature Methods, vol. 5, pp. 401-403, 2008.

[26] M. Akkouchi, "On the convolution of exponential distributions," J.Chungcheong Math.Soc, vol. 21, pp. 501, 2008.

[27] S. V. Amari and R. B. Misra, "Closed-form expressions for distribution of sum of exponential random variables," Reliability, IEEE Transactions on, vol. 46, pp. 519-522, 1997.

[28] F. Gaits and K. Hahn, "Shedding light on cell signaling: interpretation of FRET biosensors," Sci. STKE, vol. 2003, pp. PE3, Jan 14, 2003.

[29] H. Gerhardt, "VEGF and endothelial guidance in angiogenic sprouting," VEGF in Development, pp. 68-78, 2008.

[30] H. Gerhardt, M. Golding, M. Fruttiger, C. Ruhrberg, A. Lundkvist, A. Abramsson, M. Jeltsch, C. Mitchell, K. Alitalo and D. Shima, "VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia," J. Cell Biol., vol. 161, pp. 1163-1177, 2003.

[31] B. D. Gomperts, I. J. M. Kramer and P. E. R. Tatham, "Signal Transduction (2nd Edition)," .

[32] I. Goryanin, T. C. Hodgman and E. Selkov, "Mathematical simulation and analysis of cellular metabolism and regulation." Bioinformatics, vol. 15, pp. 749, 1999.

[33] J. A. Gubner, Probability and Random Processes for Electrical and Computer Engineers. Cambridge University Press, 2006.

[34] P. A. Iglesias and B. P. Ingalls, Eds., Control Theory and Systems Biology. Cambridge, MA: MIT Press, 2010.

[35] V. S. Kraynov, "Localized Rac Activation Dynamics Visualized in Living Cells " Science, vol. 290, pp. 333 <last_page> 337, 2000.

[36] H. Lodish and J. E. Darnell, Molecular Cell Biology. Wiley Online Library, 2000.

[37] T. Lu, T. Shen, C. Zong, J. Hasty and P. G. Wolynes, "Statistics of cellular signal transduction as a race to the nucleus by multiple random walkers in compartment/phosphorylation space," Proceedings of the National Academy of Sciences, vol. 103, pp. 16752-16757, 2006.

[38] S. Lu, T. J. Kim, C. E. Chen, M. Ouyang, J. Seong, X. Liao and Y. Wang. Computational analysis of the spatiotemporal coordination of polarized PI3K and Rac1 activities in micro-patterned live cells PLoS One 6(6), pp. e21293. 2011. Available: www.refworks.com.

[39] H. H. McAdams and A. Arkin, "Stochastic mechanisms in gene expression," Proceedings of the National Academy of Sciences, vol. 94, pp. 814, 1997.

[40] K. Moissoglu and M. A. Schwartz, "Integrin signalling in directed cell migration " Biol. Cell., vol. 98, pp. 547-555, Sep, 2006.

[41] M. Ouyang, J. Sun, S. Chien and Y. Wang, "Determination of hierarchical relationship of Src and Rac at subcellular locations with FRET biosensors," Proceedings of the National Academy of Sciences, vol. 105, pp. 14353, 2008.

[42] A. J. Ridley, M. A. Schwartz, K. Burridge, R. A. Firtel, M. H. Ginsberg, G. Borisy, J. T. Parsons and A. R. Horwitz, "Cell migration: integrating signals from front to back," Science, vol. 302, pp. 1704, 2003.

[43] L. D. Schmidt, "Engineering of Chemical Reactions (2nd Edition)," .

[44] B. Schoeberl, C. Eichler-Jonsson, E. D. Gilles and G. Muller, "Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors," Nat. Biotechnol., vol. 20, pp. 370-375, 2002.

[45] J. Seong, S. Lu and Y. Wang, "Live Cell Imaging of Src/FAK Signaling by FRET " Cell. Mol. Bioeng., vol. 2, pp. 138-147, Jun 1, 2011.

[46] A. Tomar and D. D. Schlaepfer, "Focal adhesion kinase: switching between GAPs and GEFs in the regulation of cell motility," Curr. Opin. Cell Biol., vol. 21, pp. 676-683, 2009.

[47] A. Tomar and D. D. Schlaepfer, "Focal adhesion kinase: switching between GAPs and GEFs in the regulation of cell motility " Curr. Opin. Cell Biol., vol. 21, pp. 676-683, Oct, 2009.

[48] A. Uhrmacher, D. Degenring and B. Zeigler, "Discrete event multi-level models for systems biology," Transactions on Computational Systems Biology I, pp. 133-133, 2005.