Regulation of T-Cell Signaling Networks

by

Arvind Shankar Prabhakar

B.Tech., Indian Institute of Technology Madras (2006)
S.M., Massachusetts Institute of Technology (2008)

Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Chemical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013

Author .....
Department of Chemical Engineering
May 21, 2013

Certified by .....
Arup K. Chakraborty
Director of The Institute for Medical Engineering and Science
Robert T. Haslam Professor of Chemical Engineering
Professor of Physics, Chemistry and Biological Engineering
Thesis Supervisor

Accepted by .....
Patrick S. Doyle
Professor of Chemical Engineering
Chairman, Committee for Graduate Students

This doctoral thesis has been examined by a Committee of the Department of Chemical Engineering as follows:

Professor Narendra Maheshri .....
Chairman, Thesis Committee
Assistant Professor of Chemical Engineering

Professor Arup K. Chakraborty .....
Thesis Supervisor
Director of The Institute for Medical Engineering and Science
Robert T. Haslam Professor of Chemical Engineering
Professor of Physics, Chemistry and Biological Engineering

Professor Mehran Kardar .....
Member, Thesis Committee
Francis Friedman Professor of Physics

Abstract

For a better understanding of how biology carries information within cells, it is not sufficient to look at individual protein or gene interactions, but to understand these networks of interactions as a whole. The goal of this thesis is to understand various aspects of how cells in general and T-cells in particular function, using models built from basic principles in chemical engineering, statistical physics and network theory, together with experiments performed by our collaborators. The ultimate objectives are to gain an insight into the mechanisms of certain key biological processes, understand the cause of certain diseases and to generate new ideas for methods of treating these diseases.

First, we look at an example of a specific network built from previously published experiments and data collected by our collaborators, which governs the mechanism of activation of the T-cell receptor (TCR) by its kinase Lck and a negative regulator of Lck called Csk. We show that the mechanism by which the cell regulates TCR levels, together with the manner in which Lck activates the TCR produces interesting behavior, such as a "perfectly adaptive" system and a high-pass filter.

Second, we look at heterogeneity in cancer cells at the level of protein signaling networks. Many common cancers are not treatable at the "source" or initial mutation, so one has to target downstream effector molecules. However, different cell lines bearing the same initial cancerous mutation exhibit varying signaling patterns due to differing secondary mutations which makes this difficult. The objective of this project is to try to characterize this heterogeneity and be able to identify molecules in the cell which would be the most effective drug targets. A general model for signaling in networks has been developed, analogous to models of neural networks, with mutations modeled as changes in the topology of this network. Keeping in mind that cancer cells are trying to maximize their growth, we are looking for patterns in secondary mutation during the directed evolution of these networks. A method for looking at free energy landscapes in topology space has also been developed. We find that lowest degree nodes along the shortest paths from the driver mutation to effector nodes tend to be the most conserved, and the frequency of multiple optima depends on the number of feedback loops.

Finally, we look at the problem of constant activation thresholds for activation of various types of T-Cells. Despite having different TCRs, T-Cells of a certain type have a fixed activation threshold in terms of a peptide-MHC interaction strength (and a corresponding time, earlier than which they do not activate). We built a reaction-diffusion model for the network involved in the search process by which a pMHC-TCR finds a coreceptor-Lck, which enables us to understand how the threshold for activation is determined by the parameters of a particular cell type. We also developed an analytical solution for a simplified Markov Chain form of the model, which predicts how the activation rate scales with the parameters of interest in the system. We find that this rate is proportional to the fraction of coreceptors with Lck,

increases (slowly) with diffusion and is independent of the number of coreceptors on the surface of the cell.

Acknowledgements

This thesis is dedicated to my parents.

Graduate school has taken up many years of my life. It has been a time of discovery for me, both scientific and personal, and it would not have been possible without the love and unconditional support of many, many people. I am grateful to everyone who has helped make me what I am today, and am going to try to list a few of the most important.

First, I would like to thank my advisor, Arup Chakraborty, for all the support he has given me. Working in the Chakraborty lab has been an eye-opener; the amount of freedom we get in terms of defining problems and solving them in our own way is fantastic. Arup is always patient, willing to help and understanding of all the problems that crop up in research. Working with my thesis committee, Narendra Maheshri and Mehran Kardar, has also been a pleasure, and they provided a lot of useful input during my meetings. I would also like to acknowledge my collaborators, particularly Jeroen Roose and Ed Palmer, who were a pleasure to work with and always answered my stupid questions about biology with infinite patience. Thanks must also go to all my previous teachers and professors, too many to list, who helped shape me intelectually.

I consider the members of the Chakraborty group some of the most intelligent, driven and fun people I've ever worked with, and I'd like to acknowledge a few of them in particular: Misha, Huan, Chris, and Steve. Huan and Chris were always there to lend an ear, listen to my problems and gripes and offer (usually sage) advice; Misha, for helping me through every computer-related problem I had (and there were a ton of them, given how I liked pushing the limits of our computing resources); and Steve, who is the best labmate one could ever hope to have, inspiring everyone with his enthusiasm and love for science.

My friends here in Cambridge through the years have been my greatest support on a daily basis. I would like to thank them all, especially Ganesh, Krupa, Murali, Anuja, Karthik, Sowmya, Anna, Aruna, Sanjeev, Sudha and Jitendra for being there for me, every time I was feeling down or even just wanted to get out and do something. I will always cherish the times we spent together, exploring the world, cooking together, or just hanging out. I would like to acknowledge my wonderful roommate for four years of graduate school, Jonathan Derocher, who has also turned out to be a great friend. The rest of the Chemical Engineering class that entered in 2006, helped me get through the first hectic year and adapt to a new culture and stage of my life. I'll miss you all, especially the nighttime trivia sessions at the Thirsty.

My means of escape from the daily grind during my graduate school years has been sport. The camaraderie on the sporting field is second to none, and all the people I have played sport with, both on the ChemE and Tang teams, have been fantastic. I loved every moment of it - the soccer and tennis, and am eternally grateful to everyone who helped me learn new sports despite a complete lack of talent - I would have never been able to learn to play softball and ice hockey otherwise.

Growing up in India, it is impossible to overestimate the influence of family members on one's life. I am eternally grateful to my parents and grandparents, for all their love and affection, in bringing me up and making me who I am today. To all my extended family, particularly those in the US: thanks for welcoming me to this new country, listening to me complain on the phone or in person, providing me delicious food and a sense of home away from home on the occasions I decided to visit. A special note of thanks must go to my cousins Visi and Rajeswari and their families, who, being in the Boston/Cambridge area, were always there when I needed them. I am also very grateful to my cousins, Lakshmi and Lalitha Sundaram, for all their long distance support - Lalitha, especially, has been great to talk to, always understanding and helpful in dealing with both the frustrations of graduate school and our family.

And finally, my work on cancer is in memory of my aunt Prabha Sundaram. Her diagnosis with breast cancer was a major factor in me taking up this project, and her subsequent passing drove me harder, and helped me always keep in mind that, at the end of the day, the ultimate objective of our work is to have a real human impact and give something back to society. I realize I am but a drop in the ocean of science, but all my years of toil will have been worth it if even in some miniscule way it leads to somebody else having a better tomorrow.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis attempts to describe several years' worth of work into understanding the immune system using models built from a knowledge of the underlying physics, computational tools and simulations. The immune system is one of the most complex areas of biology, as it involves the body's response to internal and external threats, making it central to the understanding of disease. This has led to it being at the forefront of a great amount of research, both of fundamental questions of biology as well as more applied medical-type translational research. With the emergence of better experimental techniques in biology and the collection of large amounts of data of multiple types across various experiments, it is important to be able to identify the essential features of the underlying systems which result in the observed phenotypes. Biophysics, which is the broad field in which most of the work involved in this thesis, is one such way of doing it, by combining physics-based models with computer simulations, bioinformatics and other experimental methods. This work focuses on building physics-based models of biological networks involved in the human immune system, and tries to answer some questions of biological significance.

1.1 The immune system

The immune system plays a critical role in the survival of higher organisms. It protects the host from a variety of external factors, such as invading bacteria and viruses, using a wide range of defensive mechanisms. The immune system is divided into two "parts", the innate immune system which is built to recognize common patterns of pathogens and destroy them, and the adaptive immune system which provides a pathogen-specific response along with "remembering" former invaders so that responses to repeat attacks can be efficient. One of the key components of the adaptive immune system is the T-cell, which detects antigen presented (in the form of short peptide fragments derived from the pathogen) on the surface of antigen presenting cells (APCs). The strength of the interaction between the peptide bound to the major histocompatibility complex (MHC) molecule on the APC and the T-Cell receptor (TCR) is an indicator of whether the peptide is derived from foreign or self; thymic selection tries to ensure that strong pMHC-TCR interactions imply that the peptide is foreign and a response needs to be mounted. This interaction along with a few others, leads to a chain of signal transduction events in the form of protein (chiefly kinases) interactions and phosphorylations within the T-Cell that determine the response to be mounted. In this work, we look at problems involving the signal transduction pathways post-interaction of the TCR and pMHC and try to make simple biophysical models that capture experimental results and help us understand the biology of the processes involved. Sections 3 and 4 of this thesis deal with the regulation of Lck, a key kinase involved in phosphorylation of the TCR; section 6 deals with the search process by which the TCR-pMHC finds Lck bound to a coreceptor molecule. Section 2 deals with a model of scaffolds, which are a ubiquitous feature of signaling networks. Finally, section 5 deals with a more coarse-grained model of networks and tries to understand how mutations affect them; this is inspired by the problem of signaling heterogeneity seen in T-Cell lymphomas.

1.2 TCR Signaling: Overview of the biology

During TCR-pMHC engagement, the kinase Lck which is bound to the CD4 or CD8 coreceptor gets recruited to the TCR-complex and is activated. Lck phosphorylates the immunotyrosine activation motifs (ITAMs) within the CD3- and ζ-chains of the TCR complex[1]. Doubly phosphorylated ITAMs bind the kinase ZAP-70([2][3]), which is then activated by phosphorylation, either by other ZAP-70 molecules or Lck. Activated ZAP-70 can then phosphorylate tyrosines on an adaptor molecule called linker for activation of T cells (LAT)[4]. LAT contains nine conserved tyrosines that when phosphorylated, bind to the SH2 domains of several other proteins to assemble a signaling complex called the signalosome. The LAT signalosome involves molecules such as Gads, SLP-76, Itk, Grb2, SOS, PLCγ1, Vav, etc.([5][6][7][8][9]) Some of these interactions stabilize the resulting complex (for example, PLCγ1-SLP76 and Grb2-SOS. Grb2-SOS enables the formation of LAT clusters, stabilizing the signalosome further). This SOS that is bound to Grb2 activates Ras (converting RasGDP to Ras-GTP), which in turn activates Raf and sets off the MAPK (ERK) signaling pathway. Ras can also be activated by the guanine exchange factor RasGRP[10] and is deactivated by RasGAPs[11].

Protein kinase cascades have been implicated in the processes involved in thymic selection, T-cell activation and function[13][14]. The three major groups of mitogen-activated protein (MAP) kinases in mammalian cells are the extracellular signal-regulated protein kinases (ERK), the p38 MAP kinases and the c-Jun NH2-terminal kinases (JNK)[15][16]. These protein kinases are activated by the same protein kinase cascades that lead to MAP kinase activation in other cell types. However, the mechanisms that activate these cascades are distinctive in T cells and are still not fully understood. The result of signal transduction through these cascades is the activation of certain transcription factors[17], leading to transcription, translation and effector functions such as secretion of cytokines. Another complication is that proteins are constantly created by ribosomal synthesis and degraded (mediated by Cbl, for exam-

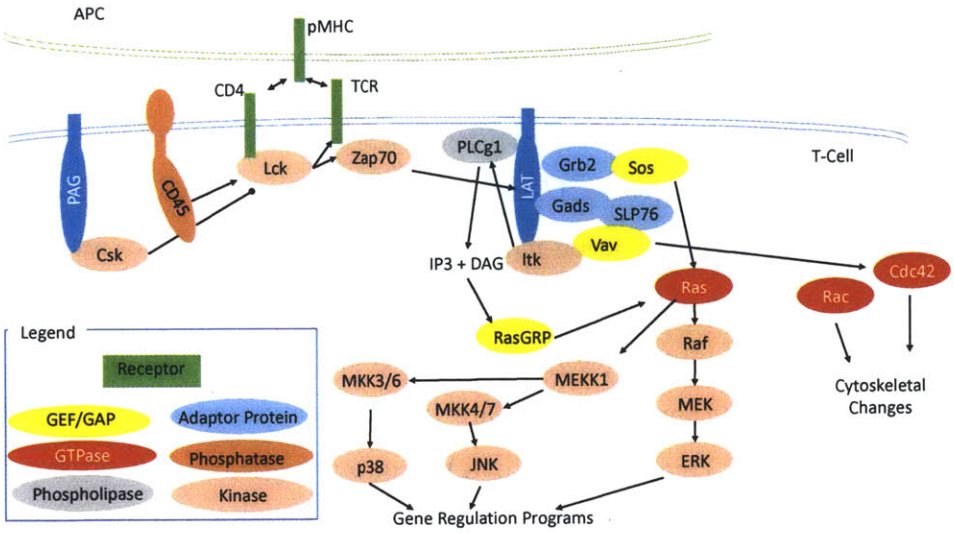


Figure 1-1: A part of the T-Cell signaling network (See [12] for a more complete figure). Most of the work done before, and the focus of this thesis, is on early (membrane-proximal) signaling events: the TCR-pMHC interaction, coreceptor CD4 and Lck regulation. These lead to formation of the LAT complex, activation of Ras, signaling through the MAP Kinase cascades and gene transcription.

ple). A cartoon of the TCR signaling pathway can be found in Figure 1.2.

1.3 Scope of the work and choice of methods

Immune responses are one of the most complex behaviors of biological systems, spanning many time and length scales, with feedback from larger to smaller scales. Interactions important in just adaptive immune responses range from sub-molecular interactions (for example, between the TCR and peptide-MHC), to interactions between proteins (in signaling networks, for example, which comprises the major portion of this thesis), between cells (cytokines) and so on. Time scales range from nanoseconds for protein conformational changes to years (lifetimes of memory T Cells). Modeling such responses, therefore, involves making mathematical models of complex processes with many interactions and feedbacks between them. There are a large number of experimental results emerging daily in this field; one of the key challenges in this field is

understanding all these different experiments and building simple models that provide a unified explanation of these phenomena. Such simple models are valuable because they strip away superfluous details and get to the key features of the system which are most important in producing the phenomena observed. These underlying models are built from basic first-principles physics, such as statistical mechanics, chemical kinetics and so on that all systems must obey; using such physics-based approaches ensures that the models developed are based on a sound theoretical footing and are not just an exercise in fitting parameters to data.

The scope of this work does not involve the development of new mathematical methods, but rather in using results and methods which were originally deployed in various fields such as financial analysis, stochastic processes, control theory, machine learning, and neurobiology to understanding aspects of the regulation of the immune systems. The synthesis of computational techniques, together with biological data and first-principles physics can produce novel insight and help unify varied biological experiments performed in slightly different contexts by highlighting the key aspects of the system which cause such phenomena. This is a critical first step in the rational design of drugs and other therapies which have real-world impact.

Biological systems consist of many interacting components: they are an example of a network. Networks are pervasive in biology, typically ranging from interactions between proteins (PPI networks[18][19]) to form molecules to interactions between species in an ecosystem (a food web[20]). Other common examples are gene interaction networks[21], metabolic networks[22] and neural networks[23].

1.4 Simulation and Analysis of Chemical Reaction Networks

Previously in the Chakraborty group, some of the chemical reaction networks involved in early T-Cell signaling events have been studied computationally. Using

lists of chemical reaction networks built from the literature, together with the corresponding parameters (both from the literature and heuristic estimates), simulations of early TCR-related signaling events have been performed[24][25][26]. For example, an enzyme binding to a substrate would be one reaction, the unbinding would be another and conversion of substrate to product would be a third reaction. Simulations could show how the rate of production of substrate varied with, say, the number of enzyme and substrate molecules. This generic scheme of an enzyme and substrate corresponds to many examples in the biology of early T-Cell activation: Csk and Lck, MEK and ERK, and so on.

Due to the fact that small numbers of molecules are involved (milli- and micromolar concentrations of proteins in cells), over small time and length scales (typically nanometers to microns), these systems cannot always be treated simply using Ordinary Differential Equations (ODEs) because of the discrete nature of molecules over these scales. Solving ODEs using tools like MATLAB gives mean concentrations that are the solutions at various time points, which may not be physically meaningful due to the unique nature of biochemical systems.

1.4.1 The Gillespie Algorithm

An approach that can be used to overcome the problems with ODE-based simulation of biochemical systems is stochastic simulation. It treats each molecule as a unit, instead of grouping them all in terms of a single "concentration". This affords us several advantages:

1. Small molecule numbers: In biological systems with small numbers of molecules, the responses to systems can be driven by stochastic effects and be very different from the mean-field solution of the underlying equations[27].

2. Many different trajectories: In the case of networks with multiple solutions (bistabilities, for example) which lead to cellular decisions[28], deterministic solutions do not provide us with an accurate description of the system. Different

cells represent individual realizations of the underlying network, and stochastic simulations can shed light on features such as the distribution of populations in each steady state and stochastic transitions between these populations.

3. Spatial distributions: For spatially distributed systems, even those without bistabilities, the exact solution involves dealing with coupled nonlinear partial differential equations (derived from the diffusion equation). This is not easy, and it can frequently be computationally easier to perform stochastic simulations.

In practice, the Gillespie algorithm[29] is used to simulate the evolution of biochemical systems due to its ability to take into account stochastic effects. For each reaction in a network, a rate equation1.1 can be written, in which a_i represents the propensity of reaction i, k_i the rate constant of reaction i, C_j the concentration of species j and ν_{ij} the stoichiometric coefficient of species j in reaction i.

$$a_i = k_i \prod_j C_j^{\nu_{ij}} \tag{1.1}$$

At each time step in the algorithm, the total propensity (a_0) is computed by adding all of the individual reaction propensities. Next, two random numbers are drawn from a uniform distribution from 0 to 1. The first random number (r_1) is used to compute the time that has elapsed since the last reaction occurred at the previous time step (τ) using equation 1.2, and the second random number r_2 is used to determine the identity μ of the next reaction using equation 1.3.

$$\tau = \frac{1}{a_0} \log\left(\frac{1}{r_1}\right) \tag{1.2}$$

$$\sum_{j=1}^{\mu-1} \frac{a_j}{a_0} < r_2 < \sum_{j=1}^{\mu} \frac{a_j}{a_0} \tag{1.3}$$

Once the identity of the next reaction is determined, the numbers of molecules for all species are changed according to the stoichiometry of whichever reaction has occurred. The total time is kept track of by summing the time elapsed between reactions. All reaction propensities are then recalculated and the process continues

until specified stopping criteria are met. The random element in which the next reaction is determined by chance (Equation 1.3) allows for realistic stochastic simulations.

Previous members of our group have used these methods to investigate certain aspects of the network involved in early T-Cell signaling. Early work in this area by members of our group focused on understanding the role of the site of interaction between a T-Cell and APC, called the immunological synapse[26][30][31]. Another problem in this field was understanding the sensitivity of this activation process, how a TCR could find a small number of foreign peptides in a sea of endogenous ones, which was looked at using a model called the dimer model[32]. Work was done on slightly more downstream parts of the signaling network as well: on Ras signaling[25] and formation of the LAT cluster[24].

Some work in development of methods in this area has been performed as well by former members of our group. Dennis Wylie developed algorithms for spatially heterogeneous systems where spatial effects of signaling matter[33]. Max Artyomov and Mieszko Lis developed an efficient tool for stochastic simulation of reaction-diffusion processes[34]. We use the spatial Gillespie algorithm to look at a problem involving threshold ligands in T Cells in Chapter 6 of this thesis.

1.4.2 Models of signaling networks

Network theory is a field of physics that aims to characterize and investigate the properties of various kinds of networks[35]. The identification of the structure of biological networks is an ongoing area of research. Networks are characterized in terms of properties such as degree distribution, clustering coefficient and so on. Due to the complexity of biological networks, they are typically divided into modules called "motifs," with each motif assumed to have certain unique characteristics[36][37]. These network motifs shape the spatio-temporal properties of the signals transmitted by these networks[38] and are thought to provide specific biological functions[39]. Models

of networks can be adapted to fit in with common statistical-mechanical models[40]; for example, one can write down the equations for the states of memories encoded by neural networks in terms of a Hamiltonian energy dependent on the interactions of the network[41][42].

One of the problems with the analysis of signaling networks is the great complexity in modeling them, both in the behavior of individual nodes and interactions between them. Also, a lot of details about these networks are not known, not just in the parameters that describe these systems but largely even in the topologies of these networks themselves. Simplified forms of nodes and interactions are therefore used to model biological signaling networks. The simplest form are Boolean networks[43][44] where the outputs to each node are just 0 or 1 and are a function of the input to the node, with edge strengths also being binary. Bayesian networks, which are a probabilistic model of activation based on parents of a node in the graph, are also commonly used[45][46][47]. These models cannot account for input-output relationships within nodes, however, and one needs to use more complex models for that purpose, for example from neural networks. Signaling models based on neural networks have therefore been used before[48] and we adapt them to study the statistical effect of mutations in signaling networks in the context of a form of cancer(Section 5). Flux-balance analysis[49][50] is another form of network modeling that is typically used in metabolic networks, but it only gives the rates of each chemical reaction rather than the concentration of each species of interest in the network. The problem of inferring networks from expression data is also a well-studied problem[51][52]. These problems usually just involve the identification of edges of the network, typically without consideration of edge strengths. Most techniques developed to solve this problem use probabilistic networks and involve machine learning and regression, usually using Bayesian networks (which cannot involve loops)[53], correlations[54], simplified dynamic models[55][56] or trees[57]. Another common problem is that of identifying similar nodes or groups of nodes[58][59]. In Chapter 5, we describe how some of these models can be used to study the effect of changing network topology

by mutations in a problem inspired by cancers of T-Cells.

1.4.3 Markov Processes

A Markov process is a stochastic process satisfying the property that the conditional probability distribution of future states of the process depends only on the current state, i.e. it is "memoryless." Conditional on the present state of the system, its future and past are independent. Markov processes can be continuous- or discrete-space and continuous- or discrete- time. In the case of interest, of chemical reaction networks where we are interested in counts of individual molecules (rather than bulk concentrations), the natural result is to model the system as a discrete-space, continuous-time Markov process (a Markov Chain).

Typically in the case of chemical reaction networks, the set of numbers and locations of molecules in the system completely specifies the propensities of all reactions that this system may undergo; this means that the current state of the system completely defines the rates of all transitions - making it a Markov process. This complete specification, however, usually requires that a lot of variables be specified, making it impractical in a lot of cases. This dramatic expansion of the state-space of the system usually requires that some sort of approximations be made in order to model a system effectively using a Markov model. Markov Chains are typically represented as directed graphs, where the nodes are different states and edges represent the probabilities (or rates) of transitions between states.

Let $x \in \{X_1, X_2, \ldots X_n\}$ be the set of states of the Markov Chain, and $P(x_i)$ the probability of the system being in state x_i. Let the rate of transition from state i to j be given by γ_{ij}. Then Equation 1.4 describes the evolution of the probability of state i.

$$\frac{d}{dt}P(x_i) = -\sum_{k \neq i} \gamma_{ik} P(x_i) + \sum_{k \neq i} \gamma_{ki} P(x_k) \tag{1.4}$$

This can be simplified to Equation 1.5, where $\mathbf{\Gamma}$ represents a matrix of transition rates.

$$\frac{d}{dt}P(\underline{x}) = \mathbf{\Gamma}P(\underline{x}) \tag{1.5}$$

Because of the form of Equation 1.5, the solution is a sum of exponentials (i.e. the system can be analyzed by looking at the eigenvalues of $\mathbf{\Gamma}$.

In Chapter 6, we use Markov Chains to coarse-grain a reaction-diffusion process and compute the rate of activation of different types of T-Cells.

1.4.4 Dynamical Systems Theory

A dynamical system is system where a fixed rule describes the time evolution of a point in a state space. At any given time a dynamical system has a state given by the coordinates of a point in an appropriate state space. The evolution rule of the dynamical system is a fixed rule that describes how the current state evolves to a future state and is deterministic[60]. Chemically reacting systems are examples of dynamical systems: the state of a system is completely defined by the set of all molecules in the system (for a homogeneous system; for a spatially distributed system, one must also specify the location of each molecule). Dynamical systems may be either continuous-state (flows) or discrete-state (maps), but converting the probabilistic nature of what reaction occurs next into a dynamical systems framework is not possible (because evolution must be deterministic in this framework). Therefore, dynamical systems theory can only be applied to systems with large numbers of molecules that are in the continuum limit (or we make such an approximation).

One can describe the evolution of a chemical reaction network in terms of the set of rate laws that govern the evolution of the system from Equation 1.6, where the a_js are propensities of individual reactions described by Equation 1.1 (this is an example

of a homogeneous reaction network).

$$\frac{dC_i}{dt} = \sum_j \nu_{ij} a_j \tag{1.6}$$

This takes the form of a set of coupled nonlinear ODEs. We can use these equations to analyze the general behavior, i.e. the number and stabilities of solutions, of the system of equations. To analyze the stabilities of a solution (fixed point) of the system, we do a linear stability analysis. Typically the evolution of the system is described the set of coupled nonlinear equations given in Equation 1.7, where $\underline{x}$ represents the states and $\underline{k}$ the parameters.

$$\frac{d\underline{x}}{dt} = \underline{f}\left(\underline{x}, \underline{k}\right) \tag{1.7}$$

Let $\underline{x^*}$ be the fixed point of interest. We expand the solution around the fixed point by an increment $\underline{\delta x}$:

$$\frac{d\left(\underline{x^*} + \underline{\delta x}\right)}{dt} = \underline{f}\left(\underline{x^*} + \underline{\delta x}, \underline{k}\right) = f\left(\underline{x^*}, \underline{k}\right) + \underline{\delta x}\mathbf{J}\left(\underline{f}\left(\underline{x}, \underline{k}\right)\right)|_{x=x^*} + \text{Higher order terms} \tag{1.8}$$

The first term is zero, as $\underline{x^*}$ is a solution (fixed point) of $\underline{f}\left(\underline{x}, \underline{k}\right) = 0$; the second term is the Jacobian of the system of equations given by $\underline{f}\left(\underline{x}, \underline{k}\right)$ evaluated at $\underline{x^*}$. Stability is indicated by the eigenvalues of the Jacobian matrix; one or more positive eigenvalues of the Jacobian imply an unstable system. Complex conjugate pairs of eigenvalues indicate oscillatory solutions.

Possible behaviors of the system are:

- Only one stable solution: The simplest case, the system has only one stable steady state.
- More than one stable fixed point: The system can exist in multiple states; this is possible when there is a switch, for example, and used in cellular decision-making[61]
- Only an unstable fixed point, with a complex conjugate pair of eigenvalues:

The system exhibits oscillations, such as those seen in Calcium in many cell types[62].

The tools of dynamical systems theory are typically used to understand the nature of the system at long times (steady-state). Biologically, most systems are not at steady-state; however, this assumption is quite frequently made in order to simplify problems and gain an understanding of the behavior of the system. Deterministic chemical systems can also exhibit chaos[63]. The framework of dynamical systems theory is used to examine two problems pertaining to the regulation of Lck in Chapters 3 and 4.

Chapter 2

The Effects of Mutations on Scaffolds

2.1 Introduction

The motif of sequential activation of multiple protein kinases is a common one in biology and is known to regulate many important cellular decisions([64], [65], [66]). A well-studied example of this motif is the MAPK cascade, in which the kinases are associated with a scaffold protein KSR. The scaffold protein is known to be required for the activation of ERK in T Cells [67], and can help amplify or attenuate the signal, depending on the context [68]. For systems involving a set of kinases modulated by a scaffold protein, there is a well-known combinatorial inhibition or "pro-zone" effect [69] whereby titration of the amount of scaffold in a system yields a bell-shaped curve in the output signal of the cascade. In this work, we present a simple model that describes the limits of protein concentrations for this effect and how mutations to this scaffold system should behave.

Lin et. al.[70] performed a series of experiments looking at how addition of mutants of KSR modified the pro-zone effect. The experiment involves over-expression of the scaffold KSR of the MAP Kinase pathway in T Cells. To a normal in vivo system containing the wild-type scaffold, extra scaffold (S') is added which is unable

to bind one of the kinases of the pathway, and the output of the scaffolded system (in terms of phosphorylated ERK, which is assumed to be a direct consequence of how much complex was formed) was measured. They found that addition of extra scaffold which did not bind MEK produced a qualitatively different sequestration curve, and in this section we attempt to understand this finding.

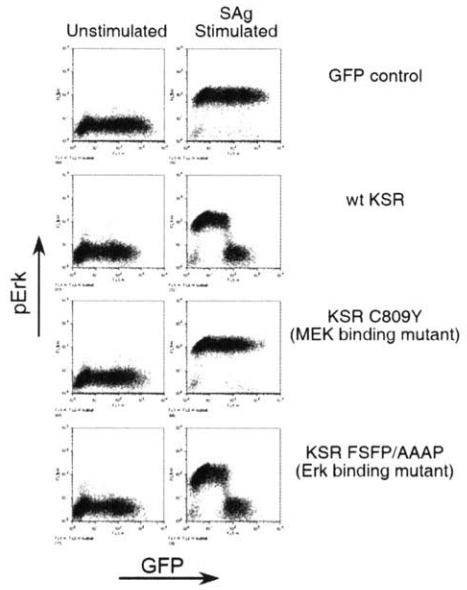


Figure 2-1: Variation of signal(pErk) with scaffold(KSR) concentration for different scaffold mutants[70]. The X-axes are amount of scaffold (tagged with GFP); Y-axes are the output of the scaffold system. KSR C809Y is a mutant that does not bind to MEK, and KSR FSFP/AAAP is a mutant that does not bind ERK. Note that the wild-type scaffold and Erk-binding mutant show the sequestration effect, but the MEK binding mutant does not.

2.2 Model

The basic limits on what sets the maximum signal of a system with a scaffold can be explained in a fairly simple equilibrium model. Consider a system with four components: a scaffold, S, and kinases A, B and C which bind to the scaffold and sequentially activate (in some arbitrary order). Without loss of generality, let the amounts of the kinases be $A < B < C$. Assume, to begin with, that all equilibrium constants for

binding of kinases to the scaffold proteins are infinity, there is no cooperativity in binding, and that a complete SABC complex (output signal) is needed for activation. The various phosphorylation steps are assumed to happen once all the components are brought together by the scaffold; this is not considered explicitly.

When S is varied:

1. For $S < A$, there is enough A, B and C to bind to all the S molecules, so the signal is equal to the amount of S

2. For $A < S < B$, all the S molecules are bound to B and C molecules, but there is insufficient A to bind to them all; only that fraction which binds to A will signal. Hence the signal is equal to the amount of A, which is the plateau in the graph.

3. For $B < S < C$, all the S molecules are bound to C; however, A and B distributes among the S molecules leading to a sequestration of A and B molecules away from each other. This reduces the signal, and it can be seen that the signal decreases (because the sequestration is more prominent) with increasing S (Figure 2.2).

4. For $S > C$, sequestration has already occurred. The amount of C (and in general, any component more numerous than the scarcest two in the system) has no bearing on the combinatorial inhibition effect.

In the case where some of the concentrations are equal, it is easy to see that the same logic still applies and the above results still hold.

We now consider the case where mutant scaffold S' is added. The case where this extra scaffold, S', is the same as the wild-type scaffold is trivial. The cases are possible for the mutant scaffold:

1. A-mutant (S' cannot bind to A): In this case, no sequestration occurs until

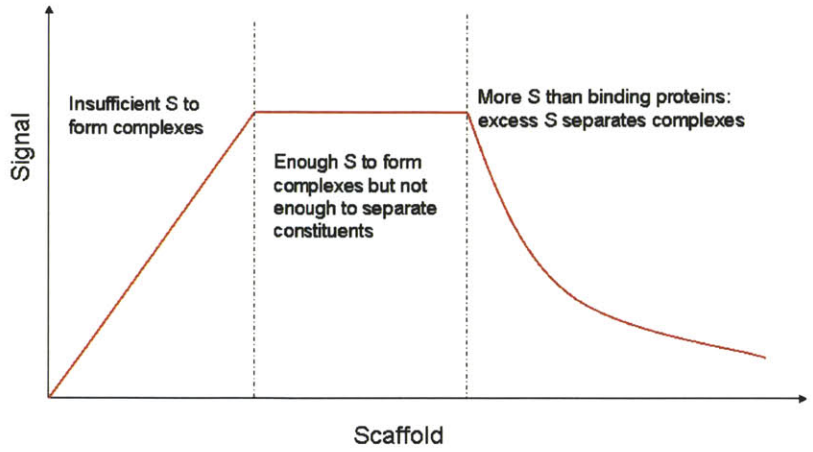


Figure 2-2: The variation of signal output from a scaffolded system as a function of concentration of scaffold. There are three regimes: Increasing output with scaffold, when the scaffold is limiting; saturation, when there is enough scaffold to form complexes but not enough to separate the constituents; and decreasing output when the scaffold sequesters the component molecules separately.

$S + S' > B$, which means the behavior of the system is the same as the wild-type case.

2. B-mutant (S' cannot bind to B): When the amount of total scaffold $S + S'$ is greater than A, sequestration occurs; therefore, the signal starts decreasing even when $A < S + S' < B$, which is earlier than in the wild-type case.

3. C-mutant (S' cannot bind to C): This is similar to the wild-type case, sequestration occurring after $S + S' > B$.

Note that it is not possible for the signal to decrease later than in the wild-type case. For the MEK binding mutant in the experiment, absence of the expected decrease in signal with an increase in the amount of scaffold cannot be explained by any stoichiometric argument – hence it seems that there is another previously unknown effect present in this system. For example, it is known that MEK is constitutively bound to KSR in certain systems[71]. Upon further investigation, it was found that the effect of KSR in this system was more complicated than initially thought, and is still not completely understood[70].

Let us look at what happens if we relax the assumption of the equilibrium constants for binding being infinity. For the sake of this analysis, drop constituent C from the system, i.e. assume C is in sufficient excess not to matter. Now our system consists of molecules A and B and scaffold S. Let the equilibrium constants for the binding of A and B to S be K_a and K_b respectively. The equations for the system are given in Equations 2.1 and 2.2.

$$\begin{aligned} S + SA + SB + SAB &= S_0 \\ A + SA + SAB &= A_0 \\ B + SB + SAB + B_0 & \end{aligned} \tag{2.1}$$

$$\begin{aligned} K_a &= \frac{[SA]}{[S][A]} \\ K_b &= \frac{[SB]}{[S][B]} = \frac{[SAB]}{[SA][B]} \end{aligned} \tag{2.2}$$

The other equilibrium relationship, $K_a = \frac{[SAB]}{[SB][A]}$ is not independent of the specified three. This represents a system of 6 unknowns in 6 equations (3 of which are non-linear), so the system is completely determined.

The system is normalized by setting total B (call this B_0) to be = 1. Total A (call this A_0) = some fraction less than 1 (taken as 0.1 for the following). The x-axis of plots in Figure 2.2 is the amount of scaffold S plotted on a log scale; the y-axis is the amount of signal SAB normalized the maximum possible amount of signal (which is A_0). Note that the natural dimensionless parameters in this system are (Concentration)*(Equilibrium constant for binding), since the dimensions of K_a and K_b are 1/Concentration. Figure 2.2A looks at the effect of the parameter $[B_0]K_b$ on the shape of the curve. For small values of $[B_0]K_b$ the amount of scaffold needed to reach a peak signal is greater than B_0. The above analysis is therefore correct as long as $[B_0]K_b$ is sufficiently greater than 1. $[A_0]K_a = [B_0]K_b$ in all cases in the above graph – so the graph appears symmetric. Figure 2.2B shows that changing each parameter but keeping the product constant maintains the shape of the graph.

Again, $[A_0]K_a = [B_0]K_b$ in all cases. The position of the graph is shifted because B_0 is changed; all cases the decrease of signal starts happening once $S > B_0$.

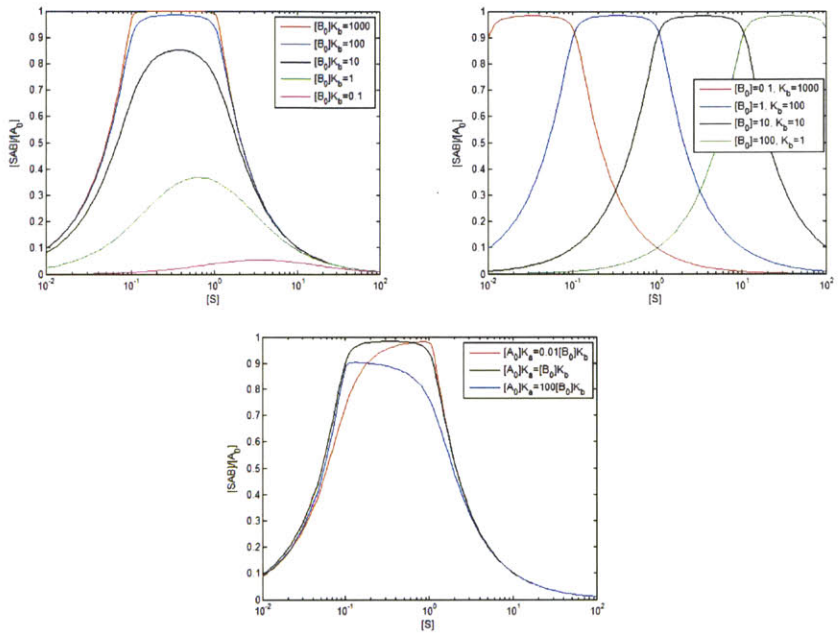


Figure 2-3: Plots of the normalized output of a scaffold process with variations in parameters. (A) Variation with the strength of scaffold binding to the second-smallest (limiting) component (B) Variation with concentration of the second-smallest component, at constant equilibrium (C) Skewing the saturation curve for different equilibrium constant for the smallest two components (by amount)

2.3 Discussion

In this section, using simple logical arguments, we have described the limits of protein concentrations at which the scaffold pro-zone effect occurs. The work also describes how simple experiments involving over-expression of the scaffold (and various mutants) in the system can reveal previously unknown cooperative or nonlinear behavior.

Chapter 3

Regulation of Src Kinases by CD45 in T and B Cells

The Src family kinases are a set of kinases with various, possibly redundant functions in the activation of lymphocytes, and each kinase is known to have multiple targets[72]. The set of Src family kinases includes Lck, Fyn, Lyn, etc. and different kinases are thought to be expressed in different cell types. Lck is the prominent kinase expressed in T cells and Lyn in B-cells[73]. These kinases have a similar structure and are thought to be regulated in a similar manner. Lck is known to phosphorylate the tyrosines of immunoreceptor tyrosine-based activation motifs (ITAMS) in T-cells, which is a crucial step in their activation. In this chapter, we primarily examine a specific aspect of how Lck is regulated, and briefly connect our findings to the regulation of Lyn.

Upon antigen recognition (sufficiently strong binding) of peptide-major histocompatibility complex (pMHC) by the T cell receptor (TCR), one of the earliest signaling events is the phosphorylation of the ITAMs of the CD3-zeta subunit of the TCR by Lck. This leads to the binding of Zeta-associated protein of 70 kDa (ZAP-70) to the TCR, initiating downstream signaling. In conditions like autoimmunity, T-cells signal through the TCR even when not activated by a strongly binding antigenic pMHC. Understanding how "upstream" molecules regulate of the activity of Lck, a key ki-

nase of the TCR, could help select targets for inhibiting spurious activation of T cells.

Src kinases have two tyrosine sites which can be phosphorylated to modulate their activity. There is an activating site (Y394) and an inhibitory site (Y505). In the most active form of Lck only Y394 is phosphorylated, and in the least active form only Y505 is phosphorylated. The form of Lck in which both sites are dephosphorylated is referred to as the "basal" state, with intermediate kinase activity. The kinase for the Y394 site is Lck itself, i.e. there is autophosphorylation[74]. It is thought that the phosphorylation of the inhibitory site changes the conformation of Lck in such a way as to make the activatory site inaccessible to phosphorylation (the "tail-bite" mechanism[75] in which the phosphorylation of Y505 causes the tail of Lck to contract and close in on itself, rendering Y394 inaccessible). It is not entirely clear, however, if the fourth state in which both sites are phosphorylated can exist; for example, Sun et al. report that phosphorylation of Y394 blocks phosphorylation of Y505[76], but Nika et al report the presence of a form of Lck phosphorylated on both activating and inhibitory sites[77]. Table 3.1 shows the various possible states of Lck.

The kinase for the Y505 site, Csk, is known to be modulated by the activity of Lck itself. It is thought that Csk is recruited to the membrane through its interaction with an adaptor protein, Cbp/PAG[78][79], and that PAG needs to be activated in order for it to recruit Csk. This activation (by phosphorylation) is thought to be performed by Lck itself, creating a negative feedback loop which tempers the activity of Lck.

The dephosphorylation of the two tyrosines of Lck, Y394 and Y505, is performed by many, possibly redundant, phosphatases. CD45 is known to be a major phosphatase involved in regulation of both these sites[80], and is critical to TCR signaling responses[81]. Since phosphorylation of the Y394 site tends to activate and phosphorylation of the Y505 site tends to inhibit Lck, CD45 has a potentially interesting role as both an activator and inhibitor of Lck. There are other molecules involved in the

system with possibly redundant roles. For example, Lyp/PEP is a protein tyrosine phosphatase that is known to be involved in the deactivation of Lck by the dephosphorylation of the activating tyrosine[82]. SHP-1 is also thought to play a similar role, for example, through a feedback loop involving ERK[14][83].

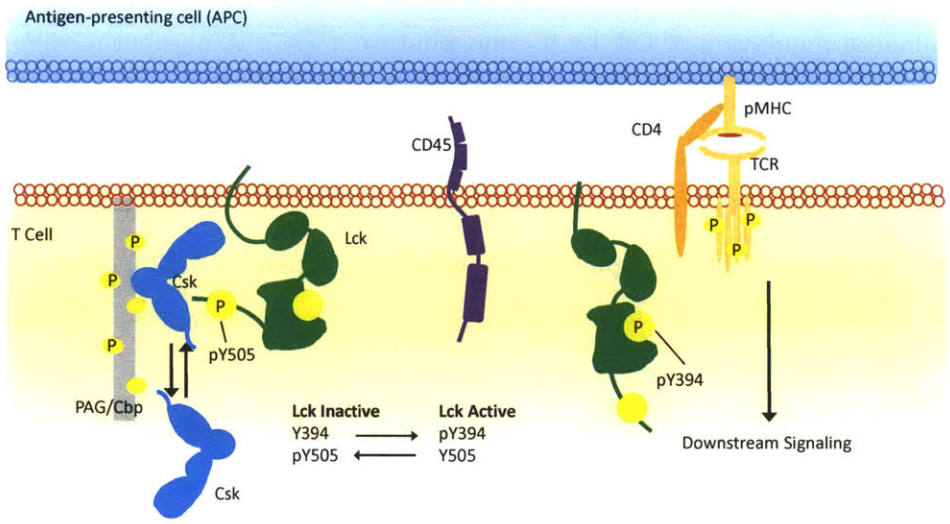


Figure 3-1: Cartoon of the molecules involved in Csk-dependent activation of T-Cells. Lck (dark green) phosphorylates TCR (yellow); Csk (blue) is a kinase for the Y505 site on Lck, Y394 autophosphorylates. Csk is recruited to the cell membrane on binding with phoshorylated PAG. CD45 (purple) is a phosphatase for both tyrosines on Lck. Phosphorylated ITAMs lead to downstream events via Zap70 and LAT.

Previous work on modeling the Lck activation mechanism has focused on two aspects. One is the types of qualitative behaviors that the system can show[84]. The presence of competing positive (activation by trans-autophosphorylation of Lck) and negative (modulation of Csk activity by phosphorylation of PAG) feedback loops in the activation scheme of Lck could lead to bistabilities, oscillations, and pulses in Lck activity, depending on the parameter regime in which the system operates. Another model of the Src kinase activation scheme focuses on the interaction with receptor[85]. Neither work looks at the interesting effect on the system of CD45, which is thought to be both an activator and repressor of Lck; this dual conflicting role of CD45 on Lck activity could have interesting biological consequences.

It is known that CD45-deficient thymocytes show diminished LAT, Akt and Zap70 and almost no ERK activation[86]. These experiments by Hermiston et al[82] provide an opportunity to study cell lines with various intermediate levels of CD45, which could help unravel how the different downstream signaling molecules are differentially regulated. Recent experiments along these lines[87] have shed new light on the mechanism of regulation of Lck by its phosphatase, CD45. A number of different cell lines with varying amounts of CD45 expressed on the surface were generated using an allelic series; the amount of CD45 varied from 5% of wild-type to 150% of wild type. It was found that the level of pY394 is maximum for the genotype with intermediate levels of CD45 (Figure 3-2A), which corresponds to 50% of the wild type level of CD45, whereas the level of pY505 monotonically decreases with increasing CD45. The location of this peak in activity of CD45 as a function of the levels of the rest of the molecules involved in the system is also of interest, as it could be shifted by changing, for example, the amount of Csk in the system. Since CD45 is a phosphatase to both the activating and inhibitory sites on Lck, it is not easy to intuit the mechanism underlying these results for the cellular response as a function of the level of CD45.

Changing the amount of CD45 in B-cells, does not produce a maximum in the level of pY394 (Figure 3-2B). In B-cells, the dominant kinase is Lyn. CD45 dephosphorylates only the activatory site on Lyn[88]. Hence, changing the level of CD45 in B-cells should produce qualitatively different responses from those seen in T-cells. Other phosphatases like PEP in T-Cells, which perform some redundant functions, can be modeled in a similar way: since PEP only has an impact on the state of pY394 in Lck in T-cells, we would expect the qualitative behavior of a system in which PEP was varied in T-cells to be similar to that of a B-cell system in which CD45 is varied.

Here we show that a simple model for the mechanism of regulation of the Src kinase can explain the effects of different levels of CD45 in the system, and this model

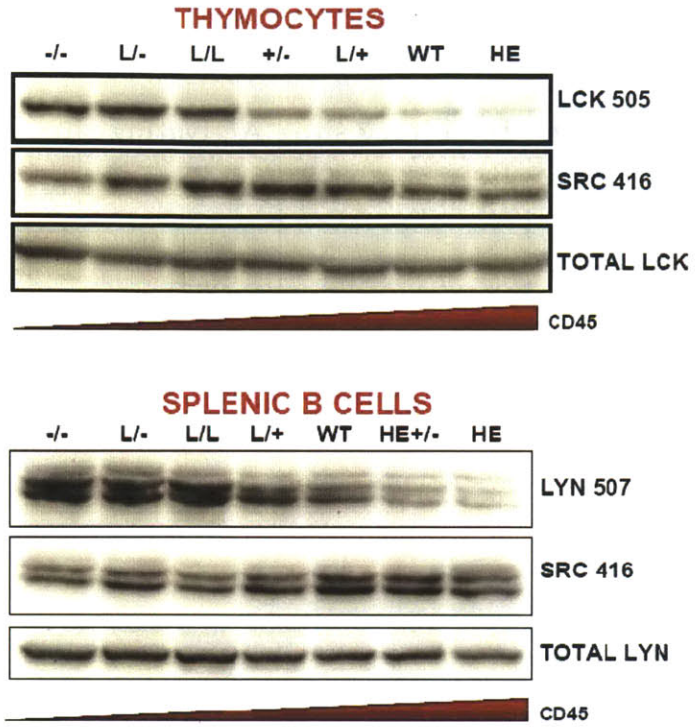


Figure 3-2: Western blots[87][88] showing the levels of the two phosphorylation sites in Src kinases as a function of the amount of CD45 expressed, (a) Lck in T Cells and (b) Lyn in B Cells. Src416 binds activating tyrosines of both Lck and Lyn[88]

is consistent across various cell types which express different types of Src kinase.

3.1 Model

The model we study consists of the following set of molecules: Lck, its kinase for the inhibitory site, Csk, and its phosphatase, CD45. Let the total amount of Lck be L, and the three possible states in which it can exist be A (activated state: pY394 phosphorylated, pY505 unphosphorylated), B (basal state, neither state phosphorylated) and I (inactive state, pY394 unphosphorylated, pY505 phosphorylated). One could add if necessary a fourth state C in which both sites are phosphorylated. The other molecules involved in modulating the activity of Lck are CD45 (D) and Csk (S). CD45 dephosphorylates both sites of Lck; Csk phosphorylates the inhibitory site

and the activating site is phosphorylated by Lck itself (Figure 3-3A).

State	Abbreviation	Inhibitory site	Activating site	Kinase Activity
Inactive	I	P		Low
Basal	B			Moderate
	C	P	P	Moderate(?)
Active	A		P	High

Table 3.1: The Various Activation States of Src Kinases

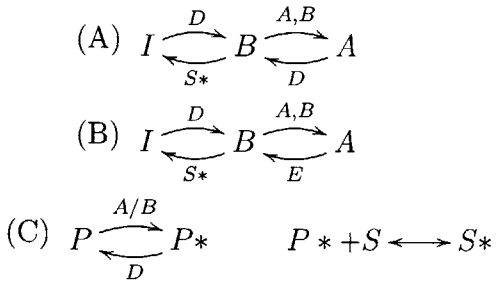


Figure 3-3: Schematic diagrams of the reaction networks simulated in the model: (A) The minimal (toy) network in T Cells, comprised of the regulation of the various states of Lck by Csk and CD45. (B) The minimal network in B Cells, comprised of the regulation of the various states of Lyn by Csk, CD45 and PEP (C) Additional reactions present in the full model: phosphorylation/dephosphorylation of PAG by Lck and CD45, and requirement of Csk to bind to PAG to make it active. The full model also contains another state C of the Src kinase with both sites phosphorylated, and all reactions are Michaelis-Menten. The labels are A:active Src, B:Basal Src, I:Inactive Src, D:CD45, S:Csk, P:PAG, E:PEP. Star denotes the "active" state for Csk and phosphorylated state of PAG.

Csk is brought to the surface of T-cells by phosphorylated Pag/Cbp, which is thought to be phosphorylated by Lck and dephosphorylated by CD45; the toy models do not contain this regulation of Csk by Lck and PAG, but both the toy and full models show the same qualitative behavior. For the case of Lyn, PEP is the phosphatase that dephosphorylates the activating (Y416) site, whereas the inhibiting site (Y507) is dephosphorylated by CD45. The rest of the network is similar to that of Lck (Figure 3-3B).

The minimal model consists of the three forms A, B, and I of Lck, CD45 and

Csk; all reactions in the minimal model are assumed to be of mass action form. In the minimal model, Csk converts LckI to LckB, LckB autophosphorylates to result in LckA, and both LckI and LckA are dephosphorylated by CD45 to give LckB. Different forms of the Src kinase phosphorylate their substrates at different rates, with the inhibiting form being unable to phosphorylate its substrate, the basal form having a low kinase activity and the activated form having a high kinase activity. We shall see that this minimal model is sufficient to recapitulate the basic features of the Lck-CD45-Csk system seen in experiments. The reactions in the minimal model are chosen to be of mass-action form. The minimal model provides us the ability to solve for the activities of the various forms of Lck analytically and provides insight into the qualitative variation of these solutions with various parameters of the system, such as the total amounts of Csk which one could potentially vary in experiments.

The full model, apart from using the more complicated and possibly more realistic Michaelis-Menten form for reaction kinetics, also contains PAG, which is activated by phosphorylation by Lck or Lyn, dephosphorylated by CD45 and acts as an adaptor to bring Csk to the surface where it can interact with Lck (Figure 3-3C). In this case as well, different forms of Src kinase phosphorylate PAG at different rates, the rate constants assumed to be the same as their kinase activities to Lck. The complete list of reactions and rate constants for the full model are noted in the supplementary material (Section A). We obtain results for the full model numerically, as it is too complex to analyze analytically. The qualitative behavior seen in the minimal model is maintained when one goes to a much bigger, and more realistic full network, for a certain choice of parameters; this qualitative behavior is also fairly robust to variations in the parameters (as seen in the parameter sensitivity analysis included in the supplementary material).

The chemical reactions for the Lck network (Figure 3-3A), written in their simplest

mass-action form, are:

$$\begin{aligned} I + D &\xrightarrow{k_1} B + D \\ A + D &\xrightarrow{k_1} B + D \\ B + S &\xrightarrow{k_2} I + S \\ B + B &\xrightarrow{k_3} A + B \\ A + B &\xrightarrow{k_4} A + A \end{aligned} \tag{3.1}$$

The parameters for this system are k_1, the rate of CD45's phosphatase activity; k_2, the rate of Csk's kinase activity; k_3, the rate of basal activity and k_4, the rate of active Lck kinase activity (the latter two during autophosphorylation). The fact that active Lck has a higher kinase rate than basal Lck is described by the constraint $k_4 > k_3$.

From these reactions we can write a set of ODEs describing the evolution of the dynamics of this system:

$$\begin{aligned} I &= L - A - B \\ \frac{dB}{dt} &= k_1 (L - A - B) D + k_1 AD - k_2 SB - k_3 B^2 - k_4 AB \\ \frac{dA}{dt} &= -k_1 AD + k_3 B^2 + k_4 AB \end{aligned} \tag{3.2}$$

For the purpose of examining the qualitative behavior of this system, we choose the following parameters: $k_1 = 1, k_2 = 1, k_3 = 1, k_4 = 2$. We then look at the steady states of this ODE system and the dependance of the steady state solutions on the amount of CD45.

For our simple model of the action of Lyn (Figure 3-3B), we can write down the

following set of equations:

$$\begin{aligned} I + D &\xrightarrow{k_1} B + D \\ A + P &\xrightarrow{k_5} B + P \\ B + S &\xrightarrow{k_2} I + S \\ B + B &\xrightarrow{k_3} A + B \\ A + B &\xrightarrow{k_4} A + A \end{aligned} \tag{3.3}$$

$$\begin{aligned} I &= L - A - B \\ \frac{dB}{dt} &= k_1 (L - A - B) D + k_5 AP - k_2 SB - k_3 B^2 - k_4 AB \\ \frac{dA}{dt} &= -k_5 AP + k_3 B^2 + k_4 AB \end{aligned} \tag{3.4}$$

3.2 Results

3.2.1 Qualitative trends for the regulation of Lck activity derived from experiments

Note that in the minimal model L = total amount of Lck = A + B + I. The qualitative feature of the experiments performed is (a) a maximum in pY394 activity as CD45 (denoted, D) is varied (b) monotonic decrease in pY505 as a function of increasing D (Figure 3-4A). From these experimental findings, if we calculate the qualitative trends in the variation of Lck states A and B with changes in CD45 expression, we find that LckA goes through a maximum with increasing D and LckB monotonically decreases with increasing D (Figure 3-4C). For the Lyn experiments, qualitatively, the level of the activating site pY416 increases monotonically and the level of the inhibitory site pY507 decreases (Figure 3-4B). In terms of the states of Lyn, this corresponds to an increase in LynA and a decrease in LynI with CD45. However, since it is not possible to obtain the qualitative trends for the variation of LynB which is total Lyn (constant), minus the sum of LynA(monotonically increasing) and LynI (monotonically decreasing), we cannot make any definitive statement about LynB (Figure 3-4D).

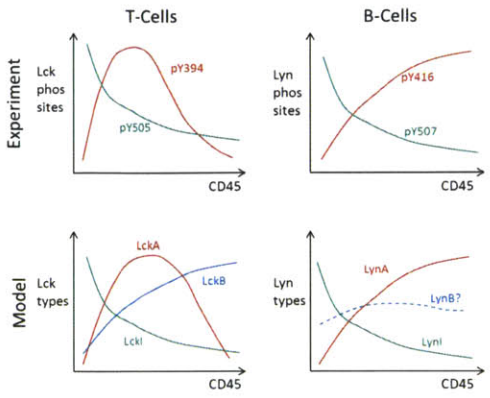


Figure 3-4: Qualitative trends of the levels of the various phosphorylation sites seen from experimental data (top row) and what that would mean for the various states (bottom row). Experimental data shows (a) a maximum in pY394 of Lck as a function of CD45 and (b) a monotonic increase in pY416 of Lyn as a function of CD45; the level of the inhibiting site pY505 of Lck and pY507 of Lyn decreases monotonically. This converts to, in terms of the states of the Src kinase, (c) a maximum in LckA and a monotonic increase in LckB and (d) a monotonic increase in LynA as a function of CD45. No qualitative prediction can be made for the level of LynB; the inhibitory states LckI and LynI decrease monotonically with CD45.

3.2.2 Solutions of Lck regulation models

We obtain the steady-state behavior by setting the left hand sides of Eqs.3.2 to zero, and then solving the resulting algebraic equations simultaneously. Solutions for the steady state levels of LckA and LckB are presented for the rate constants specified above; the complete solution for the location of the maximum in A is shown. Full solutions for steady states of LckA and LckB are presented in the supplementary material. Using the minimal model for Lck, we find that there is a maximum in LckA as CD45 is varied. We can then obtain the the level of CD45 that corresponds to the maximum in LckA. One can think of the level of Csk in the system as a proxy for the activation of the T-Cell. A low level of Csk in the cell would result in lower phosphorylation of Lck Y505, and therefore activate Lck by trans-autophosphorylation. We can vary the amount of Csk, and explore what happens upon activation (say, by receptor stimulation or as recently done by using analog sensitive Csk constructs(19)).

The steady states for LckA and LckB as a function of the amount of CD45 are given in Table 2. The level of CD45 (D_{max}) corresponding to the maximum in LckA is given by the following expression:

$$D_{max} = \frac{k_2 k_3 S - k_4 \sqrt{k_2^2 S^2 + k_2 S L (k_3 - k_4)}}{k_1 (k_3 - k_4)} \tag{3.5}$$

The biologically realistic case, when active Lck has a higher kinase rate than basal Lck, is described by $k_4 > k_3$. Substituting the rate constants chosen above,

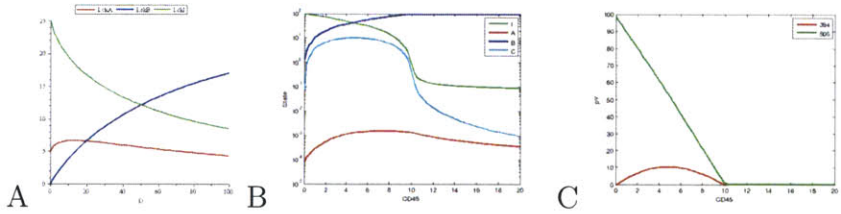
$$D_{max}(S, L) = -S + 2\sqrt{S^2 - SL} \tag{3.6}$$

If there is no positive feedback loop ($k_4 = 0$), the location of the maximum is a function of Csk only, and independent of the amount of Lck present in the system. The result of the toy model for Lck, which is a plot of the levels of the various forms of Lck as a function of CD45 for S = 50, L = 30 is plotted in Figure 3-5A. The amounts of the various states of Lck, obtained by solving the full model for Lck, are plotted in Figure 3-5. The red curves represent the activated state, green curve the inactive state, and blue the basal state of Lck. There is qualitative similarity between the two models: a maximum in LckA, monotonic increase in LckB and monotonic decrease in LckI as a function of CD45; this is similar to what one would infer from the experiment (Figure 3-4c).

In the minimal model, the levels of the two phosphorylation sites pY394 and pY505 correspond to the two states LckA and LckI respectively, since in the state B neither site is phosphorylated (and state C does not exist in this model). The levels of the various phosphorylation sites, as calculated from the full model, is plotted in Figure 3-5C, and shows qualitative experiment with the toy model and experiment (Figure 3-2A, Figure 3-4A). The presence of the maximum in pY394 as a function of CD45 is explained by the fact that CD45 has both a positive and a negative regulatory role for the system as CD45 dephosphorylates both the pY394 and pY505 sites on Lck.

Model	Src Type	Analytical Solution
Lck	A	$A(D,L,S) = L - \frac{(D+S)\left(2L+S+D-\sqrt{4L^2-4LS+D^2+2DS+S^2}\right)}{2D+4S}$
Lck	B	$B(D,L,S) = \frac{2LD+DS+D^2-D\sqrt{4L^2-4LS+D^2+2DS+S^2}}{2D+4S}$
Lyn	A	$A(D,L,S,P) = \frac{-D^2P-PS^2+2DLS-2DPS}{2D(D+2S)} - \frac{(D+S)\sqrt{4L^2D^2-4LDPS+D^2P^2+2DP^2S+P^2S^2}}{2D(D+2S)}$
Lyn	B	$B(D,L,S,P) = \frac{2DL+PS+PD}{2(D+2S)} \frac{-\sqrt{4L^2D^2-4LDPS+D^2P^2+2DP^2S+P^2S^2}}{2(D+2S)}$

Table 3.2: Analytical Solutions to the toy model for the specific choice of parameters $k_1 = 1, k_2 = 1, k_3 = 1, k_4 = 2$.



A B C

Figure 3-5: Results of (A) the toy model and (B) and (C) the full model for Lck. Red curves represent the activated state, green curve the inactive state, and blue the basal state of Lck. (A) and (B) are plots of the levels of various states as a function of CD45 for the toy and full model respectively; (C) is a plot of the levels of the phosphorylation sites from the full model. The levels of the sites for the toy model is the same as Figure (A), because in the toy model there is no site with both sites phosphorylated and so in this case red curve = pY394 and green curve = pY505

At low CD45, it seems that the system is in a state with predominantly inactive Lck, and the low amount of CD45 is not enough to overcome the inactivation caused by Csk phosphorylating the inactivating site. As a result of this, not enough basal Lck is formed and the positive feedback loop does not seem to be "ignited". With increasing CD45, more and more basal Lck is formed because the dephosphorylation of pY505 by CD45 overcomes its phosphorylation by Csk. This results in kicking off the positive feedback loop in Lck, and we see a maximum in Lck activity. With a very high amount of CD45, the dephosphorylation of pY394 is also significant and seems to keep the positive feedback loop in check, decreasing the level of Lck activity again. Hence,

the maximum in Lck activity represents a tradeoff between having enough CD45 to overcome the inhibitory role of Csk and not too much that dephosphorylates pY394. The presence of the maximum in the activity of LckA as a function of CD45 level begs the question, how much CD45 is needed to be at that maximum, and how do the amounts of other molecules in the system affect the location of this maximum? Analytically, the location of the maximum from the toy model can be calculated (see previous section). Fig 5a is a plot of this expression, position of the maximum as a function of amount of Csk (measure of the level of stimulation of the cell). We can perform the same calculation for the full model; the results are in fig 5b. The amount of CD45 required in the system for Lck to be maximally active increases with both Csk and total Lck. The Csk-CD45 balance is relatively simple to understand: by transfecting more kinase for Y505 (Csk), one would need more phosphatase (CD45) to achieve a similar balance. The reason for the maximum increasing with increasing total Lck is probably because of the positive feedback loop in Lck: increasing the total amount of Lck in the system makes it more likely to set off the positive feedback loop and phosphorylate Y394, and hence more CD45 would be needed to counter this. Unlike LckA, the solution for LckB does not show a maximum as a function of CD45 for any values of parameters.

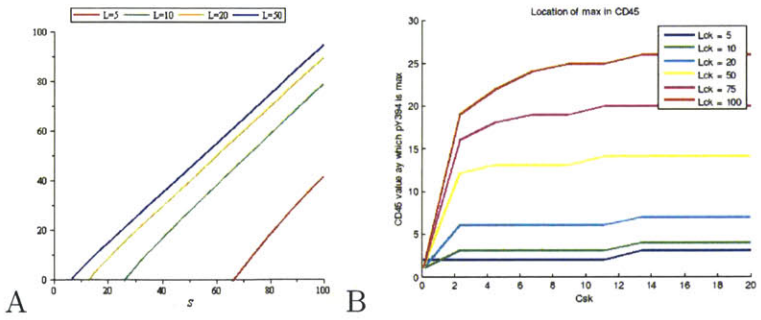


Figure 3-6: Location of the maximum in CD45 in T-Cells as a function of the amount of Csk present in (a) the toy model and (b) the full model. Different curves represent different amounts of total Lck

3.2.3 Solutions of the Lyn/PEP model

The steady states for LynA and LynB are given in Table 2. In this case, there is no maximum in activity of Lyn as a function of CD45, and the amount of active Lyn increases monotonically with CD45. The results of the toy model for Lyn, a plot of level of the levels of the various forms of Lyn as a function of CD45 for S = 50, L = 30, P=10, are plotted in Figure 3-7 and the corresponding results for the full model are plotted in e6b. We see that the qualitative features are the same as what one would predict from experiment: a monotonic increase in LynA and decrease in LynI (Figure 3-4d). There is a maximum in LynB in the full model, but not in the toy model, and we make no claim about the actual behavior for this state. In the case of Lyn, since there is no tradeoff between activating and inhibiting effects of CD45, the effects are straightforward: as CD45 inhibits only the inhibiting site Y507, the amount of active Lyn increases and inactive Lyn decreases monotonically with increasing CD45.
Fig 6c is a plot of the levels of the phosphorylation sites from the full model. The levels of the sites for the toy model is the same as fig 6a, because in the toy model there is no site with both sites phosphorylated and so in this case red curve = pY416 and green curve = pY507. This compares favorably to the experiment (Figure 3-2B, Figure 3-4B).

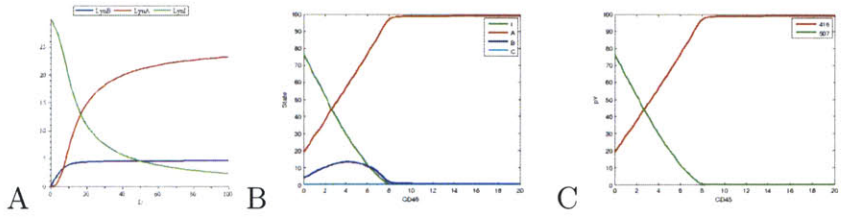


Figure 3-7: Results of (a) the toy model and (b) and (c) the full model for Lyn. Red curves represent the activated state, green curve the inactive state, and blue the basal state of Lyn. (a) and (b) are plots of the levels of various states as a function of CD45 for the toy and full model respectively; (c) is a plot of the levels of the phosphorylation sites from the full model. The levels of the sites for the toy model is the same as figure (a), because in the toy model there is no site with both sites phosphorylated and so in this case red curve = pY416 and green curve = pY507.

3.2.4 Parameter Sensitivity Analysis

A sensitivity analysis was performed for the full model. Each parameter in the original model was doubled or halved and the system was simulated. The qualitative behavior of the system (maximum in pY394 and monotonic increase in pY505 as a function of CD45 for Lck) remained for all variations in parameters. A similar procedure was carried out for the Lyn model, and the qualitative response in that case remains the same. The graphs in the supplementary material show how the activities of the two phosphorylation states and the various forms of Lck and Lyn change as parameters of the system are varied. The qualitative features are the same when parameters are varied within reasonable limits: the level of pY394 (red curve) always goes through a maximum and pY505 always decreases monotonically, even when the rate constants are doubled (dashed curves) or halved (dotted curves). For the corresponding plots for Lyn, in section SC4, the level of pY416 (red) monotonically decreases and pY507 (blue curve) increases monotonically irrespective of changes in parameters. Variations in the states of Lck and Lyn when parameters change are also plotted. The system seems to be most sensitive to the rate constants kBcat and kPDcat representing the catalytic activity of basal Lck and the rate at which CD45 dephosphorylates PAG.

3.3 Discussion

Motivated by experiments where the expression of CD45 was varied in lymphocytes using an allelic series of mice, we have studied a simple biologically-inspired model of regulation of Src kinases. The model captures and provides mechanistic underpinnings to the available experimental data. For example, it recapitulates the observation that Lck activity exhibits a maximum for a specific CD45 level, while Lyn activity varies monotonically with CD45 expression. Our model further suggests that the amount of CD45 needed for maximal Lck activity increases with both increasing Csk and total Lck in the system. This may be of considerable biological significance. Lymphocyte activation corresponds to a lower level of Csk. Thus, according to our

model, the amount of CD45 needed for maximal Lck activation is lower upon activation. This could be why CD45 molecules are excluded from the TCR cluster during activation(20). Hence, the topology of the Lck regulatory network makes it possible to "tune" for maximal responsiveness in the activated case, while still keeping a lower level of activity while Csk is high. By transfecting more Lck into the cell lines and assessing whether the amount of CD45 at which pY394 is maximum changes or is constant, one could better understand the magnitude of this positive feedback loop.

The qualitative difference in behavior between the Lck and Lyn systems is also easily explained by the model. It is a direct result of the fact that CD45 cannot dephosphorylate the Y507 site of Lyn. As a result, we do not expect that Lyn activity can be tuned for maximal signaling by CD45 exclusion. The differential ability of T cells and B cells to tune kinase activity suggested by our model could be important. For T-cell activation, upon TCR-pMHC interaction, it is known that CD45 is excluded from the TCR protein island[89]. Depending on the amount of CD45 originally present, this could result in either activation or deactivation of Lck as per our model. Removal of CD45 from the BCR rich region in B cells, however, would only result in a decrease in the activity of Lyn.

This model would also suggest that the only experimentally observable qualitatively interesting feature of the system is the maximum in pY394 of Lck as a function of CD45. The exact location of this maximum in the pY394 site of Lck is dependent on the parameters of the system and the amounts of various molecules. The other phosphorylation site pY505 of Lck and the sites of Lyn will show only monotonic behavior in these CD45 allelic series lines, irrespective of the amount of Csk or total Lck in the system.

Chapter 4

Small-Molecule (Csk) Activation of T-Cells

In the previous chapter we described how the levels of the various phosphorylation sites of Lck and Lyn can be described by a simple model. In this chapter, we use the results from the previous one in a wider context, that of the activation of T-Cells by small molecules. Inhibitors are a key class of drug used to regulate aberrant signaling, but most inhibitors interact with multiple members of the protein kinase family[90]. Therefore, it is important to understand the function of different kinases involved in cellular signaling. Doing so not only helps us understand the basic biology of the systems under consideration, but helps understand what side effects may occur as a result of using non-specific inhibitors. Knocking-in analog-sensitive kinase (as-kinase) by substituting for or in addition to the endogenous wild type kinase in cells or organisms followed by treatment with a specific inhibitor allows for the investigation of the role of a single kinase[91]. In order to examine the effect of Lck signaling, Schoenborn et al. developed an analog-specific version of Csk, a key regulator of Lck[92] together with 3IB-PP1, a pyrazolopyrimidine-1 (PP1) derivative to inhibit it[93].

Csk is a key kinase that controls lymphocyte development and prevents spurious activation of immune cells; $Csk^{-/-}$ mice are embryonically lethal[94][95]. The mechanism by which Csk influences TCR phosphorylation is through the Src kinase Lck

as described before. Various forms of Lck can phosphorylate the ITAMs on the CD3ζ region of the TCR[96]. Phosphorylation of CD3ζ leads to recruitment of Zap70, which in turn activates LAT and sets off events downstream. CD45 is known to be a major phosphatase involved in regulation Lck activity[80], and is critical to TCR signaling responses[81]. There are other molecules involved in the system with possibly redundant roles. For example, Lyp/PEP is a protein tyrosine phosphatase that is known to be involved in the deactivation of Lck by the dephosphorylation of the activating tyrosine[82]. SHP-1 is also thought to play a similar role, for example, through a feedback loop involving ERK[14][83]. In this work only CD45 is considered, and is the sole phosphatase that dephosphorylates both specified sites on Lck.

The amount of TCR in T-Cells is not constant; it is regulated by a complex mechanism. It is known that the degradation of TCR subunits is antigen- and dose- dependent[97][98]. TCR-CD3ζ complexes are constitutively internalized and recycled[99][100], and triggering of the T-Cell leads to the failure of the TCR to recycle back to the surface due to their targeting by lysosomes[101]. This degradation is thought to happen by Cbl-mediated ubiquitination[102][103]. It was also found that the rate of degradation in unstimulated T-Cells is much lower than in triggered T-Cells[104][103].

4.1 Experiments

In order to examine the potential for using analog sensitive inhibitors of Lck, experiments were performed by Jamie Schoenborn and others in the Art Weiss lab with Jurkat T-Cells[92]. Jurkats were transiently co-transfected with CskAS (or pEF6A, a control) vector and pEF-GFP. 18 hours post transfection, cells were rested in serum-free media for 20 minutes and were used for assays. Cells were first incubated with DMSO or 3IB-PP1 inhibitor for 15 minutes, then either assayed directly, either unstimulated or stimulated with C305 for 2-15 minutes. A number of variants of CskAS

were used, with different N-terminal tags that localize to different parts of the cell and thus have different effective activities towards Lck. The strongest variant is lipid-raft CskAS (Lck_{11}), the weakest the cytoplasmic versions of CskAS ($Koz2^-$ and $dPdM^-$). The non-raft membrane localizing version (dP^-) is of intermediate activity.

Figure 4-1[92] shows FACS data on the activation of T-Cells by CskAS and PP1. The left half of the figure represents the control, where PP1 is not added. The first row represents unstimulated cells, which is the main case of interest; other rows represent activation by C305 (α-CD3) for 2, 5 and 15 minutes respectively. The following observations can be made:

- Adding PP1 to unstimulated cells which had CskAS activates T-Cells
- Only the cells which expressed high amounts of CskAS could be activated in this manner
- For C305-stimulated cells (but no PP1) to which CskAS was added, cells with high levels of Csk did not activate. This is mainly true for the lipid raft-localized CskAS (the strongest version).

To test the strength of activation by CskAS/3IB-PP1 vs. activation by C305(α-CD3), Schoenborn et al[92] looked at western blots of the phospho-tyrosines (Figure 4-2). The levels of pCD3ζ for Csk/stimulation by PP1 is much higher than the levels from DMSO(control)+C305 stimulation, suggesting that activation of T-Cells using CskAS/PP1 is much stronger than the usual method of using α-CD3.

Figure 4-3 is a plot of the strengths of activation by each variant of Csk. The left panel shows histograms of phospho-CD3 for unstimulated and C305-stimulated cells (for 2, 5 and 15 min) for different variants of CskAS; the right panel is a quantification of the mean pCD3 in each case. We see that the strongest version of CskAS (Lck11) activates the strongest (has the highest value of pCD3 upon activation) when PP1 is added; the non-raft LckdP next strongest and so on - the mean level of pCD3 is directly proportional to the strength of CskAS used. We also note that the histograms

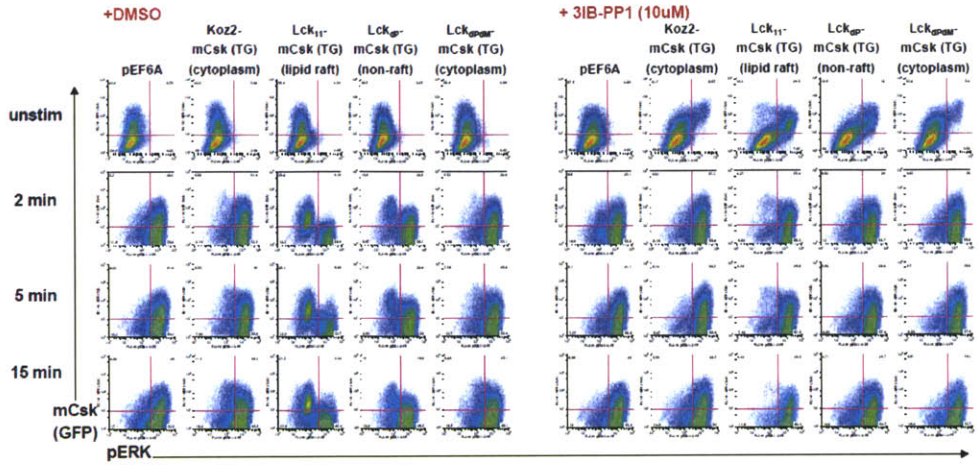


Figure 4-1: FACS data on the activation of T-Cells by CskAS and PP1[92]. The left half of the figure represents the control, where PP1 is not added. The first row represents unstimulated cells, which is the main case of interest; other rows represent activation by C305 (α-CD3) for 2, 5 and 15 minutes respectively. Each panel is a count of cells with the X and Y axes representing numbers of pERK and transfected Csk.

and mean values of pCD3 are the same for all unstimulated cells, irrespective of the strength of CskAS used. This is unexpected because presumably, different variants of CskAS should cause differences in Lck activity, which should translate to different pCD3 levels in unstimulated cells. The fact that this does not happen suggests that phospho-CD3 is regulated by a mechanism which ensures that its level at equilibrium is independent of the activity of Lck (which, as we shall see, is a consequence of the mechanism of regulation of CD3 levels).

The regulation of CD3 levels is further explored in Figure 4-4. The left panel shows how total CD3 (not just phospho-) levels vary as a function of the amount of CskAS. We are interested in the cases where PP1 has not been added (because that removes CskAS from the cell membrane). We see that the level of CD3 increases with the amount of CskAS present, and this increase (the slope of the CskAS-CD3 distribution) is greater when Raft CskAS is used (than cytoplasmic CskAS). This suggests that, in general, the higher the activity of CskAS, the more total CD3 is present on the surface of the T-Cell. The right panel of Figure 4-4 shows the levels of surface CD3

for different cell types which have different inherent expressions of CD3. For all these cell types, we see that the levels of total CD3 increase with the activity of CskAS used.

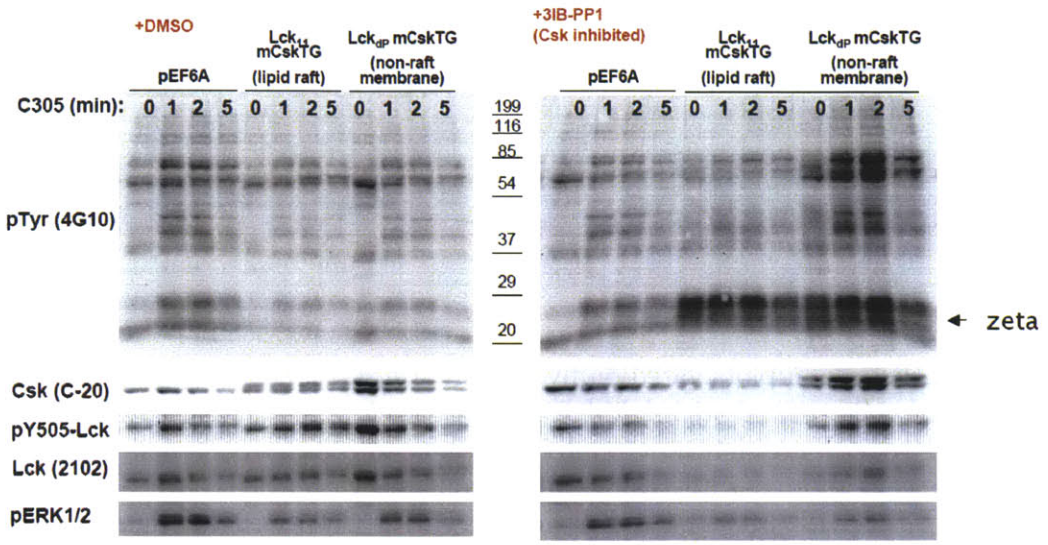


Figure 4-2: Western blots showing the induction of pCD3ζ upon activation[88]. The left half of the figure represents the control, where PP1 is not added. Columns labeled 0 represent no addition of α-CD3; Columns labeled 0, 1, 2 and 5 represent activation by C305 (α-CD3) for 1, 2 and 5 minutes respectively.

4.2 Model

The major conclusions from the experiment are:

- Activation of T-Cells by adding CskAS and the PP-1 is stronger than activation using α-TCR.
- The level of pCD3 is resting T-Cells is constant, independent of whether CskAS has been transfected in (and what the activity of the transfected CskAS is).
- The amount of total TCR on the surface of resting T Cells is directly related to the activity of CskAS added.

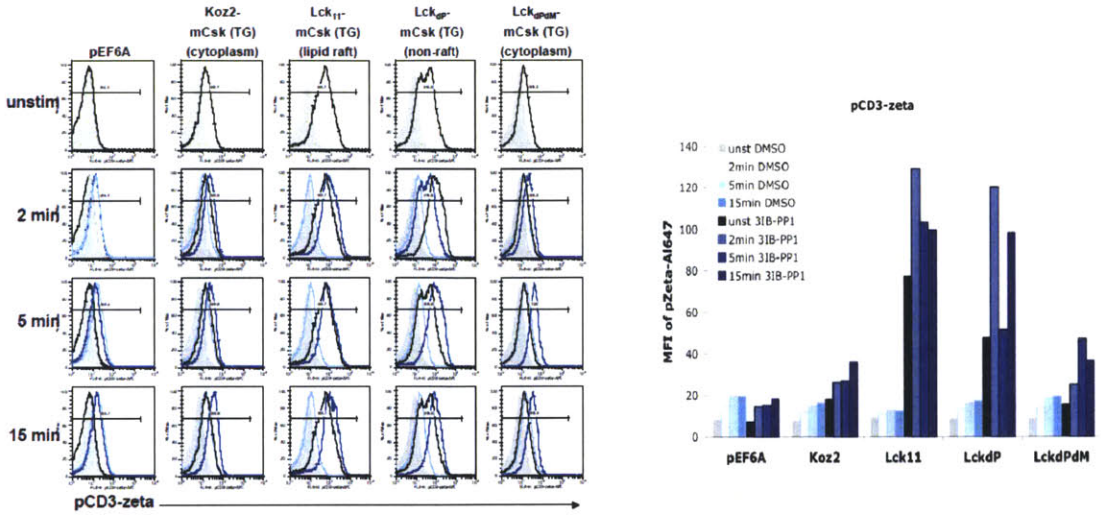


Figure 4-3: Histogram plots gated on live, GFP+ cells[92]. Left, grey histograms are unstimulated cells treated with DMSO; black overlay are unstimulated cells treated with 3IB-PP1 (both are shown in graphs of stimulated cells for reference). Light blue overlays are cells treated with DMSO and stimulated with C305, dark blue overlays are cells treated with 3IB-PP1 and stimulated with C305. Right, quantitation of mean fluorescence of pCD3-zeta.

The reasons for these experimental observations are not obvious, so we make a model involving the key molecules and interactions present in this system to try to understand the underlying mechanisms.

4.2.1 Full Model

The model consists of the key molecules in the system: Lck, CD45, Csk, and TCR. Lck has two phosphorylation sites, as described in the previous chapter, which are phosphorylated by Csk and Lck itself; the various states of Lck have different kinase activities towards CD3. CD3 is created by protein synthesis, phosphorylated by Lck, and internalized and recycled. Internalized pCD3 is degraded. Activation of the cell is measured in terms of pCD3 and RasGTP, the latter being created via Sos (which itself is created from pCD3 - this is a shortcut for a series of biological processes involving pCD3, Zap70 and LAT). The nature of the process of knocking off CskAS using PP1 yields a transient response if no positive feedback loop is present (but real responses last for tens of minutes, as the data has shown). It is known that this

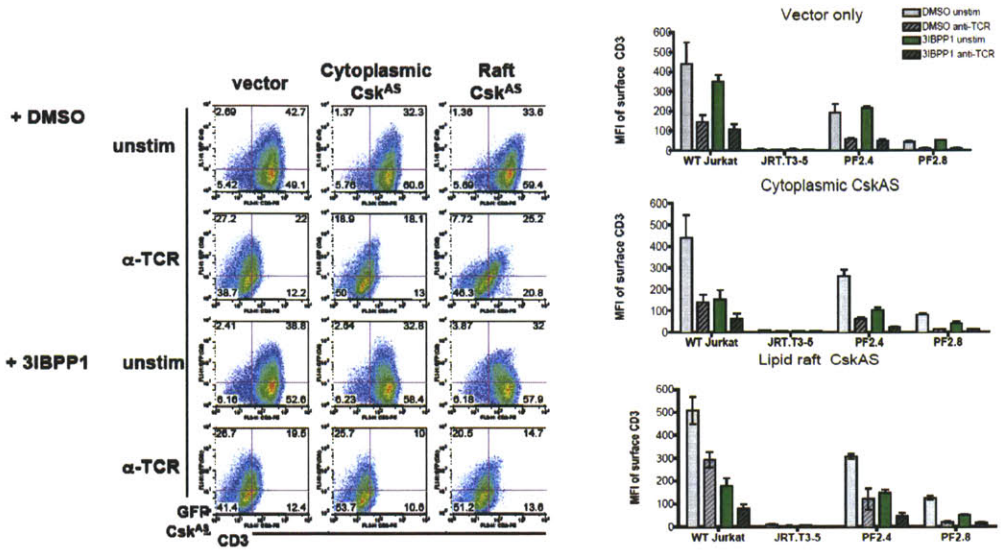


Figure 4-4: Left: FACS data showing joint population distributions of CskAS and CD3. Right: Experiments repeated in T cell lines with reduced or lacking TCR surface expression: WT Jurkats (CL1), TCR beta-deficient (JRT3.3-5), TCR beta-deficient reconstituted with TCR beta cDNA (PF2.4 = 60% of WT; PF2.8 = 15% of WT). The data represent 3 independent experiments, combined to show error bars where available[92].

activation is dependent on Zap70 and LAT ([92], data not shown), so a positive feedback loop has been put in by which Ras stabilizes the activated forms of Lck. Each reaction is modeled using chemical kinetics, and this results in a system of coupled nonlinear rate equations.

The full model consists of a set of ODEs that govern the system. We assume a plasma membrane area of 4 μm^2 and the total depth 20 nm, yielding a volume of 0.08 μm^3. Rate constants and concentrations have been converted to this volume (many of the values are estimates, and we expect the basic results to remain qualitatively the same even when rate constants are changed, but a full sensitivity analysis has not been performed). Within this box, we assume that all molecules are well-mixed, and no distinction is made between a membrane-bound and cytosolic reaction. A full list of reactions, rate constants and concentrations used are given in the supplementary material (Section B).

The simulations are performed using MATLAB's Simbiology Toolkit[105]. Cells are initialized to the normal resting state, and the system is allowed to equilibrate. At time $t = 10^4$ s, CskAS is transfected into the cell, and the cell allowed to equilibrate once again. At $t = 2 \times 10^4$ s, PP1 is added to the cell which "knocks off" the CskAS that had been added. We plot the concentration profiles of various molecules of interest as a function of time.

Figure 4-5 shows plots of concentrations of the various molecules as a function of time. Figure 4-5A suggests that the level of CD3 increases dramatically upon addition of CskAS and returns to basal after addition of PP1, but the level of phospho-CD3 at equilibrium is always the same. Figures 4-5B shows the activation of downstream signaling molecules as a result of adding PP1 at $t = 2 \times 10^4$ s. Figures 4-5C and D show the states of various forms and phosphorylation sites on Lck. The sharp transient that leads to activation at $t = 2 \times 10^4$ s is seen, and the new stabilized is stabilized by the feedback loop (the amount of active Lck after $t = 2 \times 10^4$ s is higher than the resting state, before $t = 10^4$ s.

To look at the other aspects of the experiment, we re-run the model for a range of CskAS amounts transfected in at 10^4 s (Figure 4-6). Figure 4-6A shows that in resting cells, the amount of CD3 increases with CskAS (corresponding to experiments depicted in Figure 4-4) and the amount of CD3p is constant (Figure 4-3). The strength of activation upon addition of PP1, shown in Figure 4-6C and measured as the maximum value of pCD3, increases with the amount of CskAS added (the analog of experiments shown in Figure 4-3, right panel).

The model shows a couple of interesting features. First, there is a sharp spike in the activity of Lck upon addition of PP1. This is possibly due to the mechanism of regulation of the various states of Lck. Upon decreasing the level of Csk (by PP1), there will be an accumulation of basal Lck, and due to its activation by autophos-

phorylation, a spike in the level of active Lck as well before it returns to the level in resting cells (this spike is partially stabilized by the feedback loop that has been put in, in the model). The second feature is the amount of CD3 is high just before activation. The combination of these two effects, a sharp increase in the activity of the kinase (Lck) as well as an abundance of its substrate (CD3) seems to be what leads to the extremely strong activation of T-Cells by this method.

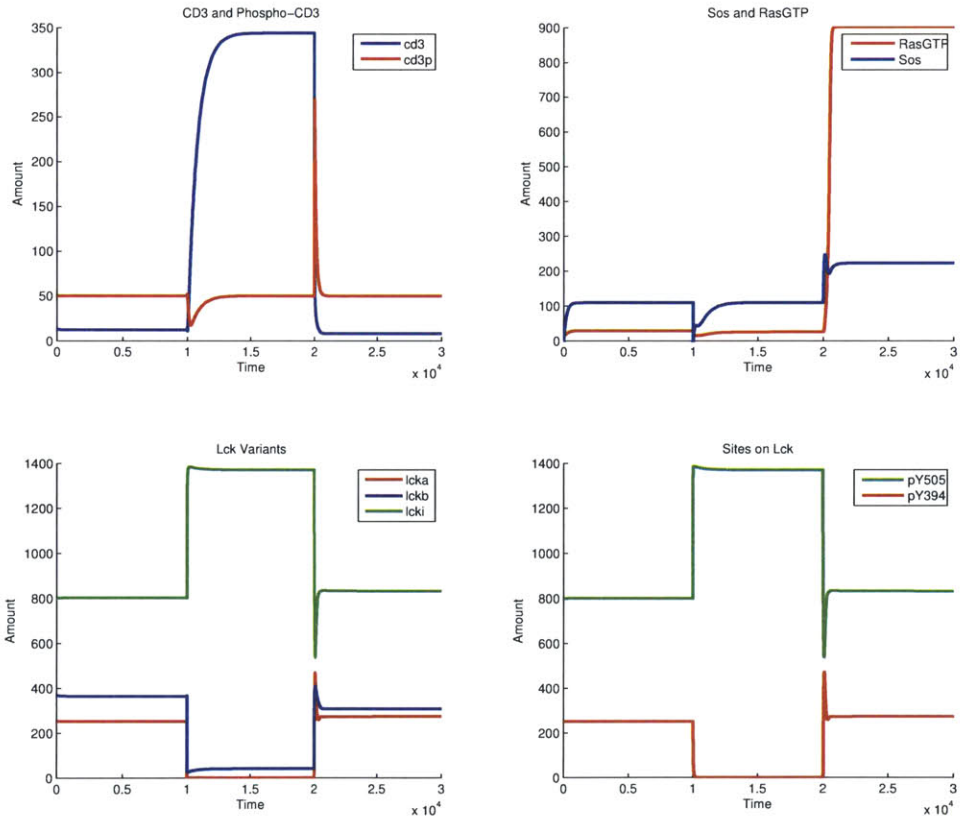


Figure 4-5: Concentration profiles from the full model: plots of (A) CD3, phospho-CD3 (B) Recruited Sos and RasGTP (C) Levels of the various states of Lck (D) Levels of the phosphorylation sites on Lck. CskAS was added at $t = 10^4$ s and PP1 at 2×10^4 s. The numbers on the axes are estimates and these qualitative features are representative.

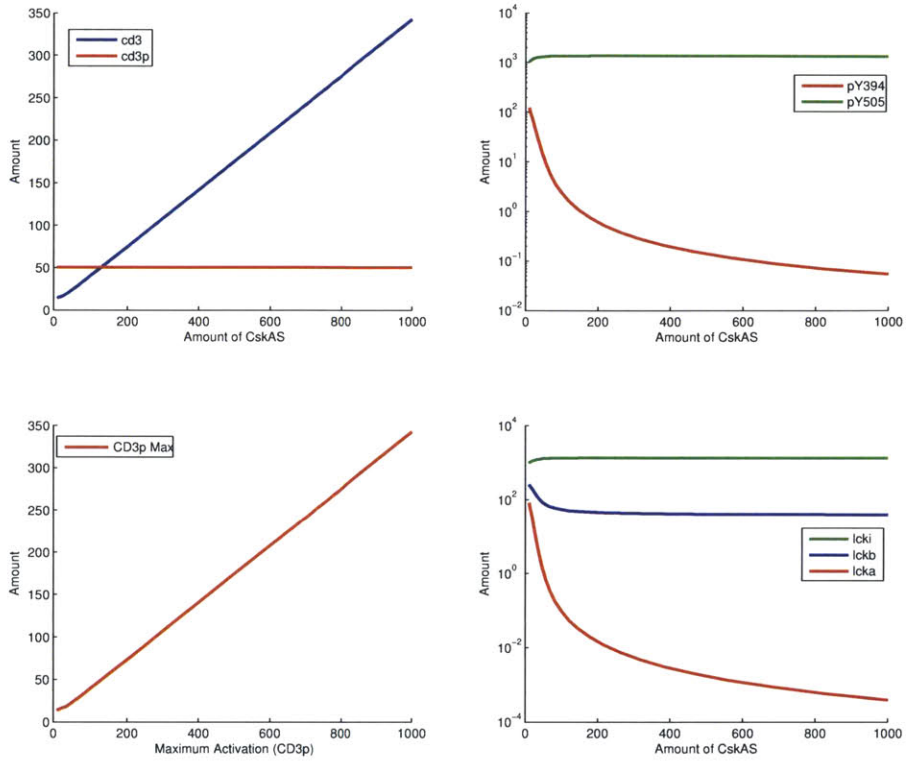


Figure 4-6: How the results of the full model vary with the amount of CskAS. Concentration profiles are (A) CD3, phospho-CD3 (B) Levels of the phosphorylation sites on Lck (C) Strength of activation, measured by the maximum of the CD3p peak (D) Levels of the various states of Lck. The numbers on the axes are estimates and these qualitative features are representative.

4.2.2 Toy Model

So far, we have shown that a model with many of the biological reactions present in the system successfully capture the essence of the experiments that were depicted. In this section, we use simple toy models to gain a deeper understanding of the mechanisms involved.

In general, the CD3 network can be described thus: CD3 is produced at some rate β (which may be dependent on the state of the cell, for example). It can be internalized (at rate k_I, say) and recycled (at rate k_R). It can be phosphorylated by Lck, and phospho-CD3 may be dephosphorylated(at rate k_u). Phospho-CD3 is internal-

ized similarly to unphosphorylated CD3 (actually, both are internalized by the same mechanism), but internalized pCD3 is then degraded (at rate k_d, say). The network is described in Figure 4-7.

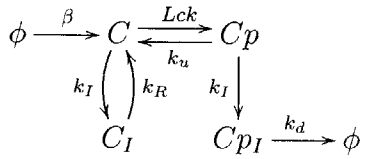


Figure 4-7: A toy model of CD3/TCR regulation and activation. CD3 (C) is created at a constant rate β by protein synthesis, is phosphorylated by Lck (L) to pCD3(C_p). C_p can get dephosphorylated; both C and C_p get internalized at rate k_I and recycled at rate k_R; internalized C_p degrades at rate k_d.

We can write a set of ODEs that describe the evolution of the system:

$$\begin{aligned}
\frac{dC}{dt} &= \beta\left(C_p\right) - k_I\left(C\right) + k_R\left(C_I\right) - k_P\left(C, L\right) + k_u\left(C_p\right) \\
\frac{dC_I}{dt} &= k_I\left(C\right) - k_R\left(C_I\right) \\
\frac{dC_p}{dt} &= k_P\left(C, L\right) - k_I\left(C_p\right) - k_u\left(C_p\right) \\
\frac{dC_{pI}}{dt} &= k_I\left(C_p\right) - k_d\left(C_{pI}\right)
\end{aligned} \tag{4.1}$$

In these equations, the brackets imply "a function of", so for example, the rate of creation of CD3, β, is a function of the amount of phospho-CD3 (which presumably specifies the state of the cell, whether it is activated or not). The steady state solution to this system of equations is: $\beta\left(C_p\right) = k_I\left(C_p\right), \quad \beta\left(C_p\right) = k_p\left(L, C\right)$

The first of these solutions implies that the steady state amount of pCD3 is given by the solution of one equation, which is independent of the amount of Lck or unphosphorylated CD3, i.e. the steady-state level of pCD3 is a constant, given by the solution of that equation. The second of these two describes the relation between pCD3, Lck and total CD3 present in the system. The relation between the various molecules is not evident, however, from these solutions, so we simplify them to uncover the qualitative trends.

$$\phi \xrightarrow{\beta} C \xrightarrow{L} C_p \xrightarrow{k_d} \phi$$

Figure 4-8: A simplified toy model of CD3/TCR regulation and activation. CD3 (C) is created at a constant rate β by protein synthesis, is phosphorylated by Lck (L) to pCD3(C_p), which degrades at rate k_d.

A simplified form of the model which keeps all the necessary detail is as follows: CD3 is created at a constant rate β, is phosphorylated in a second-order reaction by Lck (L) with a rate constant k_p to form phospho-CD3 (C_p); pCD3 degrades with rate constant k_d. This is shown in Figure 4-8.

The set of ODEs describing the simplified system is given in Equation 4.2.

$$\begin{aligned} \frac{dC}{dt} &= \beta - k_p \cdot L \cdot C \\ \frac{dC_p}{dt} &= k_p \cdot L \cdot C - k_d C_p \end{aligned} \tag{4.2}$$

The steady-state solutions to these equations are $C_p = \frac{\beta}{k_d}$ and $C = \frac{\beta}{k_p L}$. This suggests that the steady-state amount of pCD3 is given by the ratio of its production and degradation rate (even though it is CD3, not the phospho-form which is created by protein synthesis), and is a constant. The total amount of CD3 (unphosphorylated) is inversely proportional to the amount (or in general, activity, since there are many states) of Lck present in the system. In this case, since the amount of Lck present in the system is a function of (inversely related to) the activity of Csk.

This CD3 creation-degradation machinery is an example of a biological network that displays "perfect adaptation"[106], which means that the output of the network (CD3p) is constant, independent of the input (Lck). Network topologies exhibiting such perfect adaptation are found in many biological systems, including bacterial chemotaxis[106], migration of neutrophils[107], osmo-response in yeast[108], sensory cells in mice[109] and calcium homeostasis in mammals[110]. This topology is an example of perfect adaptation by integral feedback control[111].

4.2.3 Consequences of the model

It is interesting to explore a few of the properties of a network whose structure is as given in Figure 4-8. We use the simplified system to look at dynamics (these are difficult to test experimentally, however). The question here is, if the activity of Lck L is a function of time, how does the amount of phospho-CD3 behave?

Assume we start with the system described before, at the steady state $C_p(0) = \frac{\beta}{k_d}$, $C(0) = \frac{\beta}{k_p L_0}$. We examine three conditions for how L varies with time:

1. **Step change**: L changes instantaneously to L_1, i.e. $L(t) = L_1$ (Equation 4.3).

$$\begin{aligned} \frac{dC(t)}{dt} &= \beta - k_p \cdot L_1 \cdot C(t) \\ \frac{dC_p(t)}{dt} &= k_p \cdot L_1 \cdot C(t) - k_d C_p(t) \end{aligned} \tag{4.3}$$

These equations can be solved with the initial conditions given above, and the solution is given in Equation 4.4.

$$\begin{aligned} C(t) &= \frac{\beta}{k_p L_1}\left(1 + \left(\frac{L_1}{L_0} - 1\right) e^{-k_p L_1 t}\right) \\ C_p(t) &= \frac{\beta}{k_d} + \frac{\beta}{k_d - k_p L_1}\left(\frac{L_1}{L_0} - 1\right)\left(e^{-k_p L_1 t} - e^{-k_d t}\right) \end{aligned} \tag{4.4}$$

This step change models the extreme limit of an infinitely fast change in the activity of Lck (in reality, even if the change in Csk level is extremely rapid, say upon PP1 introduction, the response of Lck will not be so). However, we can use this limit to ensure some basic checks are satisfied: the system goes back to the steady state for $C_p = \frac{\beta}{k_d}$, and is dependent on the rates of phosphorylation and degradation.

2. **Ramp**: L varies as $L(t) = L_0 + \alpha t$ (Equation 4.5).

$$\begin{aligned} \frac{dC(t)}{dt} &= \beta - k_p \cdot (L_0 + \alpha t) \cdot C(t) \\ \frac{dC_p(t)}{dt} &= k_p \cdot (L_0 + \alpha t) \cdot C(t) - k_d C_p(t) \end{aligned} \tag{4.5}$$

with the initial conditions given above. These equations are hard to solve exactly analytically; however, we can calculate the long-term limit of the solution (as $t \to \infty$) using the final value theorem of Laplace transforms[112]:

$$\lim_{t\to\infty} C_p(t) = \lim_{s\to 0} s\tilde{C}_p(s) = \frac{\beta}{k_d} \tag{4.6}$$

where $\tilde{C}_p(s)$ is the Laplace transform of $C_p(t)$. The final steady state is the same as the initial state and is independent of the rate of change of Lck activity α; however, the transient depends on α as shown in Figure 4-9A. The interesting thing about his result is that even though the activity of Lck is growing without bound, the steady state of C_p is still reached. Due to the nature of the final value theorem and the limit, we can say that any increase in L that is linear or sub-linear with time will reach the steady state. Figure 4-9B shows that the faster the change in L (steepness of the ramp, in this case), the higher the magnitude of the response.

3. **Sinusoid**: We can calculate the linearized frequency response of $L(t)$ when it is a sinusoid, say with frequency ω (Equation 4.7).

$$\begin{aligned} \frac{dC(t)}{dt} &= \beta - k_p \cdot L(t) \cdot C(t) \\ \frac{dC_p(t)}{dt} &= k_p \cdot L(t) \cdot C(t) - k_d C_p(t) \end{aligned} \tag{4.7}$$

In this case $L(t) = L_0 + a \sin \omega t$, and the linearization approximation means we set $C(t) = \frac{\beta}{k_p L_0} + \delta c(t)$ and $C_p(t) = \frac{\beta}{k_d} + \delta c_p(t)$. The linearized equations are

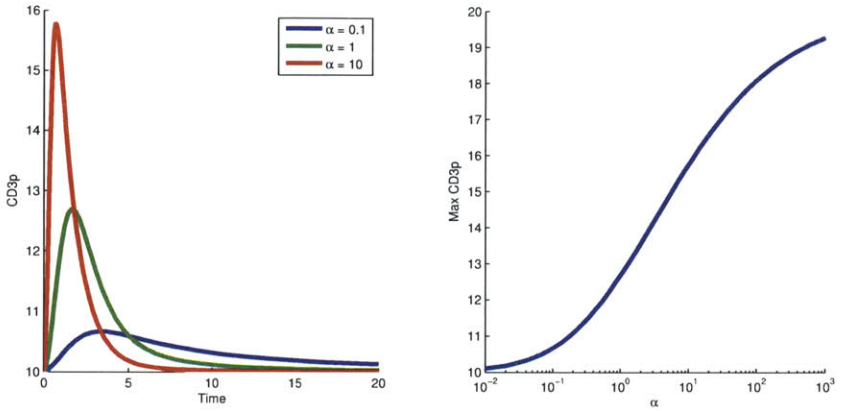


Figure 4-9: (A) Time profiles of CD3p for different activation rates α for $L(t) = L_0 + \alpha t$. Faster activation leads to faster but higher-magnitude transient responses. The final value of CD3p is always the same as the initial value. (B) Variation of the magnitude of the response as a function of steepness of the ramp α.

given in Equation 4.8.

$$\frac{d}{dt}\begin{bmatrix} \delta c\,(t) \\ \delta c_p\,(t) \end{bmatrix} = \begin{bmatrix} -L_0 & 0 \\ L_0 & k_d \end{bmatrix} \cdot \begin{bmatrix} \delta c\,(t) \\ \delta c_p\,(t) \end{bmatrix} + \frac{\beta}{k_p L_0} \cdot a \cdot \sin(\omega t) \begin{bmatrix} -1 \\ 1 \end{bmatrix} \tag{4.8}$$

The amplitude of the response as a function of frequency of Lck variation can be analytically calculated:

$$\left|\frac{\delta c_p}{a}\right| = \frac{\frac{\beta}{k_d} \cdot \omega}{\sqrt{\left(k_p^2 L_0^2 + \omega^2\right)\left(k_d^2 + \omega^2\right)}} \tag{4.9}$$

This is a band-pass filter, meaning it has a maximum amplitude at intermediate frequencies. Biologically, the reason is the following: if the Lck concentration fluctuates too rapidly, there isn't enough time for it to affect the rate of formation of phospho-CD3; if the changes in Lck concentrations are too slow, the CD3 creation-degradation mechanisms have time to adjust and hence the pCD3 concentrations, again, do not change much. Intermediate rates of change of Lck activity, too fast for CD3 creation-degradation to adjust but slow enough that it does not average out the phosphorylation of CD3, yields responses of the

highest possible magnitude.

These results illustrate a key feature of the model: not only is the level of pCD3 a constant, but sharp and high-magnitude responses are required to produce significant amounts of pCD3. If we assume that there is a barrier to be crossed in terms of number of pCD3 molecules to start downstream signaling, this mechanism ensures that the height of the barrier is independent of the concentration of CD3 and activity of Lck (and other parameters which affect the cell through CD3 or Lck). Small-magnitude but extremely high-frequency changes (noise) and very slow, sustained changes in Lck activity are filtered out by this regulatory network, allowing only rapid, high-magnitude changes in Lck activity to activate the T-Cell.

4.3 Discussion

We have built a model of activation of T-Cells using the negatively-regulating kinase CskAS and its inhibitor PP1. The really strong activation of T-Cells using CskAS/PP1 is likely a combination of two effects: an increased concentration of unphosphorylated CD3 molecules on the cell surface before addition of PP1 and a sharp spike in the activity of Lck due to the nature of its regulation. This behavior is captured not only by an ODE model consisting of many of the relevant reactions of these molecules, but also by a simple toy model consisting of just CD3 and Lck. These models also predict other features seen in the experiment: increasing the activity of Csk increases the amount of CD3 present on the surface of the T-Cell before activation, and the level of phospho-CD3 levels at long time is constant. This toy model suggests that the topology of the network involved in the Lck-CD3 regulation mechanism displays a form of perfect adaptation, and ensures that the barrier to activation (in terms of the amount of pCD3 needed to set off signaling) is independent of amount of CD3, activity of Lck and other parameters in a cell which are highly variable. This mechanism also ensures that slow changes in activity of Lck, even if sustained, does not set off downstream signaling, and that rapid and high-magnitude

changes in Lck activity are required to activate T-Cells.

Chapter 5

Properties of scale-free signaling networks under a directed evolutionary pressure

5.1 Introduction

A common approach to the treatment of many cancers involves the targeting of specific molecules in signaling pathways[113]. The main problem with this approach is the great degree of heterogeneity shown by even a single type of cancer, for example, in recent studies by Hartzell et. al.[114] of Ras signals in T-cell leukemia. Given the heterogeneity of the downstream signaling responses, it is unclear which proteins would make the optimal drug targets. A knowledge of the specific mutations has been shown to be useful: for example, personalizing medicine using gene-expression patterns in cancers has been attempted[115]; however, complete sequencing of cancers is a time-consuming and expensive process and in this work we ask whether there are constraints or patterns in these mutations.

Each specific instance of a tumor is fundamentally characterized by a set of mutations which have occurred in the development of that tumor; since mutations are

stochastic, different realizations of the same cancer have different sets of mutations. Mutations are changes the nucleotide sequence of DNA; some changes in DNA affect the protein sequence (through the processes of transcription and translation) and hence protein structure. These changes in turn affect protein-protein interactions and therefore intracellular signaling, and the response of the cell is a function of the levels of certain signaling molecules. Going from a change in DNA sequence to cellular response, while fundamentally sound, is impractical on many levels and in this work we look to coarse-grain out some of the steps to simplify the approach of going from mutations to cellular function.

Normal cells in the human body are optimized for certain functions, with an "error-correcting" machinery to set the cell right in case anything untoward happens; however, cancerous cells have somehow managed to bypass these error corrections, and evolve to maximize their own growth[116]. The growth of cells depends on the levels of certain downstream effector molecules (such as c-Myc) within their metabolic and signaling network. In this way, cancer cells are selected for growth and not just allowed to mutate freely. This selection for growth should constrain the patterns of mutations in cancerous cells in such a way as to make them more predictable and easier to understand. The question is whether the mutations of proteins in a signaling network that aim to optimize some function of the network (in this case, growth) lead to predictable patterns (or probability distributions thereof) of mutations. Network evolution models have been built from scratch for specific functions [117], but as of now, there exists no known study trying to connect mutations in the topologies of general classes of signaling networks to function. Studies have also been performed of the importance of nodes to the overall viability of the network[118], but such studies are too blunt for modeling cancer mutations, where there are more subtle changes to network topology in terms of only changes in certain protein-protein interactions, rather than wholesale removal of proteins.

The fact that there is heterogeneity in cancer falls out naturally from the above

framework: since mutations are stochastic, there might be many patterns of mutations which lead to optima in growth, i.e. if the growth function (energy landscape) is rough. Each optimum in topology space would therefore represent a different pattern of mutations which maximizes growth of the cell. One of the questions we ask in this model is, can one understand what the heterogeneity of cancers depends on? This approach assumes several things, which are currently not very well known: a knowledge of the underlying signaling network, which one may have a first guess at from, say, protein-protein interaction data; a set of possible changes in the network (mutations); a model for propagation of signal for each node (an input/output relation for each node) and a model for growth as a function of signals. We put in a set of biologically motivated first approximations to these underlying features and see how the results vary with them, and explore if this model has any general detail-independent features that we can understand.

•

In this work we have built a biologically motivated model to try to predict patterns in mutations in cancers, focusing on changes in protein interactions rather than at the level of DNA. We attempt to quantify the changes the topology of signaling networks due to mutation in terms of the basic metrics of the network (degrees, loops, paths) that differ between optimized and randomly mutated networks. This general framework would be applicable not just to protein kinase signaling networks, but also to gene regulatory networks, cytokine signaling, etc. which share the same basic idea of directed interactions between components which can change as a result of mutation. The key results are (1) degree distributions of nodes (proteins) in the network are not a good metric for measuring mutation likelihoods (2) the more constrained the system, the clearer the pattern in mutations in the system (3) loops and paths are modified differently in optimized networks as compared to randomly mutated networks (4) the likelihood of seeing multiple optima, which is hypothesized to be the analog of variants of the same cancer, increases with the number of loops and paths in the networks, and is a stronger function of feedback loops.

5.2 Building the Model

5.2.1 Connecting mutations to intracellular signaling

To try to understand how the mutation-selection evolution process could lead to patterns of mutations in cancers, we shall try to build the simplest possible model that gives this behavior. The key, therefore is to try to connect structure to function: to figure out how changes in DNA sequence lead to changes in growth of the cell. Generally speaking, this happens in a sequence of two stages: mutations in DNA change the structure of proteins that they transcribe/translate to, and changes in structure affect the function of these proteins. Protein functionality, in general, happens through interactions with other proteins, predominantly by changing the characteristics of that protein-protein binding process, which in turn leads to a change in the "signals" being transduced through the cell, rates of metabolic processes, and so on. We assume that growth of cells is ultimately determined by the level of certain downstream signaling molecules ("effector" molecules) within these cells. The mutations as described above change the PPI network, which in turn changes the levels of these effector molecules to change the growth. Hence, to relate mutations to growth, we need a two-part model: to relate mutations to the signals being transduced within these cells, and to connect these signals to the growth.

To look at how mutations or changes in PPI networks affect the cellular signals being transduced within cells, we need a way of relating network structure (topology) to signaling. Neural networks provide us with a way of connecting network topology to signals. The basic equation of neural networks is:

$$\tau_i \frac{dx_i}{dt} + x_i = f_i\left(\sum_{j=1}^{N} W_{ij}x_j + c_i\right) \tag{5.1}$$

In neural network models, x_i represents the signal at node i (neuron i), τ_i is the time constant of node i, W_{ij} is the connection from node j to node i, c_i is the basal signal of node i and f_i is a non-linear function representing the input-output relationship for

node i. These classes of models have an easy way of representing topology W_{ij}, and relating them to signals x_i. For the purposes of our model, we use x_i to represent the output of protein i, which is the "signal level" of node i. This could typically be the level of the phosphorylated state of protein i, for example. W_{ij} would represent the activity of protein j on i, which would be zero if protein j does not influence protein i. Note that matrix W does not need to be symmetric, i.e. in general $W_{ij} \neq W_{ji}$ and feedback loops are allowed. Positive values of W_{ij} represent activation, and negative values represent inhibition. We set c_i, which represents the basal signal of node i, to zero for all nodes except the node with the driver mutation, and set the basal signal of that node to a high value, i.e. c_1 is large. Note that, if there was no driver mutation, all c_is would be zero, and all our choices of input/output relationship for a node have to satisfy the condition that $f(0) = 0$, meaning that if an intermediate node does not have any net signal input, it does not have any output as well. The input/output relationship for each node, $f(x)$, is in general a form that saturates in some way. We are also looking at the limit of long-time responses, in which the cancerous cells have time to fully adjust to a mutation before making another. In this limit, the cellular signaling mechanism is assumed to have reached a steady state, so the equation that governs the signaling output of protein i is:

$$x_i = f_i\left(\sum_{j=1}^{N} W_{ij}x_j + c_i\right) \tag{5.2}$$

To find the entire signaling profile within the cell, we solve simultaneously the set of equations, one for each component of the network:

$$\underline{x} = \underline{f}\left(\mathbf{W}\underline{x} + \underline{c}\right) \tag{5.3}$$

The solution to this set of coupled nonlinear system of equations describe the signaling profile within the network of topology $\mathbf{W}$ and basal signal $\underline{c}$.

Values of the individual elements of W_{ij}, in general, would be representative of strengths of protein-protein interactions within this network. These values can be

either positive, representing activation, negative, representing inhibition, or zero if protein j has no effect on protein i. This model inherently assumes linear effects of proteins: for example, there is no $x_i x_j$ term within f (which could, for example, model a dimerization step), or higher (trimeric) complexes etc. Another model for concisely representing biochemical reaction networks are S-systems[119], but does not have such a simple representation of topology.

5.2.2 Mutations on this framework and the loss-of-function approximation

Mutations in this framework are changes in elements of W. In general, elements of W could change in magnitude, sign, or both. Changes of sign would imply that an interaction that was previously activating has now become inhibitory, or vice versa, and are unlikely; changes in magnitude could be a measure of, for example, changes in binding strength between proteins before and after the mutation. In principle, one could go from structure (mutation in DNA), leading to change in codons transcribed and protein translated, to change in protein structure as a function of sequence, to change in protein interactions. In a complete framework, each element of W_{ij} would be allowed a set of possible values obtained from how mutations change the structure of proteins involved in that interaction. However, this is intractable at this point, so we just coarse-grain out these details and say that each element of W_{ij} can have two values, its initial value (unmutated interaction) and zero. The zero value is the limit where mutations completely destroy protein binding, i.e. the loss-of-function limit; it is much more common than the other limit where mutations can create interactions between proteins.

5.2.3 Correlated Systems

When each element of **W** is allowed to vary independently, the mutations to various links connected to one node are uncorrelated, implying that the various protein-protein interactions that one protein participates in can be modified independently by mutation. Any correlation between different effects of one protein due to a mutation are modeled by changing groups of interactions of the same protein, i.e. sets of elements of W_{ij} with the same i or j, together. To model this, we divide each protein up into a set of domains, each domain consisting of a subset of all the edges which that node is a part of. A mutation to that domain affects the entire subset of links which comprise the domain; for a link between two proteins to be retained, the assumption made is that both domains of which that link is a part of has to be wild-type. A mutation in either of the domains of the two proteins at the ends of that edge destroys that interaction. Mutations to a domain also destroy all links which that domain is involved in; which is the origin of correlations between effects on protein interactions due to a mutation. The number of domains increases with the degree of that protein in the interaction network, and input and output links to a protein are considered to be parts of separate domains.

5.2.4 Growth as a function of certain effector nodes

Growth is considered to be determined by the levels of activation of certain "effector" signaling nodes. The number of effector nodes is a variable; one effector node is always chosen to be the greatest distance away from the source mutation so that the maximum fraction of the network affects signaling. The other effector nodes in the network are chosen randomly. In general, with an increasing number of effector nodes, a greater fraction of the network contributes to signal modulation and mutations to nodes or links within this fraction of the network affects growth.

5.3 Solving the model

5.3.1 Constructing the network

Protein interaction networks are generally scale-free[120], with a degree distribution of $P(k) \approx k^{-\gamma}$. The specific choice of the exponent γ depends on the system; different networks have different values of γ. For this work, γ has been chosen to be equal to 1.78 [121].

Scale-free networks of size N=50 with exponent $\gamma = 1.78$ are constructed using the method described in [122]. Once edges have been assigned according to this algorithm, we assign directions and signs (activating or inhibiting) randomly with equal probability. An example of a network generated in this way is shown in Figure 5-1.

The initial mutation is chosen to be the node in the network which has the maximum of nodes downstream to it, either directly or indirectly. This helps in reducing the edge effect mentioned in the previous section, and makes sure that the largest number of mutations are not neutral or useless to the signals in the cell. The initial mutation is chosen to be the only node for which $c_i \neq 0$, so that there is a signal in the network.

5.3.2 Total model size and trade-offs

Real protein interaction networks are of the order of thousands of nodes, the number of distinct proteins in the human proteome being of the order of 30000 [123]. However, it is not efficient to simulate such large networks computationally for the present purpose for reasons that will be described. However, if one goes to too small a network, most of the nodes will be "edge nodes", and will not affect others within the network. To look at this metric for our simulations, we plot a histogram of the number of nodes downstream of a particular node in these networks, as a function of network size in figures 5-2A-C. The figure shows that the network can be divided into two parts, a

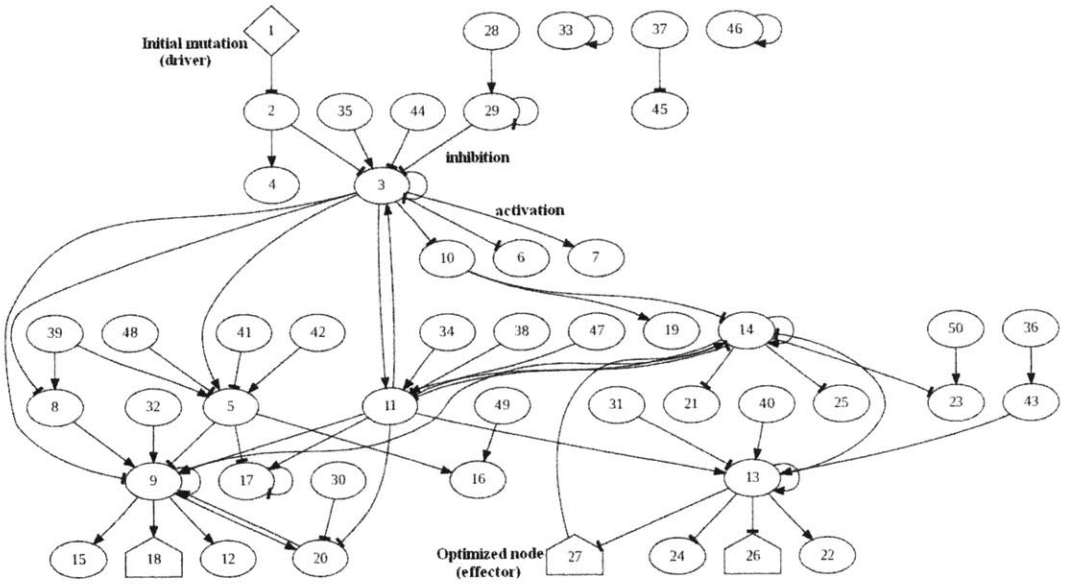


Figure 5-1: A sample signaling network generated by the algorithm described in text. Each node or oval represents a protein; directed arrows represent protein influences on each other. The model includes activating (arrows) and inhibitory (bars) protein-protein interactions. The driver mutation is indicated by a diamond, and effector nodes (which are optimized) as pentagons.

"bulk" where most of the network is connected internally, and a set of edge nodes, which are not connected to most others. With increasing total size of the network, we see that the separation between the two becomes clearer; so if a network is chosen that is too small in size, one would expect most of the nodes to be edge nodes, unlike in a real PPI network, which often consists of thousands of proteins. However, the simulations described later are computationally very expensive, and simulation time scales very strongly with the size of the network. For the purposes of the rest of this work, we have used networks of size 50 as they represented a reasonable tradeoff between reasonable network size and ability to simulate in a reasonable time.

5.3.3 Initial Mutation

The initial mutation is chosen to be the node in the network which has the maximum of nodes downstream to it, either directly or indirectly. This helps in reducing the

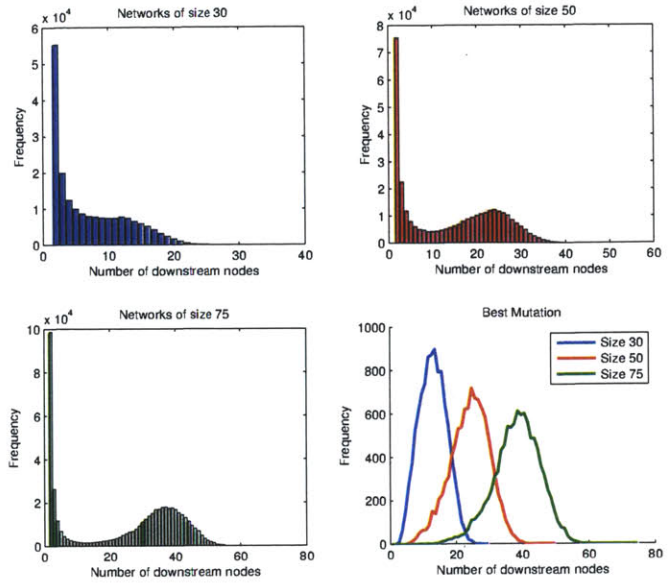


Figure 5-2: A, B and C: plots of histograms of number of nodes downstream of any given node. There are broadly two parts to these networks, the bulk of the network and the edge; the edge nodes in these networks become less pronounced as the size of these increases. figure D is a plot of the maximum effectiveness of a mutation in these networks, i.e. the maximum of figures A-C.

edge effect mentioned in the previous section, and makes sure that the largest number of mutations are not neutral or useless to the signals in the cell. The initial mutation is chosen to be the only node for which $c_i \neq 0$, so that there is a signal in the network.

5.3.4 Choices for functional forms

Like in the case of neural networks, one needs to make a choice for the input/output behavior of each node. Two choices are made in this work. The first one is a sigmoidal function, as a large number of cancerous mutations are disproportionately present in protein kinase domains[124], and switches are a ubiquitous feature of kinase signaling cascades:

$$f(x) = \begin{cases} \frac{x^2}{1+x^2} & \text{if } x \geq 0 \\ -\frac{x^2}{1+x^2} & \text{if } x < 0 \end{cases} \tag{5.4}$$

This is representative of many kinds of signaling motif behavior like multiple phosphorylation motifs which yield sigmoidal behavior. Another assumption that is made is that each node is a Potts spin, with an output of -1, 0 or 1 depending on the input:

$$f(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases} \tag{5.5}$$

This Boolean-type approximation is also commonly used to model signaling networks[125].

Let the set of effector nodes in this system be denoted by M. The functional form of growth is also chosen; in general, increasing the level of the effector molecule should increase growth, but only up to a point. Too high a transcription/translation of some proteins will cause, for example, resource limitations in other parts of the cell, and hamper growth. High levels of certain signaling molecules also cause pathways that induce senescence. We select a functional form for the growth that is convex and has a maximum. The simplest such form, which is chosen for the base model, is:

$$g\left(\underline{x}\right) = \sum_{k \in M} \left(x_k - x_{k,max}\right)^2 \tag{5.6}$$

We have also examined the result of coupling between nodes in the growth function. For this case, we have assumed g to be of the form (where m is the number of effector nodes):

$$g\left(\underline{x}\right) = \frac{1}{m}\left(\sum_{k \in M} x_k - \sum_{k \in M} x_{k,max}\right)^2 \prod_{k \in M} \delta\left(x_k\right) \tag{5.7}$$

In this case, the signals in all the effector nodes are coupled and the sum of these signals contributes to the growth function. The last term containing delta functions is implemented in order to make sure that the network does not get completely disconnected by mutations (which was allowed in the previous case).

For the Boolean model, since the values that the signal at any node can take are discrete rather than continuous as before, the previous approach to a growth function has to be slightly modified. Since Boolean networks may exhibit regular periodic oscillations at long times, rather than just steady behavior, the average value of the signal at the effector node over a cycle has been used in place of x_k in equation (5.6)

The four classes of models that have been examined are:

1. Sigmoidal f(equation (5.4)), uncoupled g(equation (5.6)), referred to as the "base model"
2. Sigmoidal f(equation (5.4)), coupled g(equation (5.7))
3. Potts f(equation (5.5))
4. Correlated mutations

5.3.5 Solving the model

In order to look for optima using this framework, we use a Metropolis Monte Carlo algorithm. We set it up in a manner similar to that of the evolution of real cancers: start with many realizations of a given (initial) network, make mutations stochastically, compute the steady state signals in the network after each mutation, and then calculate the new growth function for systems with that mutation. If the growth increases, the mutation is accepted; otherwise, it is accepted with some probability proportional to the Boltzmann factor of the change in the growth rate.

- Generate an initial $\mathbf{W}$

- Pick a "sufficiently influential" node for the initial mutation, $c_i \neq 0$
- Pick a set of effector nodes
- For each of the above, run a set of trajectories with the same starting state
- Each trajectory consists of a set of mutation-selection steps:

  1. Solve for $\underline{x}$:

$$\underline{x} = \underline{f}\left(\mathbf{W}\underline{x} + \underline{c}\right) \tag{5.8}$$

  2. Calculate $g\left(\underline{x}\right)$
  3. Mutate $\mathbf{W}$ to $\mathbf{W}'$
  4. Solve for $\underline{x'}$: $\underline{x'} = \underline{f}\left(\mathbf{W}'\underline{x'} + \underline{c}\right)$
  5. Calculate $g\left(\underline{x'}\right)$
  6. Probability of accepting the mutation:

$$p_{accept} = min\left(e^{\beta g(\underline{x'}) - g(\underline{x})}, 1\right) \tag{5.9}$$

Figure 5-3 shows an example of a starting network evolving to two different final realizations (optima) after the Monte Carlo process.

In this framework, to calculate the steady state of the system, one needs to solve a set of coupled nonlinear equations(Equation (5.8)). Since each step involves the calculation of a new steady-state signal after making a mutation, we use the previous steady state as the initial guess and solve the system in two parts: (1) evolve the system for a fairly long time, as a set of ODEs, as in equation (5.1). Then we solve the set of equations simultaneously using a standard optimization package, MATLAB's fsolve[105].

There are various parameters within this framework that one might choose. Ten trajectories have been run with each starting network, and each trajectory consists of

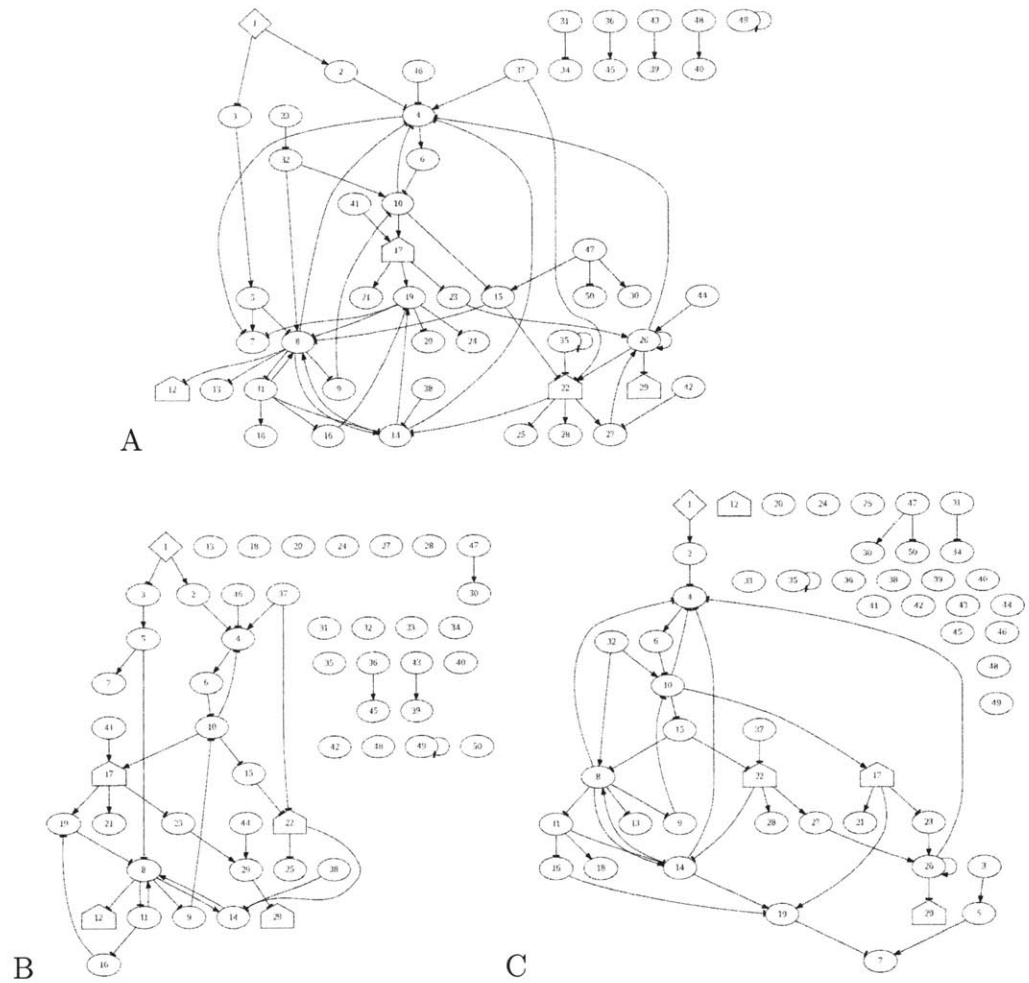


Figure 5-3: An example of a network evolving to multiple different structures. The starting network is depicted in A, and two final realizations in B and C.

10^5 attempted mutations. This Metropolis MC scheme attempts to find the topology **W** that maximizes growth g; it essentially evolves the system in topology space in the framework of an energy landscape defined by the negative of the growth function. The fitness of each topology is given by $e^{-\beta - g(x\_)}$, and one can use the parameter $\beta$to control how likely it is that unfavorable mutations will be accepted - it is an analog to inverse temperature in statistical mechanics. Simulations performed at high temperature (low β) have a greater chance of unfavorable mutations being accepted, and at high $\beta$are analogous to zero temperature where only favorable mutations are accepted. Physically, one may compare $\beta$to various things that are present in a real cancer system but ignored in this model: since this system is monoclonal, unlike real

tumors which are polyclonal, beta can be thought of as an analog to the size (number of cells) of the tumor, where larger tumors are much harder for an unfavorable mutation to take over. Of course, in real systems, even favorable mutations may not get fixed in the population, but that consideration is ignored here.

Real cancer tumors typically contain a few driver mutations, and many passenger mutations[126]. In this model, however, we are evolving the system for many steps until it does not change anymore, i.e. presumably reaching an optimum. Drivers and passengers in this scheme are not treated differently from each other. All mutations made are thought to be functional in the sense of affecting a protein-protein interaction; some of these do affect the growth and others (for example, those which are completely disconnected from the part of the network containing the initial mutation) do not. In the base model, which is the least constrained, typically 30-40 % of the initial links have been removed at the optimum, and this seems to be an overestimate. The coupled growth model, with its constraint that effector nodes cannot be disconnected completely from the network, has a much smaller number of mutations present finally.

The exact determination of whether a network topology is an optimum is computationally very expensive, so we monitor the functional value and ensure that it does not change with a large number of attempted mutations being made. The number of attempted mutations is on the order of 10^5, for networks with typically 50 nodes and a hundred or so links. We can see in Figure 5-4 that the growth usually reaches its final value in the first 10% of the attempted simulation steps after going to zero temperature, which supports the hypothesis that the simulation method is driving network topology to a local optimum. Results are similar for the other models (data not shown).

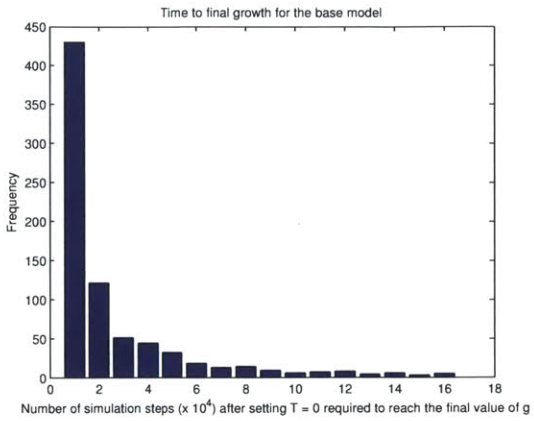


Figure 5-4: Number of simulation steps required for the network to reach its final growth value for the base model.

5.3.6 Modeling Inhibition and Escape

We can use the same framework to model the effect of adding drugs, such as inhibitors, to a signaling network. Assume, for example, that an inhibitor is added which decreases the signal output at node q by a factor β. The input-output relation for that node becomes Equation 5.10, and remains Equation 5.2 for all other nodes. We can then solve the entire system of equations, with this one equation modified, as described in Section 5.3.5.

$$x_q = \frac{1}{\beta} f_q \left(\sum_{j=1}^{N} W_{qj} x_j + c_q \right) \tag{5.10}$$

This gives us the signal profile of the network after inhibition. Since cancers can mutate subsequent to the addition of a drug, we let mutations occur on the inhibited system and the network evolve as described earlier. Ten different trajectories are run after the inhibition of each node; we look at the value of the growth function after the escape mutations occur. Only the base model has been explored in the inhibition/escape mutation context.

5.4 Results

To identify the patterns of mutations which are caused by selection toward optimal growth, we compare with a baseline of a randomly mutated network (which just consists of the original network with the same fraction of edges removed). In the correlated model, this is the probability that a domain was mutated in the final network. The randomly mutated network is then constructed by starting with the initial network and removing edges (or domains, in the fourth case) with that probability. Once that is done, we can use various metrics of networks to see how non-random the mutations are after selection, and compare the evolved cases with the randomly mutated networks.

5.4.1 Probability of Multiple optima

Multiple optima are important as a measure of the diversity of the cancer; broadly, the higher the probability of multiple optima, higher will be the likelihood of seeing heterogeneous behavior in the cancerous system. In these simulations, we look at the final growth of the ten trajectories obtained by evolving each starting network, and see how many values of the growth function are present to calculate the probability of multiple optima. Figure 5-5 is a plot of the probability of multiple optima as a function of the number of effector nodes, loops and paths in the starting network. We see that the probability of multiple optima increases with the number of optimized nodes, but then saturates; potentially, in these small systems, the number of effector nodes is a proxy for what fraction of the network is "in play", i.e. involved in signaling, and a few nodes are enough for the whole of the network to be involved. The probability of multiple optima also increases with the number of feedback loops and paths in the system, but is most affected by the number of loops. This suggests that in some way, feedback loops are inherent in the heterogeneous responses of cancer evolution with mutation. Results for other models are presented in the supplementary material (Section C).

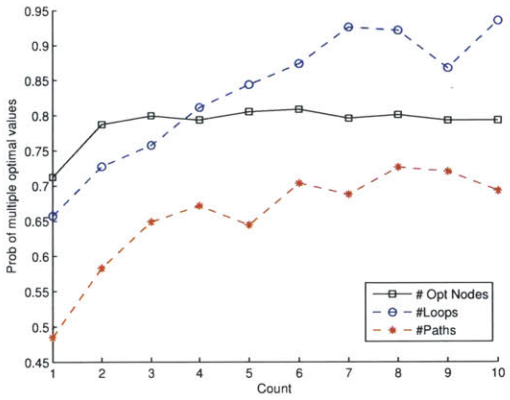


Figure 5-5: Probability of multiple optima as a function of the number of optimized nodes.

5.4.2 Degree Distributions

The most basic metric to consider when dealing with a network is the degree distribution. The degree distributions represent the probability that a given node has k connections; a comparison between optimized (evolved) networks and randomly mutated networks by this metric would reveal whether nodes of a certain degree are mutated more frequently in cancers. Figure 5-6 shows overall degree distributions for the base model. We see that the initial degree distribution (thick black) is a straight line, representing a scale-free network. The other curves are for degree distributions for the optimized networks and the randomly mutated network, with the number of effector nodes varying between one and ten. The initial degree distribution has nodes of degree one and greater; however, after mutations (removal of edges), it is possible that in networks with completely disconnected nodes result. So the other curves start at a degree of zero. The greatest difference between the initial and mutated networks is present at high degree - because of the large number of links from high-degree nodes, it is probable that mutations affect such nodes disproportionately. We also see that such mutations seem to affect the optimized and random cases similarly with regard to this metric, so it is not possible to differentiate between them using this metric.

A similar situation is seen for the Boolean model. However, in the case of coupled effector nodes, we see that a lot fewer mutations occur, and they are mostly at high-degree nodes. This is probably due to the constraint that enforces that none of the effector nodes get completely decoupled from the rest of the network. Even in these cases though, it is hard to distinguish between the randomly mutated and optimized networks. The correlated mutation model shows similar results to the coupled growth model: the constraint that links are coupled seems to result in much fewer, high-degree mutations (see Supplement).

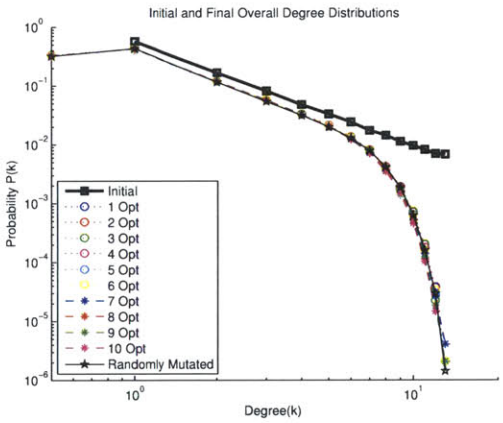


Figure 5-6: Overall degree distributions for the base model. Plot is the probability $P(k)$ of having a node with degree k. This distribution is plotted for the initial network (black), Randomly mutated network, and networks with 1 to 10 optimized nodes.

Apart from the overall degree distribution, one can look at various specific degree distributions. Since the network under consideration is a directed network with signed edges, one can look at the in- and out- degree distribution, and the degree distribution of activating (+) and inhibiting (-) nodes. Figure 5-7 shows these degree distributions for the base model. The initial distribution in this case contains half the links of the overall distributions given earlier, and since there may be nodes with only incoming (no outgoing) or only activating links, nodes of degree zero are possible in this case, even initially. Again, the relative difference between initial and final distri-

butions is greatest at high degree, and this metric does not help us tell the difference between the optimized and randomly mutated networks. Even though the growth function has to be maximized, and this would involve the value of $\underline{x}$ at specific nodes be positive, we do not see a significant bias in activating nodes vs. inhibiting nodes. This is potentially because of the large system size and the fact that inhibiting an inhibitor can result in activation. The specific degree distributions for the Boolean, coupled growth and correlated mutation models are shown in the supplement and also show the same feature: they are not good metrics for distinguishing between optimized and randomly mutated networks.

5.4.3 Degree - degree probability maps

In order to examine more closely the changes in degrees, we can look at the probability distributions of degree changes, i.e., if a node starts out with a certain degree, say k, what are the probabilities that it will end up with degrees of 0 to $k-1$? Figures 5-8A and 5-8B show these distributions for different initial degrees for both the optimized and randomly mutated cases. To be able to pick out the differences between the optimized and randomly mutated networks, we show the difference between the two: optimized - random, in Figure 5-8C. This figure shows which kinds of mutations are more likely in the optimized networks (red) than in the randomly mutated network. The differences are small and unclear. To parse out the finite size edge effects and the fact that many mutations might take place in nodes that are not between the driver and effector nodes, we can look at the nodes along the shortest path from the driver mutation to the effector node (in the case of multiple effector nodes, the union of all such shortest paths is considered). The low- degree nodes along the shortest path from driver to effector are most likely to be conserved, and this effect decreases with the degree of the node. This is potentially due to the fact that removing links from low-degree nodes would completely disconnect the network.

Results for the other models are presented in the supplementary material. For the case of the coupled growth model, apart from the conservation of low degree nodes,

there seems to be a trend in the evolution of the higher degree nodes: a small decrease in their degree, meaning that they get slightly mutated. This is reasonable because it is another way of avoiding the complete disconnection of the network. The Boolean model shows similar trends to the base model, but clearerly. In the correlated mutation model, single mutations can have drastic effects on network structure, and even though the growth function is not coupled in this case, the effects of the constraints seem to be similar to that of the coupled growth model. However, one does not see a preservation of the pattern for the low degree nodes along the shortest path.

5.4.4 Loops and paths

Another basic network structure characteristic is the number of feedback loops in the network and the number of paths (similar to feed-forward loops). Feedback loops are typically used to regulate signaling, and so one would expect feedback loops to play a role in achieving optimal growth in this framework. The sign of the feedback loops are measured by the overall polarity of the loop, i.e. by multiplying the signs of the various edges comprising the loop. These networks are small enough for a brute-force method of counting all loops in the system to be viable. The probability distributions of number of feedback loops in the system for the initial, optimized and randomly mutated cases are shown in Figure 5-9A-C. Figure 5-9A shows all feedback loops, 5-9B shows positive feedback loops and 9C shows negative feedback loops. We see that the number of positive feedback loops in the optimized networks is slightly higher, but the number of negative feedback loops is very similar to randomly mutated networks. This slight prevalence of positive feedback loops is potentially due to needing to control the value of signal at the optimized nodes. Since the feedback loops could, irrespective of their sign themselves positively or negatively impact the effector nodes, the difference between positive and negative feedback is mitigated and not strongly seen. Francois and Siggia[127][128] find a similar result of feedback loops influencing complexity of evolved networks in the case of a model of development.

Similarly, the number of paths from driver mutation to all effector nodes is shown in figure 5-10. Once again, we calculate the sign of these paths as the product of the signs of the individual nodes comprising the path. Figure 5-10A is for all paths, 5-10B for paths with positive polarity, and 5-10C for paths with negative polarity. Figure 5-10B shows that there are many more paths with positive polarity in the optimized networks than by random chance; this is because there is a positive value of the signal at the effector node that gives maximum growth.

Similar results of path and loop evolution for the other models are shown in the supplementary material. For the case of the coupled growth model, growth seems to be maximized by decreasing the number of negative feedbacks rather than increasing the number of positive feedbacks and a preservation of positive polarity paths. The Boolean model seems to preserve loops and paths in all cases.

5.4.5 Inhibition and Escape

To evaluate where it is optimal to inhibit these networks, we look at the values of growth of these trajectories at three points in time: (1) After the cancer has evolved to reach its "optimal" growth (2) When an inhibitor is added to the fully evolved cancer (3) After the inhibited cancer is allowed to make mutations in order to escape from the pressure imposed by the drug. We use the metric of the total degree to characterize each node: this can be the initial degree before the cancer (which is presumably known for each node in a real network), the degree at inhibition and the degree after escape mutations occur. Figure 5-11 shows results for inhibition and escape; the top row depicts the change in growth upon addition of the inhibitory drug. All values in these two plots are negative, referring to the decrease in growth upon addition of the inhibitor. Figures 5-11 A and B show that the maximum decrease in growth upon inhibition happens for high-degree nodes in general; nodes with the highest degree after cancer evolution are the best targets for inhibitors. Since these may not be easy to measure practically, nodes with the highest degree at the beginning represent a

suitable proxy. To measure how much these cancers can mutate around the effect of these inhibitors and regain their growth, we plot the difference in growth as a result of making escape mutations. Figures 5-11 C and D plot the difference in growth upon making escape mutations; these are all positive showing that making escape mutations increases the growth of these systems. The maximum increase in growth occurs for nodes that have a high degree at the time of inhibition(after the cancer has evolved).The magnitude of decrease of growth after inhibition is roughly equal to the increase after escape mutations, suggesting that these systems are recovering most or all of their ability to grow despite addition of the drug.

To look in more detail at whether adding an inhibitor is beneficial at all, we plot the difference in growth between the cancer after escape mutations and before inhibition (difference of the previous two figures). Figures 5-12 A and B shows a plot of this difference in growth as a function of node degrees; the values of this difference are slightly smaller than zero in most cases suggesting that addition of the inhibitor has a beneficial but small effect in general. The best effect, once again, is for nodes with a high degree at inhibition (Figure 5-12A); it is difficult to make statements about the degree after escape due to the small number of data points and large amount of noise (Figure 5-12B).

5.5 Discussion

We have built a simple model of the signaling consequences of mutations in cancers, in which a scale-free network is optimized(natural selection) for a function (growth) of signaling in the network by means of changes in topology (mutations). We asked the question: are there patterns of mutations in cancers that are derived as a result of this constraint? We attempt to quantify the changes in topology of signaling networks due to mutation in terms of the basic metrics of the network (degrees, loops, paths) that differ between optimized and randomly mutated networks. This basic

computational framework can also be used with other models of networks and interactions: for example, using a chemical kinetics model of reactions in the network where mutations represent changes in rate constants.

The key results from this model are (1) degree distributions are not a good metric for measuring mutation likelihoods (2) the more correlated the system, the clearer the trend in mutations in the system (3) loops and paths are modified differently in optimized networks as compared to randomly mutated networks (4) low-degree nodes along the shortest path from driver mutation to optimized node are more likely to be conserved and (5) the likelihood of seeing multiple optima, which is hypothesized to be the analog of variants of the same cancer, increases with the number of loops and paths in the networks, and is a stronger function of feedback loops. This framework has also been extended to test the statistical efficacy of drugs for different nodes and cancers to generate escape mutations to bypass the drug. We find that inhibiting high-degree nodes provides the greatest decrease in growth, but is also the easiest to escape from.

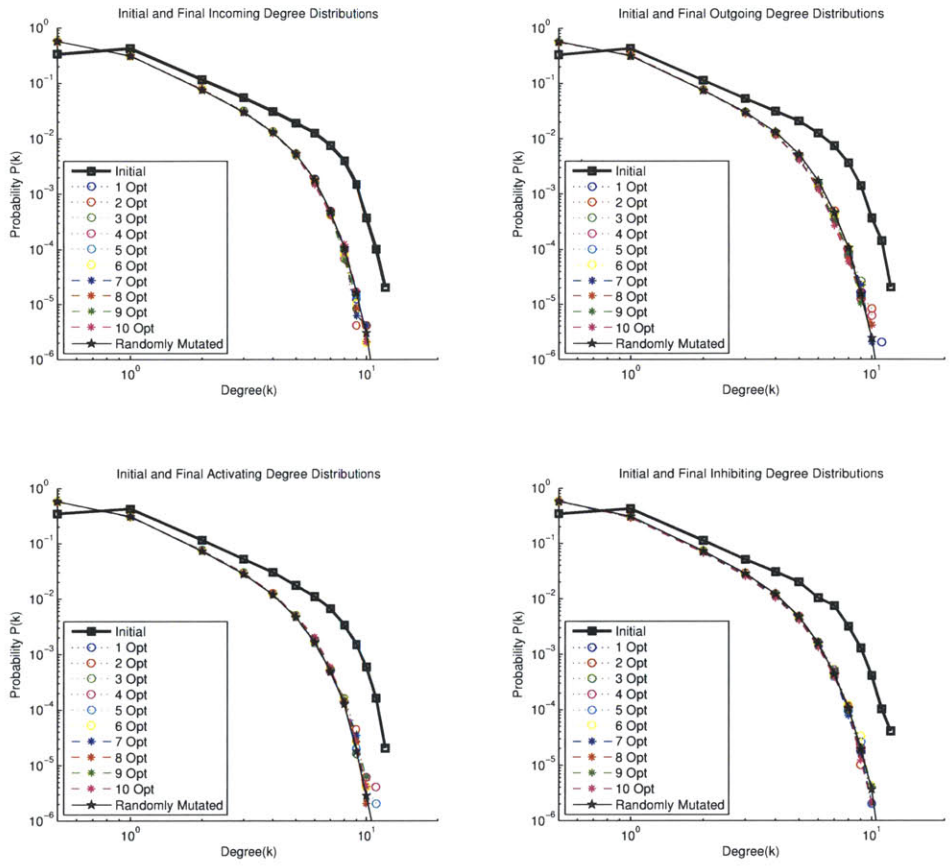


Figure 5-7: Specific degree distributions for the base model. Plot is the probability $P(k)$ of having a node with degree k. These can be the degree distributions of (A) edges going in to a given node; (B) edges going out of a given node (C) Activating edges; (D) Inhibitory edges. Note that the distributions for optimized and randomly mutated networks coincide.

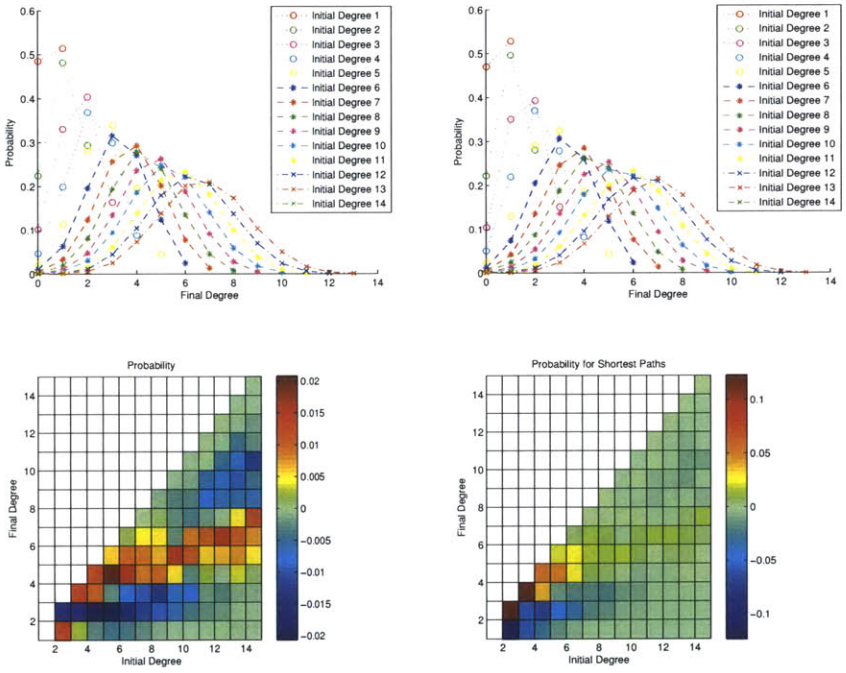


Figure 5-8: Probability distributions of final degree for different initial degrees for (A) Optimized networks (B) Randomly mutated networks. Figure (C) is a plot of the difference in probability, optimized - random. Red shading represents cases which are more probable in optimized networks than by random; blue ones represent less probable. (D) is an analog of (C), but only nodes along the shortest path from initial mutation to optimized node are considered: we see that low-degree nodes along the shortest path tend to be conserved. This trend is seen across different model types.

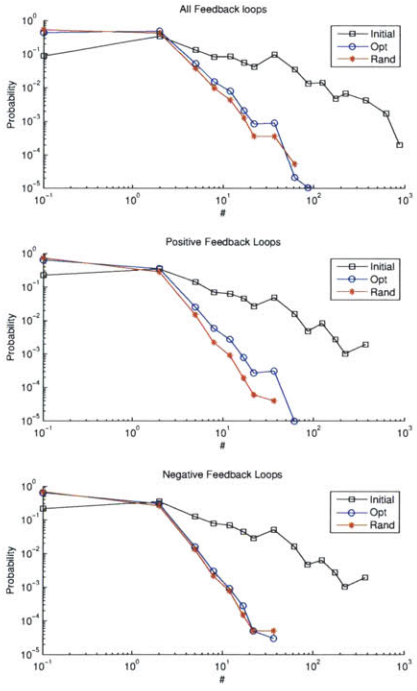


Figure 5-9: Plots of probability distributions of number of loops present in the network for (A) all (B) positive and (C) negative feedback loops. Plots are made for the initial (black), optimized (blue) and randomly mutated (red) networks.

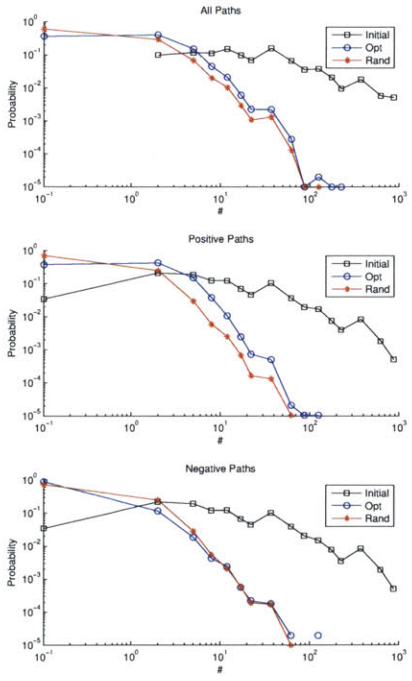


Figure 5-10: Plots of probability distributions of number of paths (without any cycles) present in the network for (A) all (B) positive and (C) negative polarity paths. Plots are made for the initial (black), optimized (blue) and randomly mutated (red) networks. The polarity of a path is defined as the product of the signs of the various edges that make up the path.

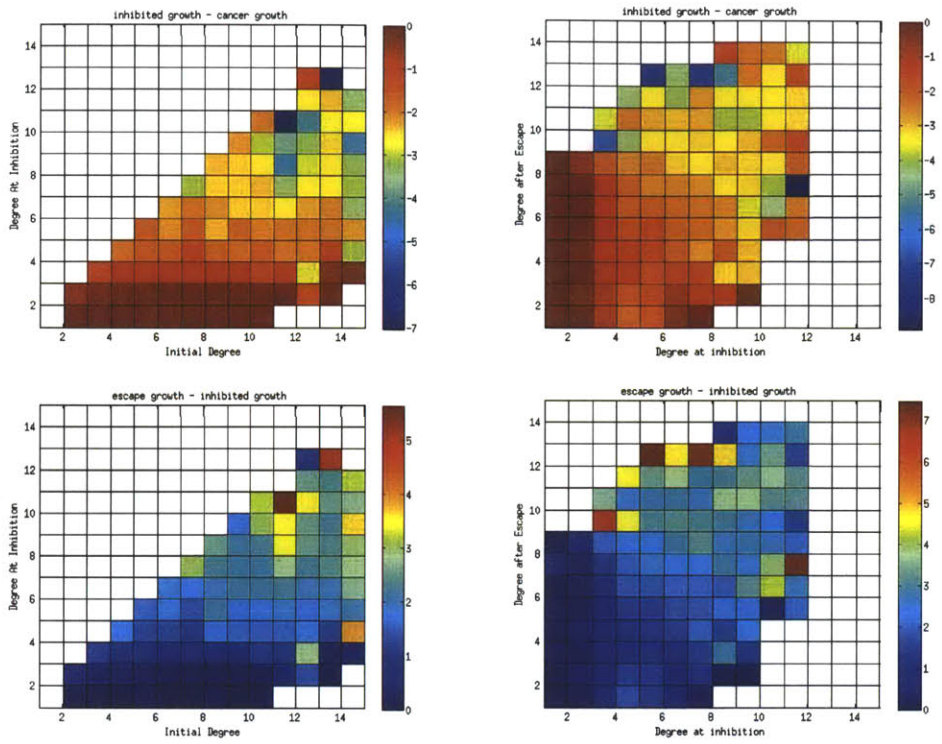


Figure 5-11: Change in growth of model systems upon inhibition and escape. Plots consist of change in growth as a function of degree of the node inhibited. (A) and (B) Change in growth upon inhibition. (C) and (D) Change in growth as a result of escape mutations. The largest effects are for high degree nodes.

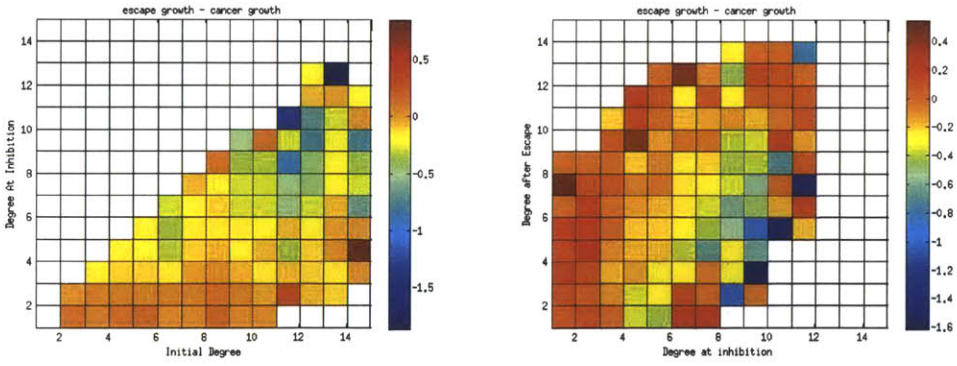


Figure 5-12: Overall effectiveness of adding an inhibitor. Plots consist of change in growth, the difference between growth after escape and growth before inhibition, as a function of degree of the node inhibited. Inhibitors in this framework are only marginally effective.

Chapter 6

Threshold ligands for different cell types

6.1 Introduction

Developing thymocytes pass through a series of distinct stages, marked by changes in expression levels of the TCR and other cell-surface proteins. The levels of cell surface markers CD3 (TCR) and coreceptors CD4 and CD8 are typically used to identify which stage a cell is in, and differentiate between different T-Cell lineages[129][130][131]. After VDJ gene rearrangement, $\alpha\beta$ T-Cells cells start off as "double positive" $CD4^+CD8^+$[132]. These undergo positive selection, in which TCRs which either do not bind to MHC at least weakly die due to neglect, eliminating those T-Cells which would not be functional. These then undergo lineage commitment, in which they become either CD4($CD4^+CD8^-$) or CD8($CD4^-CD8^+$) "single-positive" T-Cells. Then, single positive cells undergo a process called negative selection in which those that bind too strongly to self-pMHC apoptose, so that immune responses towards self can be avoided in the periphery. The processes of selection are driven by the affinity of TCR to pMHC: some affinity is required to be positively selected but too strong an affinity leads to negative selection.

It has been proposed that a T-Cell detects ligand affinity by measuring how long peptide-MHC complexes remain bound to a TCR (called the dwell time); this is

called the kinetic proofreading model of T-Cell activation[133][134]. Since the pMHC-TCR complex can also bind to coreceptor-Lck, there is in effect a larger complex that is stabilized by multiple interactions[135]. This prolongs the dwell time and results in an apparent increase in the affinity of the TCR-pMHC interaction[136][137]. There is a narrow apparent affinity threshold between the possible outcomes of thymic selection[138]. This threshold has been measured experimentally for different MHC Class I restricted TCRs, and was found to be a K_D of around 6 μM (half-life of around 2 seconds) in all cases[139]. Similar results were found for other T-Cell types, with different thresholds for each cell type[140]. The question still remains as to what is the biophysical basis for this constant affinity threshold for each cell type.

6.2 Experiments

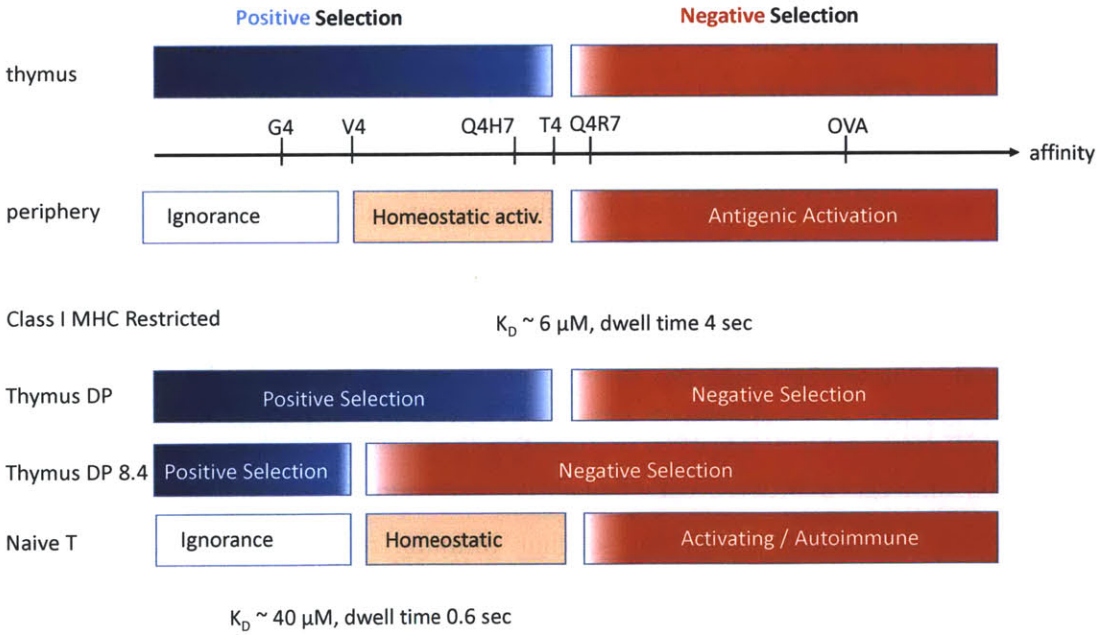


Figure 6-1: An overview of the experimental results for CD8s. A series of peptides were used to measure if different TCRs would activate, and the threshold K_Ds and dwell times for each T-Cell type were found to be constant.

Cell Type	**$K_D(\mu M)$**	**$t_{1/2}$(ms)**	**Dwell Time (ms)**
MHC I restricted DP Thymocyte	6	2000	4000
MHC I restricted Naive Peripheral T Cell	40	300	600
MHC II restricted DP Thymocyte	300	40	80
MHC II restricted Naive Peripheral T Cell	300	40	80

Table 6.1: Experimental results: The "universal" threshold for activation, for different cell types. Equilibrium constants for TCR-pMHC binding K_Ds and dwell times.

The typical diameter of a T-Cell is about $d = 5\mu m$, which gives a surface area (assuming a spherical cell) of about $\pi d^2 = 78.54\mu m^2$. Total numbers for various molecules on the surface of the T-Cell is given in Table 6.2.

Molecule Type	DP Thymocyte	Peripheral CD8	Peripheral CD4	Chimera CD8.4
Coreceptor CD8	273000	320000		256000
Coreceptor CD4	112500		218000	91000
TCR	2500	53000	91000	8000
% of CD8 with Lck	1.4%	15.1%		9.8%
% of CD4 with Lck	15.9%		24.1%	19.4%

Table 6.2: Concentrations of various molecules on the surface of T Cells

Simulations are performed on a $1\mu m \times 1\mu m$ patch of membrane; the diffusion constant is assumed to be $D = 0.08\mu m^2 s^{-1}$ for all molecules. A "lattice spacing" l of $0.01\mu m$ is assumed, and molecules can interact if they are closer than this distance (This is needed to convert second order rate constants and diffusion).

Kinetic constant	Symbol	Value	Reference
Unbinding of MHC and CD8 without Lck	k_u	$20s^{-1}$	[135]
Binding of MHC and coreceptor	k_b	$1000s^{-1}$	[135]
Diffusion	k_{diff}	$800s^{-1}$	See text
pMHC-TCR binding rate	k_{on}	$150s^{-1}$	[135]
pMHC-TCR unbinding rate	k_{off}	varied	See text
Phosphorylation rate of TCR by Lck	k_p	$0.05s^{-1}$	[135]; see text

Table 6.3: Rate constants for the T-Cell early activation system

6.3 Complete Models: Simulations to analyze qualitative behavior

The model consists of the four main molecules involved in the activation of T Cells: the coreceptor CD8 or CD4, Lck, TCR and pMHC. The coreceptor and TCR are present on the surface of the T-cell, pMHC on the surface of the APC and Lck is a cytosolic protein within the T Cell. However, to simplify the model, we assume that they are all present on a single membrane. A square patch of membrane is looked at, sized 1 μm by 1 μm, and this is divided into a 100×100 grid in order to provide spatial resolution.

The following pairs of molecules can interact with each other: TCR and pMHC, MHC and coreceptor, coreceptor and Lck, and Lck and TCR. Each of these binding/unbinding reactions is assumed to follow mass action kinetics and can happen only when the molecules are in the same grid point. The following species are assumed to diffuse on the membrane: Coreceptor, Lck, TCR, pMHC, Coreceptor-Lck pair and TCR-Lck pair; larger complexes do not diffuse. All species diffuse at the same rate and the diffusion constant has been varied (and this is one of the parameters of interest in understanding the behavior of the system). TCR has two phosphorylation sites and the simulation readout is fully phosphorylated TCR. TCR bound to Lck can get phosphorylated, whether or not it is bound to pMHC; the effective stabilization of the TCR-pMHC-coreceptor-Lck complex is what makes the result dependent on the other rate constants of the system. Free TCR can get dephosphorylated by first-order kinetics. The simulations are run for a long time (10^5 seconds) and we analyze the long time, steady-state behavior.

In different cell lines, the fraction of coreceptor that is bound to Lck varies significantly. In order to vary this fraction in our system, we change the off rate between coreceptor and Lck. The relationship between fraction of coreceptor that has Lck and off rate is quite stable and is independent of other parameters in the system, like the

off rate between TCR and pMHC (Figure 6-2).

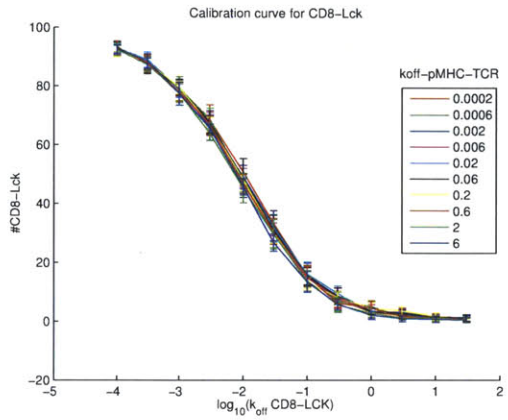


Figure 6-2: Fraction of CD8 bound to Lck as a function of k_{off} between CD8 and Lck. Different curves are for different values of k_{off} between MHC and TCR

6.4 Results of the complete model

6.4.1 Threshold Ligand Strengths

We first examine the effect of changing the diffusion constant in the system. Diffusion could have multiple opposing effects in this system: increasing the diffusion constant would bring molecules together much quicker, but it would also mean that they are close to each other for a shorter period of time and hence less likely to react. A plot of the amount of fully phosphorylated TCR as a function of TCR-pMHC affinity and diffusion (Figure 6-3) shows two regimes: for low values of diffusion constant, the amount of fully phosphorylated TCR increases with diffusion, whereas for high values of diffusion, the reverse is true. The variation with TCR-pMHC binding strength is easy to understand: stronger TCR-pMHC binding (lower k_{off} between TCR and pMHC) results in stronger activation.

We speculate that in the low diffusion coefficient regime, the bottleneck is the ability of different molecules to find each other within the relevant timescale. Hence, the activation level increases with the diffusion rate (at a fixed pMHC affinity). In

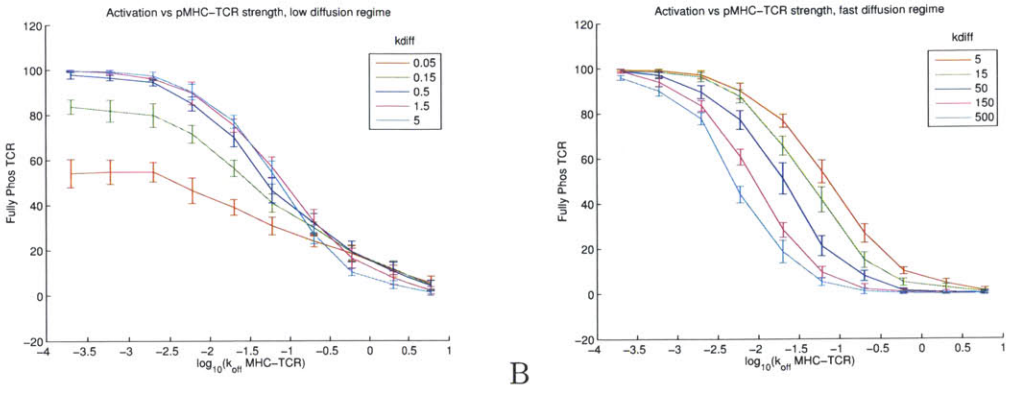


Figure 6-3: Amount of fully phosphorylated TCR as a function of TCR-pMHC strength and diffusion. The X-axis is the $K-off$ of TCR-pMHC binding, so weaker ligands are to the right. The location of the inflection point in these curves is a measure of threshold ligand strength needed to activate these systems. This threshold is highest (weakest ligands can activate) at intermediate diffusions of around 5 (SSC units)

the high diffusion coefficient regime, species diffuse fast enough to find each other within the relevant timescale. However, a very high diffusion coefficient means that species are more likely to diffuse away when they are each other, and the probability of a binding event is lower, which leads to lower a probability of reaction. This seems to be the regime in which the biological system is operating.

We can then plot the number of fully phosphorylated TCRs as a function of k_{off} TCR-pMHC and the fraction of coreceptor with Lck in Figure 6-4.

In order to understand the composite effect of the three independent parameters of the system (TCR-pMHC strength, diffusion and fraction of coreceptor that is bound to Lck) we can choose an arbitrary level of the amount of fully phosphorylated TCR as the "threshold" at which the T-cell activates (modifying this value does not change the qualitative results). In Figure 6-4, three possibilities for the required threshold level of phosphorylation have been chosen: TCRpp = 25, 50 and 80. In each of these three cases we calculate the maximum k_{off} between TCR and pMHC that enables the cell to reach that level of activation. For example:

- At threshold TCRpp of 25: a k_{off} (MHC-TCR) of 0.006 is required to activate at less than 5% Lck; k_{off} of 0.02 can activate at 15% Lck; and k_{off} of 0.06 can

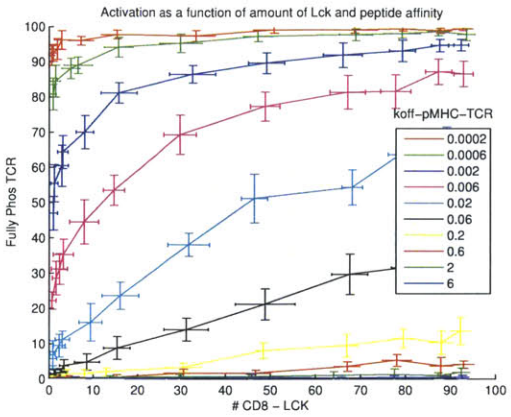


Figure 6-4: Amount of fully phosphorylated TCR as a function of TCR-pMHC strength and fraction of coreceptor that is bound to Lck. The number of fully phosphorylated TCR (a measure of activation strength) increases with the amount of Lck bound to CD8 and decreases with greater k_{off} of the TCR-pMHC interaction.

activate at 60% Lck.

- At threshold TCRpp of 50: a k_{off} (MHC-TCR) of 0.002 is required to activate at less than 5% Lck; k_{off} of 0.006 can activate at 10% Lck; and k_{off} of 0.02 can activate at 50% Lck.

- At threshold TCRpp of 80: a k_{off} (MHC-TCR) of 0.0006 is required to activate at less than 5% Lck; k_{off} of 0.002 can activate at 15% Lck; and k_{off} of 0.006 can activate at 65% Lck.

Since only the qualitative behavior is of interest, it does not matter what the threshold level of TCRpp is. For Figure 6-5, a threshold TCRpp of 50 has been chosen. We then plot of the threshold k_{off} (defined here as the k_{off} between TCR and pMHC where amount of phosphorylated TCR = 50) as a function of both the diffusion rate and the amount of CD8 bound to Lck:

With increasing amount of CD8 bound to Lck, the threshold k_{off} increases for all diffusion rates, i.e. weaker ligands can activate. There is a maximum in the curves with respect to the diffusion rate: at too low diffusion, molecules do not have time to find each other, and at too high diffusion, molecules move past each other before they are able to react. This qualitative behavior is independent of the fraction of

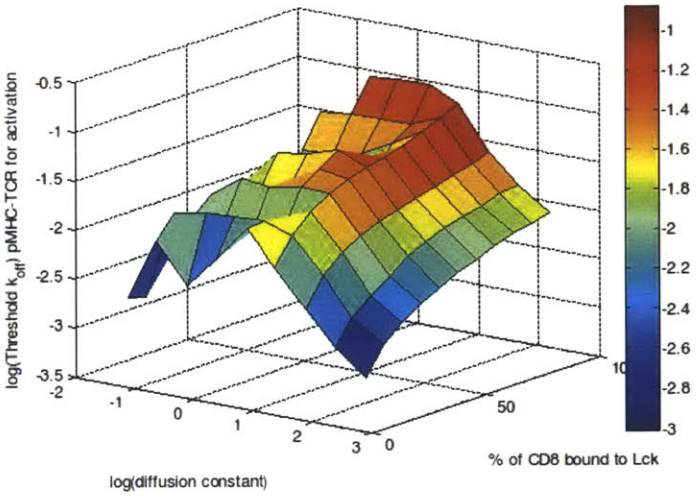


Figure 6-5: Maximum k_{off} between pMHC and TCR for which the cell can activate (50 TCRs phosphorylated) as a function of diffusion and fraction of coreceptor bound to Lck. The threshold k_{off} increases (weaker ligands are sufficient) with increasing fractions of CD8 bound to Lck, but there is a maximum effectiveness for intermediate values of diffusion.

CD8 bound to Lck.

6.4.2 Activation Timescale

The other question of interest is how long the TCR needs to be bound to the APC for the cell to activate. In the limit that all reactions are fast and the rate-limiting step is the time taken for a TCR-pMHC to find a coreceptor with Lck, we can calculate a diffusion-limited rate. For the case where concentration of TCR = Concentration of coreceptor/MHC = $c \approx 100$ molecules/μm^2, the mean distance between molecules r is obtained from $\pi r^2 = \frac{1}{c}$, so $r = \frac{1}{\sqrt{\pi c}} = \frac{1}{17.7} = 0.056\mu m$. For a diffusion coefficient of $D = 0.01\mu m^2/s$, the diffusion time to find another molecule is $\tau \approx r^2/D = \frac{(0.056\mu m)^2}{0.01\mu m^2/s} = 0.31$ sec. This value will vary with the diffusion coefficient as $\frac{1}{D}$ (the solid line in Figure 6-6).

We calculate from simulations the fist-passage time distribution (FPTD) to phosphorylating n TCRs, where $n = 2, 5, 10, 25$; and vary the parameters.

To quantify the effects of various parameters on the first passage time distribution, we plot the mean of those distributions as a function of various parameters. Figure 6-6A shows the MFPT for 2 TCRs to be fully phosphorylated. The solid line is the diffusion limit scaling, $\tau = \frac{b^2}{d}$. Scaling for the limit of the real data is different, suggesting that a process other than diffusion is limiting the rate of phosphorylation of TCR.

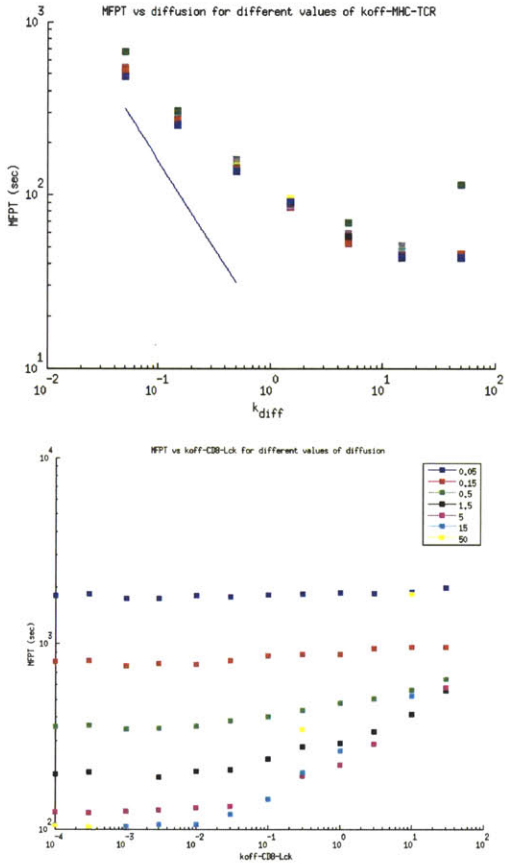


Figure 6-6: (A) Mean First-passage time for phosphorylation of 2 TCRs as a function of diffusion, for different pMHC-TCR binding strengths. The solid line is the diffusion limit, $\tau = l^2/D$. (B) MFPT as a function of fraction of coreceptor bound to Lck.

Similar plots for different thresholds for MFPT of activated TCR: TCRpp = 5, 10 (top row) and 25, 50 (bottom row) are plotted in Figure 6-7. There is a threshold strength of ligand above which MFPT is independent of k_{off}, but MFPT increases

with weaker ligands. We also see that there is an intermediate value of diffusion at which the system activates fastest, and this is den-pendant on the strength of the ligand. Figure 6-6B suggests that the MFPT for activation is not a strong function of the fraction of CD8 that is bound to Lck.

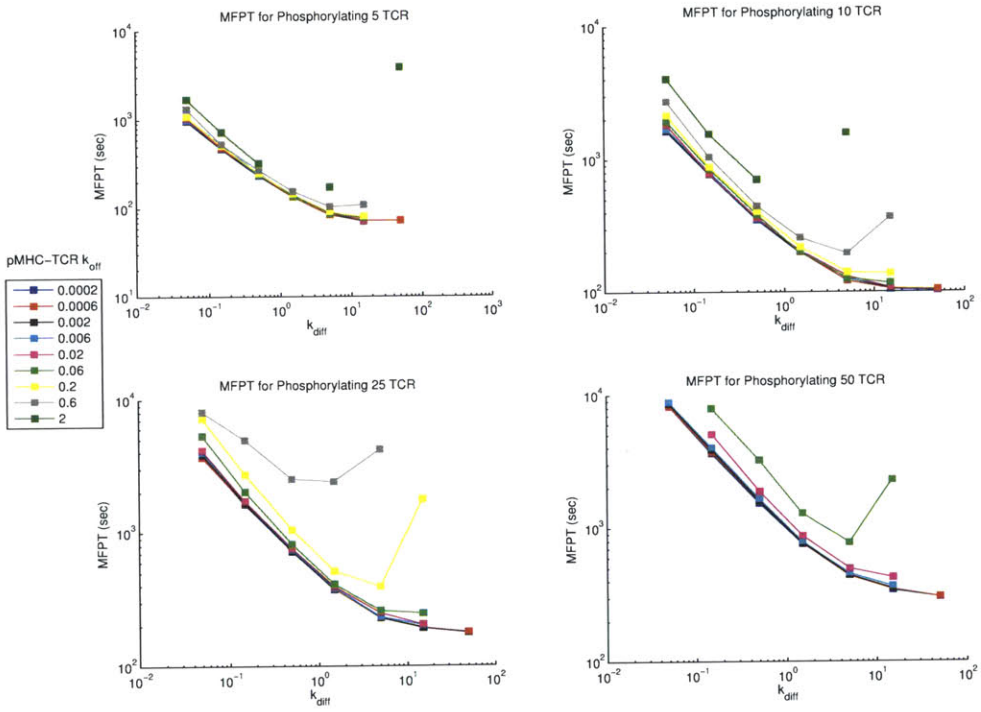


Figure 6-7: The Mean First-Passage Time (MFPT) for activation of 5, 10, 25 and 50 TCRs as a function of pMHC-TCR k_{off} and diffusion. There is a minimum in the MFPT for an intermediate value of diffusion, suggesting that moderate diffusion rates are most efficient. The location of this minimum varies with ligand strength and number of TCRs phosphorylated.

6.5 Markov Chain Model

We illustrate the analytical model with the example of CD8-restricted MHC Class I with peptide activating the DP Thymocyte. On the surface of the T Cell there are 2500 TCR and 273000 CD8 (of which on 1.4% has Lck). For a T-Cell of diameter

$d = 5\mu m$, this works out to $c = \frac{273000}{78.54} = 3476$ coreceptor molecules (of which 49 have Lck) and 32 TCRs on a $1\mu m \times 1\mu m$ patch of membrane. Assuming a lattice spacing of $0.01\mu m$, the diffusion constant of $0.08\mu m^2 s^{01}$ works out to a hopping rate of $k_{diff} = \frac{D}{l^2} = \frac{0.08}{0.01^2} = 800s^{-1}$. For the initial model, we assume there is only one pMHC-TCR complex in this patch, and this does not dissociate (the limit of a very strong ligand) so we can analytically determine a formula for the time required to activate. The average distance between the pMHC-TCR and a coreceptor is $r = \frac{1}{\sqrt{c \times \pi}} = \frac{1}{\sqrt{3475\pi}} = 0.0096\mu m$, which yields an average time taken to find the target $t = \frac{r^2}{D} = \frac{0.0096^2}{0.08} = 0.00115s$ or a corresponding "finding rate" of $k_f = \frac{1}{t} = 868s^{-1}$.
A physical description of the process involved in a pMHC-TCR being activated is as follows: the pMHC-TCR diffuses around on the surface of the T Cell, where it can encounter coreceptors either with or without Lck. It can bind to and unbind from either, but only a coreceptor bearing Lck can phosphorylate it. This process can be modeled as a Markov Chain as follows: A free pMHC-TCR (state "TM"), diffuses around on the surface of a T Cell and can encounter either coreceptor without (C) or with Lck (LC). There are rates at which these processes happen, described by rate constants k_{f0} and k_{f1} respectively. If pMHC-TCR and coreceptor are close by (in the same lattice site), they can either bind or diffuse apart; these processes happen with rate constants k_b and k_{diff} respectively. Each of the bound states can unbind with a rate constant of k_u, and the pMHC-TCR-coreceptor-Lck complex can get phosphorylated at rate k_p. This is visualized in Figure 6-8, and the states are listed in Table 6.4. The rate constants for transitions between these states are given in Table 6.5.

$$TM : C \underset{k_{b0}}{\overset{k_{u0}}{\rightleftarrows}} TM + C \underset{k_{f0}}{\overset{k_{diff}}{\rightleftarrows}} TM \underset{k_{diff}}{\overset{k_{f1}}{\rightleftarrows}} TM + LC \underset{k_{u1}}{\overset{k_{b1}}{\rightleftarrows}} TM : LC \xrightarrow{k_p} T_p$$

Figure 6-8: The full Markov Chain model for activation of T Cells. A description of states and parameters are given in Tables 6.4 and 6.5

The initial State of Markov Chain is free TCR-pMHC (state "TM"). The activated state(T_p) is an absorbing state. We can numerically evolve the set of ODEs that

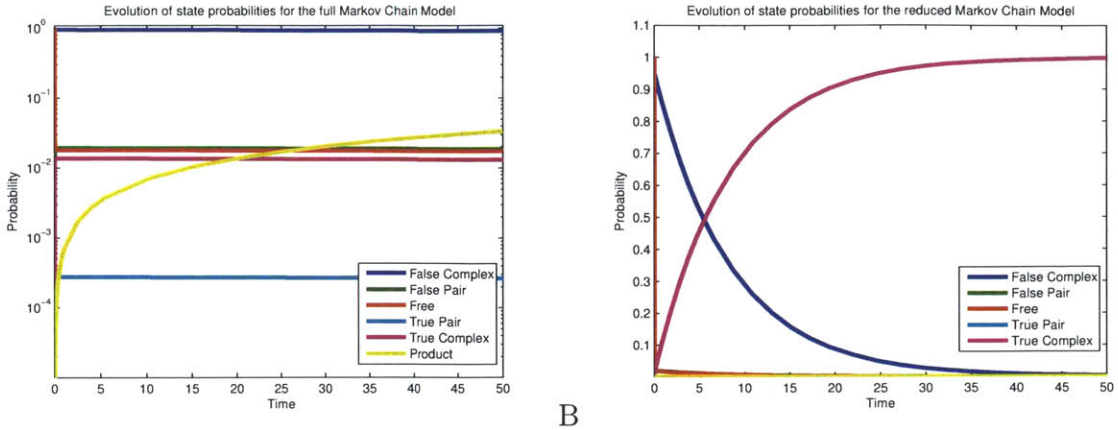
State	Description
TM:C	TCR-pMHC bound to coreceptor without Lck. "False Complex"
TM + C	TCR-pMHC next to coreceptor without Lck. "False Pair"
TM	Free TCR-pMHC. "Free"
TM + LC	TCR-pMHC next to coreceptor with Lck. "True Pair"
TM:LC	TCR-pMHC bound to coreceptor with Lck. "True Complex"
Tp	Product

Table 6.4: The states of the Markov Chain Model

Description	Symbol	Value
Unbinding of MHC-coreceptor without Lck	k_{u0}	k_u
Binding of MHC-coreceptor without Lck	k_{b0}	k_b
Diffusion	k_{diff}	k_{diff}
Rate of pMHC-TCR finding CD8 without Lck	k_{f0}	$k_f \times \frac{3475-49}{3475} = 0.986k_f$
Rate of pMHC-TCR finding CD8 with Lck	k_{f0}	$0.014k_f$
Unbinding of MHC-coreceptor without Lck	k_{u1}	see text
Binding of MHC-coreceptor without Lck	k_{b0}	k_b
Phosphorylation	k_p	see text

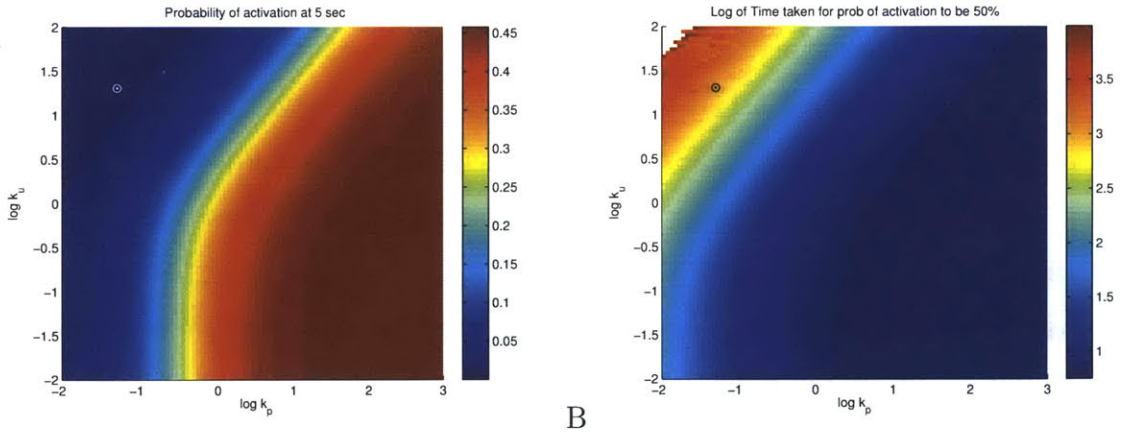
Table 6.5: The transition rate parameters of the Markov Chain Model

describe the evolution of probabilities of the various states with time; the result is plotted in Figure 6-9A. By varying the two unknown parameters (stability of the full complex, described here by k_{u1} and rate of phosphorylation of TCR k_p), we can plot the probability of activation in 5 seconds (Figures 6-10A and 6-10B respectively). These suggest that the maximum probability of activation in five seconds is around 0.46 and the fastest time for 50% activation is about 6.3 seconds ($10^{0.8}$). The result for the original parameters (CD8 without Lck having the same values as CD8 with Lck) is represented by the black dot near the top left. The fact that the probability of activation never goes above 0.5 in five seconds, even at the limit $k_p \to \infty$ and $k_{u1} \to 0$, suggests that the rate-limiting step is the formation of the TM:LC complex. To look at this limit, we can re-run the simulations with $k_p = 0$ and $k_{u1} = 0$, and plot the probability of the various states as a function of time. (Now the "True Complex" effectively becomes an absorbing state, instead of the product, and this simplified Markov Chain model is shown in Figure 6-11).



A B

Figure 6-9: A: Time courses for various states of the full Markov Chain model. Note that the probabilities of many of the states are relatively constant with time. Ordinate is log-scaled. B: Time courses for various states of the reduced model.



A B

Figure 6-10: (A) Probability of activation of the full Markov Chain model after 5 seconds, as a function of two parameters (TCR phosphorylation rate k_p and stability of the full complex $k_u 1$). The result for the original parameter set is given by the dot in the upper right corner of this plot. (B) Time taken for the probability of activation of the full Markov Chain model to be 0.5. Again, the original parameter set is given by the dot in the upper right corner.

6.5.1 Approximate Analytical Solution of the Markov Chain Model

The full set of ODEs describing the system is given in Equation 6.1.

$$TM:C \underset{k_{b0}}{\overset{k_{u0}}{\rightleftharpoons}} TM + C \underset{k_{f0}}{\overset{k_{diff}}{\rightleftharpoons}} TM \underset{k_{diff}}{\overset{k_{f1}}{\rightleftharpoons}} TM + LC \xrightarrow{k_{b1}} TM:LC$$

Figure 6-11: The reduced Markov Chain model for activation of T Cells. A description of states and parameters are given in Tables 6.4 and 6.5.

$$\begin{aligned}
\frac{dP_{TM:C}}{dt} &= -k_u P_{TM:C} + k_b P_{TM+C} \\
\frac{dP_{TM+C}}{dt} &= k_u P_{TM:C} - (k_b + k_d) P_{TM+C} + k_{f0} P_{TM} \\
\frac{dP_{TM}}{dt} &= k_d (P_{TM+C} + P_{TM+LC}) - (k_{f0} + k_{f1}) P_{TM} \\
\frac{dP_{TM+LC}}{dt} &= k_{f1} P_{TM} - (k_d + k_b) P_{TM+LC} \\
\frac{dP_{TM:LC}}{dt} &= k_b P_{TM+LC}
\end{aligned} \tag{6.1}$$

In Figure 6-9, we see that the probabilities of the False Pair and True Pair states are low and approximately constant through the simulation. We make a pseudo-steady state assumption[141] to try to simplify and analytically solve - set the 2nd and 4th equation to zero (PSSA) which gives us the ODEs in Equation 6.2. This approximation essentially means that the system spends most of its time in the states where pMHC-TCR is bound to CD8 without Lck or in the "free" state; the states where pMHC-TCR is next to a coreceptor are short-lived and the molecules very quickly either bind or diffuse apart.

$$\begin{aligned}
\frac{dP_{TM:C}}{dt} &= \frac{k_b k_{f0} P_{TM} - k_u k_d P_{TM:C}}{k_b + k_d} \\
\frac{dP_{TM}}{dt} &= \frac{k_d k_u P_{TM:C} - (k_{f0} + k_{f1}) k_b P_{TM}}{k_b + k_d} \\
\frac{dP_{TM:LC}}{dt} &= \frac{k_b k_{f1} P_{TM}}{k_b + k_d}
\end{aligned} \tag{6.2}$$

The initial conditions are $P_{TM:C}(0) = 0, P_{TM}(0) = 1, P_{TM:LC}(0) = 0$. This set of equations can be solved analytically but results in complicated expressions. This analytical simplified solution seems to be close to the full Markov Chain solution (Figure 6-12A).

As the peptide-MHC seems to extremely quickly find a coreceptor without Lck and the probability of being in the free state is also quite low, we can assume that

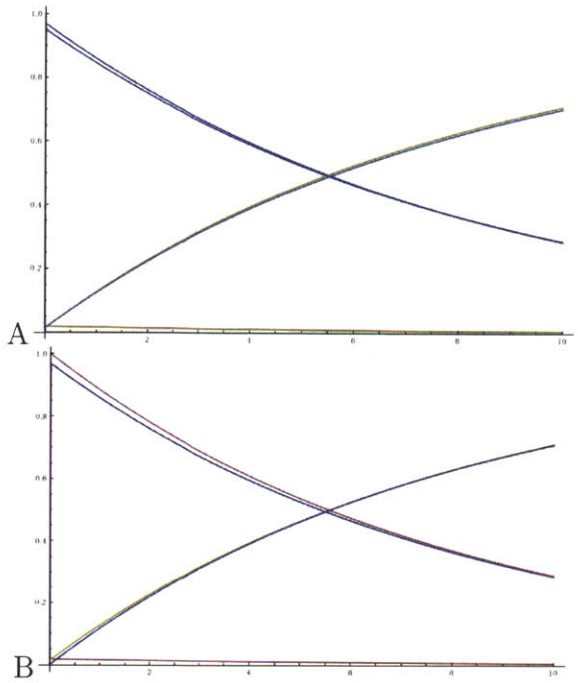


Figure 6-12: Comparison of solutions of the Markov Chain model. Comparison of (A) the approximate analytical solution (P_{TM+C} = constant, P_{TM+LC} = constant) with the full numerical solution (B) the even-more approximate analytical solution (P_{TM+C} = constant, P_{TM} = constant, P_{TM+LC} = constant) with the full numerical solution. The approximations are reasonably good.

the initial state of the system is state $P_{TM:C}$ in the above model, and that PSSA can be applied to free TCR-pMHC also. Simplified equations are shown in Equation 6.3.

$$\begin{aligned}
\frac{dP_{TM:C}}{dt} &= -k_u P_{TM:C} + k_b P_{TM+C} \\
\frac{dP_{TM+C}}{dt} &= k_u P_{TM:C} - (k_b + k_d) P_{TM+C} + k_{f0} P_{TM} = 0 \\
\frac{dP_{TM}}{dt} &= k_d (P_{TM+C} + P_{TM+LC}) - (k_{f0} + k_{f1}) P_{TM} = 0 \\
\frac{dP_{TM+LC}}{dt} &= k_{f1} P_{TM} - (k_d + k_b) P_{TM+LC} = 0 \\
\frac{dP_{TM:LC}}{dt} &= k_b P_{TM+LC}
\end{aligned} \tag{6.3}$$

The solution of the middle three equations for P_{TM+C}, P_{TM} and P_{TM+LC} is shown

in Equation 6.4.

$$
\begin{aligned}
P_{TM} &= \frac{k_d k_u P_{TM:C}}{k_b(k_{f0}+k_{f1})} \\
P_{TM+LC} &= \frac{k_d k_{f1} k_u P_{TM:C}}{k_b(k_b+k_d)(k_{f0}+k_{f1})} \\
P_{TM+C} &= \frac{(k_b k_{f0}+k_d k_{f0}+k_b k_{f1}) k_u P_{TM:C}}{k_b(k_b+k_d)(k_{f0}+k_{f1})}
\end{aligned}
\tag{6.4}
$$

We use the expressions in Equation 6.4 to simplify Equation 6.3, and get Equation 6.5.

$$
\begin{aligned}
\frac{dP_{TM:C}}{dt} &= -k_u P_{TM:C} + \frac{(k_b k_{f0}+k_d k_{f0}+k_b k_{f1}) k_u P_{TM:C}}{(k_b+k_d)(k_{f0}+k_{f1})} = \frac{-k_d k_{f1} k_u P_{TM:C}}{(k_b+k_d)(k_{f0}+k_{f1})} \\
\frac{dP_{TM:LC}}{dt} &= \frac{k_d k_{f1} k_u P_{TM:C}}{(k_b+k_d)(k_{f0}+k_{f1})}
\end{aligned}
\tag{6.5}
$$

The initial conditions $P_{TM:C}(0) = 1, P_{TM:LC}(0) = 0$. Solutions to Equation 6.5 are $P_{TM:C}(t) = e^{-\lambda t}$, $P_{TM:LC}(t) = 1 - e^{-\lambda t}$, where $\lambda = \dfrac{k_d k_{f1} k_u}{(k_b+k_d)(k_{f0}+k_{f1})}$

A comparison of plots, this very approximate analytic solution with the full MC model, is shown in Figure 6-12B. We see that the approximation is quite good. For a solution of the form $P_{TM:LC}(t) = 1 - e^{-\lambda t}$, where $\lambda = \dfrac{k_d k_{f1} k_u}{(k_b+k_d)(k_{f0}+k_{f1})}$, the short-time behavior(small t) is $P_{TM:LC}(t) = \lambda t$. This short-time behavior gives us a way to understand the scaling of the response with parameters.

6.5.2 Scaling

We can use this analytic solution to see how the results scale with various physical parameters, which differ between the various cell types:

The rate constant in three dimensions is $k_2 = 10^5 M^{-1} s^{-1}$. To convert the 3D value of k_2 into the two-dimensional $k_{A+B\to C}$, we have to know a characteristic length h, corresponding to the confinement of the proteins to the 2D membrane. Usually we assume that membrane proteins can move in a direction perpendicular to the membrane within the distance of $h = 10\mathring{A} = 10^{-3}\mu m$. The lattice spacing l in μm,

so 1 μm^2 area has $\frac{1}{l^2}$ lattice chambers.

The conversion is given in Equation 6.6.

$$
\begin{aligned}
& k_2\left(\frac{L}{mol \times sec}\right) \text{ to } k_{A+B\to C}\left(\frac{chamber}{molecules \times sec}\right) \\
& k_2\left(\frac{L}{mol \times sec}\right) = k_2\left(\frac{10^{-3}m^3}{6 \times 10^{23} molecules \times sec}\right) = k_2\left(\frac{10^{-3} \times 10^{18}\mu m^3}{6 \times 10^{23} molecules \times sec}\right) \\
& = \frac{k_2}{6 \times 10^8}\frac{\mu m^3}{molec.sec} = \frac{k_2}{6 \times 10^8}\frac{1}{l^2}\frac{chambers}{molecules \times sec} \times 1\mu m \\
& = \frac{k_2}{6 \times 10^8}\frac{1}{l^2 \times h}\left(\frac{chamber}{molecules \times sec}\right)
\end{aligned}
\tag{6.6}
$$

The scaling of the diffusion constant is $k_{diff} = D\left(\frac{\mu m^2}{sec}\right) \times \frac{1}{l^2}\left(\frac{1}{\mu m^2}\right)$.

Scaling of other rates with concentrations of molecules is calculated as follows:

- C = number of coreceptor molecules on cell surface
- A = surface area of T-Cell
- $c = \frac{C}{A}$ = number of coreceptors on patch under consideration
- f = Fraction of coreceptors bound to Lck

We can calculate the rate of finding coreceptor using the average distance to the next coreceptor r. r is related to c as $\frac{1}{c} = \pi r^2$, so $k_f = \frac{D}{r^2} = \pi c D$
We can calculate the rate of finding coreceptor with and without Lck by multiplying this rate by the fraction of coreceptors with and without Lck: $k_{f1} = \pi f c D$ and $k_{f0} = \pi (1-f) c D$

Therefore, the scaling of rates in the Markov Chain works out to: $k_d = \frac{D}{l^2}$, $k_{f1} = \pi f c D$, $k_{f0} = \pi (1-f) c D$, $k_b = \frac{\gamma}{l^2}$. The rate of activation at short times λ therefore scales as shown in Equation 6.7.

$$
\lambda = \frac{\frac{D}{l^2} \cdot \pi f c D \cdot k_u}{\left(\frac{\gamma}{l^2} + \frac{D}{l^2}\right)(f + (1-f))\pi c D} = \frac{f D k_u}{\gamma + D} \tag{6.7}
$$

Equation 6.7 suggests that the initial rate of activation λ is independent of l and c and is $\propto \frac{D}{D+const}$ and is $\propto f$. The results for the reduced Markov Chain model (numerical solutions) when D, l, C and f are varied are plotted in Figures 6-13. The black dot in each figure represents the original parameter set, and we look at how good the PSSA results are around this point. The analytic solution suggests that probability of activation is independent of number of coreceptors C and lattice spacing l, which shows good agreement with the full MC solution (Figures 6-13A and D). The model suggests that probability of activation increases linearly with fraction of coreceptors with Lck f, like in Figure 6-13C (note that the X-axis is log-scaled; saturation at high f is because the activation probability is close to 1). The variation with diffusion is not as large (Figure 6-13B), corresponding to the $\frac{D}{D+const}$ scaling of the model.

The scaling with k_u should be the same as the scaling with f, as the probability of activation is proportional to k_u (at fixed k_b). This is seen Figure 6-14A (subject to saturating at a probability of 1). Since $k_b = \frac{\gamma}{l^2}$, where γ appears in the Equation 6.7, and l is the lattice spacing, we have $P \approx \frac{1}{\text{const}+k_b}$, which is the scaling on the right hand side of the plots in Figure 6-14. For small k_b, P increases with k_b. In the model, increasing k_b makes TCR-pMHC bind faster to coreceptor both with and without Lck; binding to coreceptor without Lck slows down the search process, but binding to coreceptor with Lck is required. These two behaviors are seen in the figure (but not in the analytical solution, which captures only one of them - this suggests that the approximation made is not valid in this case). For small k_b, the TCR-pMHC will not bind to coreceptor without Lck and will mostly be in the free state. However, the model assumes that it is always bound - probably where the approximation breaks down. If only the equilibrium constant k_u/k_b is known, we have $k_b = \theta k_u$. In this case, the analytical solution simplifies to $\lambda = \frac{fDk_u}{\theta k_u l^2 + D}$ or $\lambda \approx \frac{k_u}{k_u + \text{const}}$. This is the scaling seen in Figure 6-14C.

To understand why the PSSA fails at low C and high l, we plot the time courses

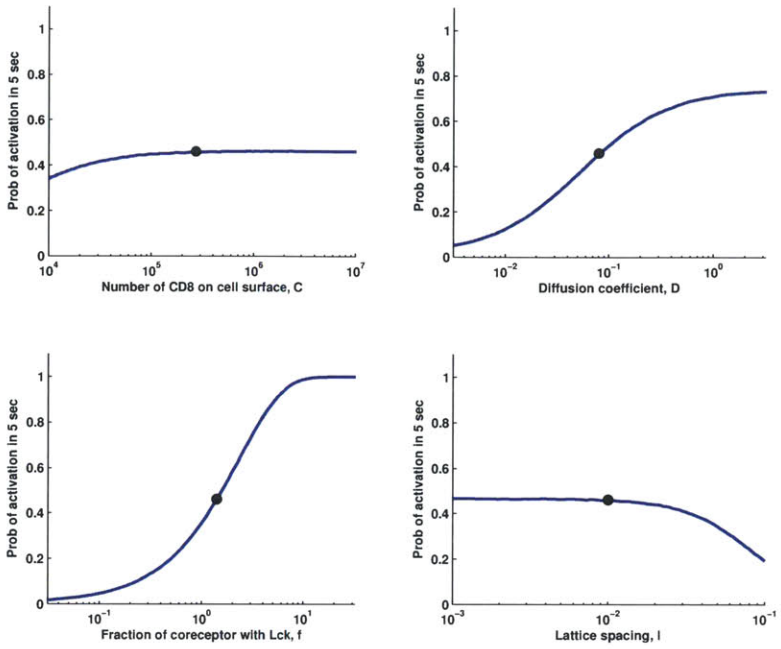


Figure 6-13: Plots of the activation probability in 5 sec as a function of (A) Number of CD8 on cell surface (B) Diffusion coefficient (C) Fraction of coreceptor with Lck and (D) Lattice Spacing. The black dot in each figure represents the original parameter set.

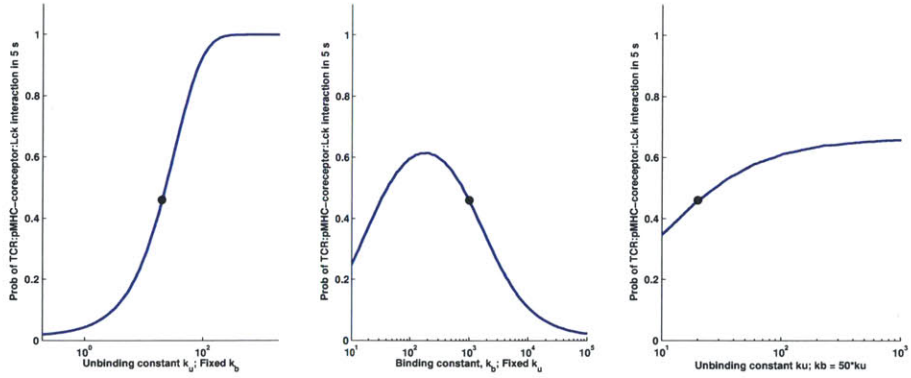
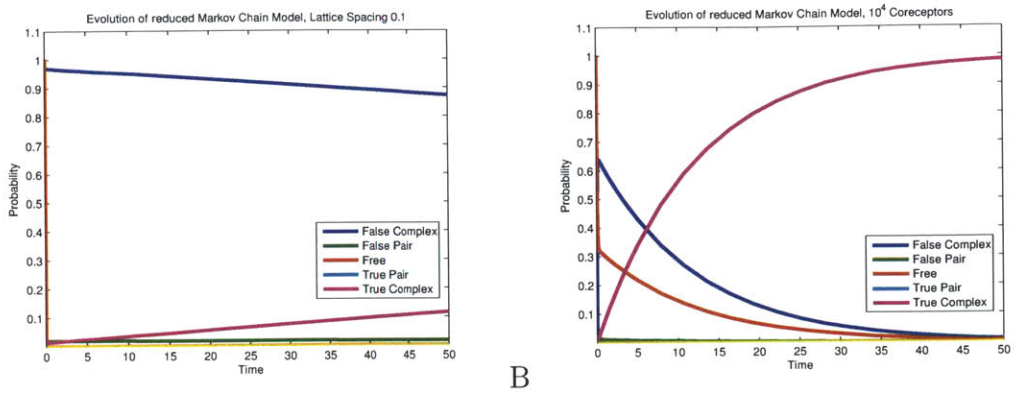


Figure 6-14: Plots of the activation probability in 5 sec as a function of MHC coreceptor (A) unbinding rate (B) binding rate (C) unbinding rate, if the equilibrium constant is fixed. The black dot in each figure represents the original parameter set.



A B

Figure 6-15: Time courses (probability of each state vs. time) for situations where the PSSA fails (A) For large lattice spacing (B) Small number of coreceptors on the cell surface. The PSSA assumes that the probabilities of the False Pair, Free and True Pair states are constant with time.

for the full Markov Chain models (Figure 6-15) for these cases. For the case of low c ($C = 10^4$ molecules on the surface of the T-cell, as compared to 273000 normally), we see that the probability of the system being in the "free" state is high (as fewer coreceptors are present) and the PSSA fails. For the case of large lattice spacing l ($l = 0.1\mu m$, ten times its default value), the reason for PSSA failing is unclear.

6.6 SSC Simulations of the Markov Chain

There are two major approximations made while formulating the analytical model. They are (1) TCR-pMHC dissociation is ignored and (2) Spatial effects are neglected. We cannot analytically solve the system without making these approximations, so we run stochastic reaction-diffusion simulations of the system using SSC[34]. The SSC simulations are similar to the previous MC model. Molecules can diffuse around on a $1\mu m \times 1\mu m$ patch of membrane which is divided into a 100×100 grid, and molecules on the same lattice site can react. On this patch of membrane, there are $C = 3425$ coreceptors without and $CL = 49$ coreceptors with Lck. There is one TCR-pMHC which can dissociate and associate with rates k_{off} and k_{on}. The diffusion constant is $k_{diff} = 800s^{-1}$; MHC-Coreceptor on and off rates are $k_b = 1000s^{-1}$ and $k_u = 20s^{-1}$. Reactions are similar to that of the Markov Chain model, and we look at the for-

mation of the TCR-pMHC-CD8-Lck complex (product). 1000 trajectories were run for each parameter set and statistics collected. The difference between the previous model looking at evolution of probabilities and the SSC simulation should be the spatial effect. In the simulation, rebinding should happen more frequently because "escape," i.e. jumping one lattice site away, does not necessarily mean the particles are now well mixed. However, there are now 10000 lattice sites and about 3500 coreceptor molecules, so the density of coreceptor molecules is high.

The first two plots are histograms of (a) Time at which the TCR-pMHC first finds a CD8 with Lck and (b) Free time distribution for pMHC-TCR (histogram of times for which it is actually free and diffusing around): We can also from the SSC results look at the efficiency of the search, i.e. the number of coreceptors that it must sample to first find one with Lck: The simulation also tells us, on average, how many times each coreceptor was sampled. The x axis of this figure represents the total number of different coreceptors bound at any point of time during the simulation, and the y axis represents the number of times a coreceptor to pMHC-TCR binding event occurred in that time. The data is (roughly) clustered around the line $y = 1.3x$, meaning that each coreceptor was on average bound 1.3 times by the TCR-pMHC.

6.7 Comparison of different cell types

6.7.1 CD8s

We can perform the same set of calculations for various cell types in the experiment. There are 3 CD8 cell types: DP restricted with MHC class I, peripheral (naive) CD8s and the CD8.4 chimera hybrid cell line. The main differences between these cell types are (1) DP CD8s have only 1.4% of coreceptor bound to Lck, as compared to the peripheral or chimeras which have 15.1% and 9.8% respectively(Table 6.2 - since activation rate scales linearly with this fraction, we should expect these cell lines to activate about an order of magnitude faster than the DPs. This is indeed the case,

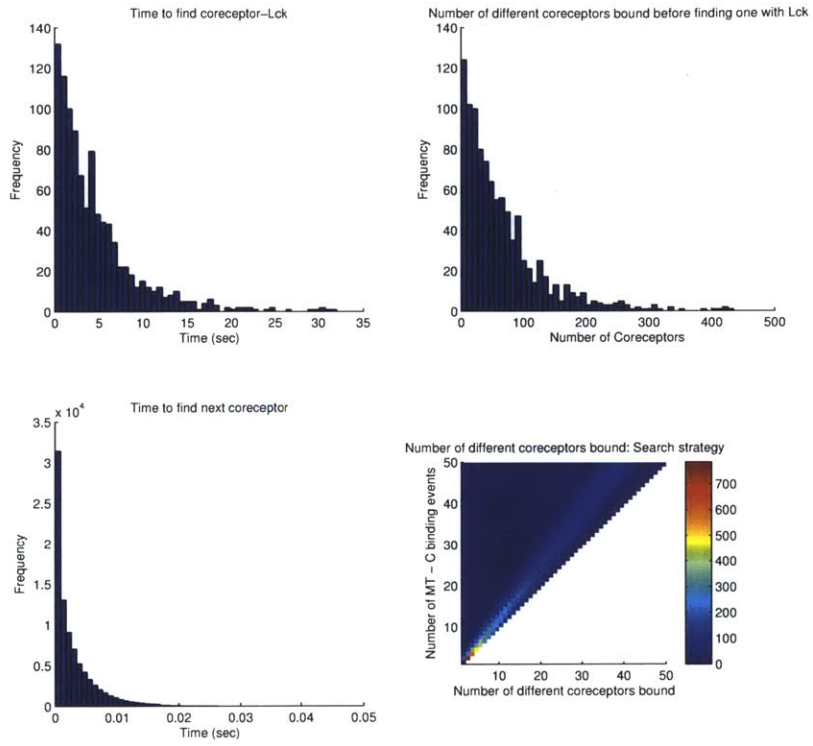


Figure 6-16: Quantification of various features of the stochastic search process. (A) Distribution of times for TCR-pMHC to find coreceptor-Lck for a strong ligand. (B) Number of different coreceptors bound before finding one with Lck. Mean of this distribution is around 66. (C) Distribution of times for which the TCR-pMHC is free, i.e. not bound to any coreceptor. These times are small compared to the overall search time. (D) A plot of number of binding events vs. number of coreceptors bound. The data is clustered around the line $y = 1.3x$, meaning that each coreceptor is bound 1.3 times on average.

both in experiment (4 sec vs. 600 ms dwell times) and in the simulations (Figure 6-19). The other major difference is that the number of TCRs is much higher, but the effect of this is not directly captured in the model, only in simulations.

Figure 6-17 shows the result of simulations: the probability of activation for various CD8s and various pMHC-TCR k_{off} as a function of time. DP CD8s activate much slower than peripheral or chimeric T-Cells; strongest ligands activate fastest at short to moderate times, as expected. Some curves seem to cross at long times; this

crossing is more prominent in Figure 6-18 and is probably because the simulations start with pMHC and TCR bound. In the small number of cases that these unbind, fast diffusion means that they diffuse apart and it is harder for pMHC to find TCR again on these timescales, which leads to a decrease in the activation probability. This is potentially the reason that we see an optimal diffusion rate in the first part of this work (for example, Figures 6-3 and 6-5). The $\frac{D}{D+const}$ scaling for diffusion for short times is also seen in Figure 6-18.

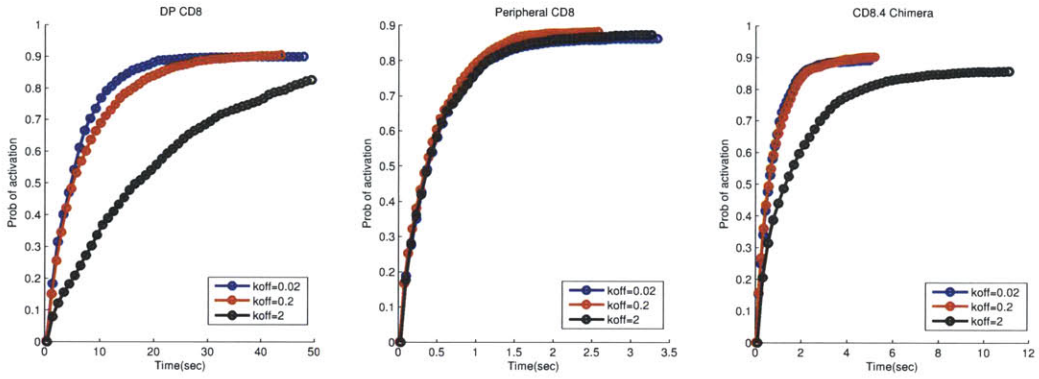


Figure 6-17: Activation curves for various cell types as a function of pMHC-TCR strength: DP CD8 (A), Peripheral CD8 (b) and the CD8.4 Chimera (C). Note the differing time scales on the X-axis; activation of the DP CD8 much slower than the other cell types.

A comparison of theory, experiment and simulation for CD8s is shown in Figure 6-19. The plots show activation curves from simulation (black, red and blue curves for various TCR-pMHC strengths), the theoretical curve in the limit of zero k_{off} (green curve), and experimentally determined dwell time necessary for activation (vertical magenta line). The dwell time lines intersect the theoretical and simulation curves at probabilities between 0.4 and 0.6. This suggests that the time for which TCR and APC need to be in contact in experiments (the dwell time) correspond, in theory and simulation, to the time taken for the same cell to have an activation probability of 50%. This shows excellent agreement between model and experiment, suggesting that the dwell time needed for activation is largely determined by the fraction of coreceptor that is bound to Lck.

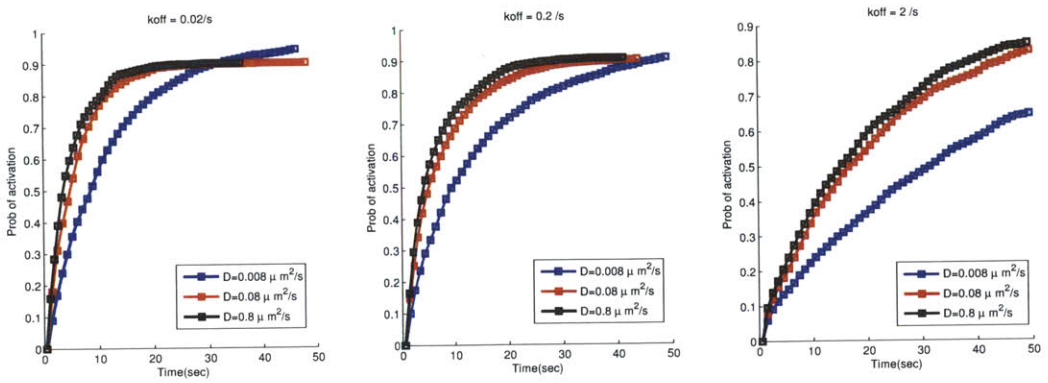


Figure 6-18: Effect of diffusion on activation of DP CD8s. Plots are for D = 0.008, 0.08 and $0.8\mu m^2 s^{-1}$ for different k_{off}s. Broadly, increasing diffusion speeds up activation up to a point, similar to the $\frac{D}{D+const}$ scaling suggested by the analytical method. However, some curves cross at long times.

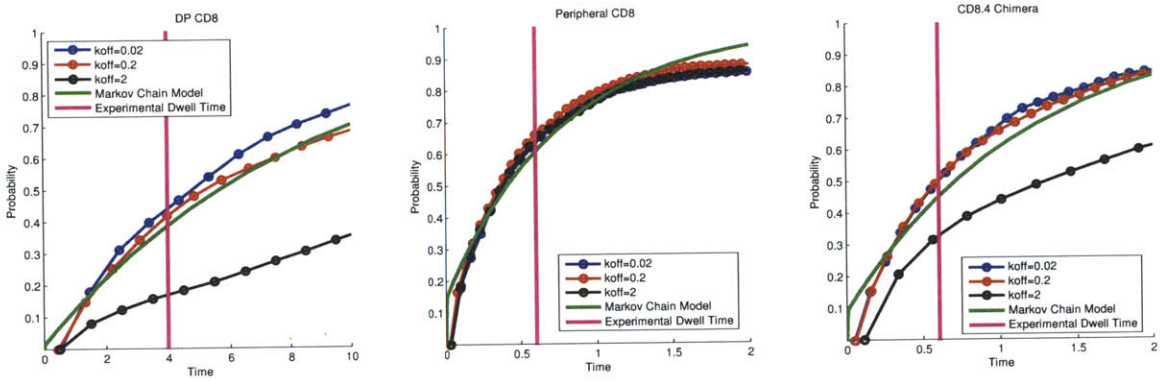


Figure 6-19: Comparison of the Markov Chain model, simulations and experimental dwell times for (A) DP CD8s, (B) Peripheral CD8s and (C) CD8.4 Chimera cells. The blue curves (strongest ligands) should correspond most closely to the model, and black curves least closely.

6.7.2 CD4s

It is known that the binding between coreceptor and MHC for CD4s is a factor of 2 to 20 weaker than CD8s[142]. We can model this decreased binding in two ways, by changing either the MHC-coreceptor on (k_b) or off (k_u) rates. It is also known that CD4s activate much faster than CD8s, by a factor of 10 or more. The model

predicts that the activation rate varies as $\lambda = \frac{fDk_u}{\gamma+D}$, where γ is proportional to k_b. So decreasing the binding constant k_b would not have as great an effect on activation rate as increasing the unbinding rate k_u. This is because increasing k_u makes only the the "false" TCR-pMHC-coreceptor complex (without Lck) less stable (the true complex is assumed not to dissociate on the timescale during which phosphorylation happens), and decreasing k_b affects formation of both complex with and without Lck. So in the following, we vary k_u and we can see how much weaker the CD4-MHC interaction must be as compared to CD8s in order to get activation times of around 80 ms.

Figure 6-20A and B consist of probability of activation of DP CD4s and peripheral CD4s as a function of time, and are for the strongest ligand ($k_{off} = 0.02s^{-1}$). All k_us described are in units of s^{-1}. Plots are for various strengths of CD4-MHC unbinding; the value for CD8s is $20s^{-1}$. Since pMHC-CD4 strengths are thought to be a factor of 2- to 20- fold lower than for CD8s, the k_u should be a factor of 2- to 20- fold higher. Dots are SSC simulations and solid lines are Markov Chain model results. With increasing unbinding from the false pair state, the cell activates faster; we can compare k_us by using the time for 50% activation. Figures 6-20 represent time taken for 50% activation as a function of k_u. The analytical model gives the result for probability of activation for short times as $P(t) = \lambda t$, where $\lambda = \frac{fDk_u}{\gamma+D}$. Since we are looking at the case where $P = 0.5$(constant), we expect the scaling of activation time to be $t \propto 1/k_u$. The plots are therefore of time for 50% activation vs. $1/k_u$, and we can get the value of k_u for which t = 80 ms = 0.08 sec (experimental dwell time). For DP CD4s, we get $k_u = 513$ and 154 from simulations and model respectively, about 8- to 25- fold weaker than CD8s, whereas for Peripheral CD4s, we get $k_u = 66$ and 59 from simulations and model respectively, about 3- fold weaker than CD8s. These are all similar to the range of the 2- to 20- fold less affinity as thought previously. (Note that we are estimating k_u from a line that scales as $1/k_u$, near zero, so the errors in this estimation can be pretty large.) It is not clear why the $k_u = 200$ curve for DP CD4 is quite far off from theory. Also, these calculations assume all the Lck is capable of phosphorylating TCR (basal or active Lck). The probability

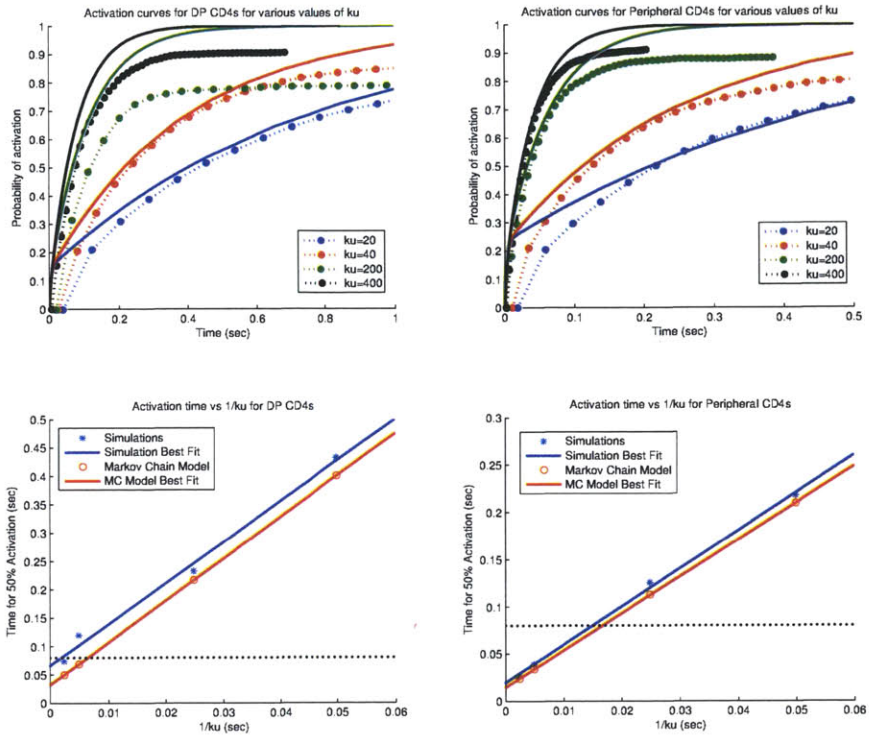


Figure 6-20: Activation probabilities and times for CD4s. The top row of figures are plots of activation curves: probabilities of activation vs. time, for (A) DP CD4s and (B) Peripheral CD4s for various values of the MHC-coreceptor unbinding constant, k_u which is an unknown parameter. Lines indicate model results and dots are from SSC simulations. Bottom row: Plots of time for which probability of activation = 0.5 vs. $1/k_u$ for (C) DP CD4s and (D) Peripheral CD4s. The points are data from simulations, lines are the best-fits. Experimentally, the dwell time is about 0.08 sec, so we can calculate the value of k_u that would correspond to these dwell times.

of activation P is proportional to the fraction of coreceptor with Lck (f), so if only about 26% of the Lck is capable of phosphorylating TCR (as some of the new data suggests), the model and simulations will be about 4 times slower than what it is now.

6.8 Discussion

In this section we have shown how the threshold activation strength and dwell time required for different T Cell types to activate is determined by the search process of a TCR-pMHC finding a coreceptor with Lck. Even for a pMHC that binds extremely strongly to a TCR (infinite half-life), the cell cannot activate before the TCR has found an Lck molecule to phosphorylate it. The first requirement for this is the formation of the four-body TCR-pMHC-coreceptor-Lck complex. We have built a model for the rate of formation of this complex, and find that this timescale closely corresponds to the dwell time required for the T Cell to activate, for many different cell types. This seems to indicate that the formation of this complex, rather than the subsequent phosphorylation step, is rate-limiting and determines the dwell time. Since the TCR and pMHC have to remain bound during this process, this also sets a threshold k_{off} for a TCR-pMHC interaction for each cell type above which the cell cannot activate.

We find that this search process depends on various parameters that differ between T Cell types. Running spatial SSC simulations of the full system suggest that the rate of activation should increase with the fraction of coreceptor bound to Lck and is maximum at some intermediate value of the diffusion constant. We built a simple Markov Chain based model and solved this analytically using the pseudo steady-state approximation. The solution, which is valid in the limit of a small fraction of coreceptor bound to Lck, suggests that the rate of activation increases linearly with the fraction of coreceptor bound to Lck and saturates with diffusion. The activation rate is also independent of the total amount of coreceptor and the lattice spacing used to discretize the problem. The activation rate being independent of the total amount of coreceptor is easy to rationalize: the total amount of coreceptor increases both the number of targets and temporary traps by the same factor, and therefore has no net effect on the total rate of activation.

The behavior with diffusion is contrary to what the full simulations show; the difference is probably in the fact that the first set of simulations is allowed to run for a long time and we are looking at steady-state behavior. The decrease in activation with diffusion (beyond the optimum) in the long-time limit is presumably because in this case, all when complexes unbind, fast diffusion means that components diffuse apart before they can rebind. This is also seen in the SSC version of the simplified system: at long time, the slower- diffusing cases activate to a larger extent than the faster- diffusing ones. Again, this is probably a consequence of the way these models are set up: since we start with the TCR bound to pMHC and let them unbind, slow diffusion means that rebinding to the same TCR (rather than the pMHC having to search for another TCR) is more likely; however, this effect is important only in the case where the TCR-pMHC has (1) not yet found a coreceptor with Lck and (2) has unbound, so the time much be larger than $1/k_{off}$. The $\frac{\mathcal{D}}{\mathcal{D}+\text{const}}$ scaling of the analytically approximate solution also warrants some comment: at the infinite diffusion limit, one might naively expect no complex to form because the components would diffuse away as soon as they are brought together; however, the rate of bringing together these components also becomes very large. The effective result seems to be that the rate of formation of the complex saturates with diffusion to some finite limit.

These results suggest that the primary reasons why the CD4s activate much faster than CD8s are twofold. First, CD4s have a larger fraction of coreceptor that is bound to Lck than CD8s (this is also why naive peripheral CD8 T Cells tend to activate faster than DP CD8 thymocytes). Second, the weaker MHC-coreceptor binding in CD4s means that the temporary traps where TCR-pMHC is bound to coreceptor without Lck are much less effective, and hence the Lck-bearing coreceptors are found much faster.

Chapter 7

Conclusions

This thesis attempts to take a "system-wide" or "network" view of some key reaction networks involved in T Cell signaling. Many problems in immunology in particular (and biology in general) are just beginning to be thought about in this sort of framework, where we consider networks as a whole and try to understand how these particular networks (topologies) of interactions define the behavior of the system. The major downside to this approach is, in many cases, the lack of detailed knowledge of these interactions, and statistical-mechanical approaches may help shed light on such problems. The objective of these models is three-fold: (1) to understand the mechanisms of the underlying biological processes (2) to describe these systems from the principles of physics (such as statistical mechanics and network theory), which ensures that these models have a sound theoretical basis and (3) to make testable predictions, which can help in model discrimination. The key benefit of such mechanistic insights, of course, is that in enables us to tweak the system in more logical ways, for example, as the first step towards rational drug design.

In the previous chapters, we have looked at a variety of problems involving the modeling of different aspects of protein kinase reaction networks. Chapters 2, 3 and 4 describe the modeling of spatially homogeneous reaction networks, and the relation between network topology and behavior. These chapters show how one may use simple dynamical systems-based models to describe chemical reaction networks and infer

details about biological mechanisms from these models. In Chapter 3, we use these models to analyze how the amount of the CD45 changes the activity of Src kinases. The experimental observation that T-Cells show a maximum activity of Lck at an intermediate concentration of CD45, but B-Cells do not with Lyn, can be attributed to a difference in the corresponding reaction networks. The dependence of this activity on the amount of CD45 also suggests that this process may have a function consequence in enabling the activation of Lck by clustering CD45.

Chapter 4 is a model of TCR activation by up-regulation and knocking off the negative regulator Csk. Our models suggest that the extremely strong activation is a consequence of two effects: an increase in the amount of unphosphorylated TCR due to effect of increased Csk (and the TCR creation-degradation mechanism), and a temporary surge in the activity of Lck upon removing the excess Csk which was transfused into the system. This reaction network is an example of a "perfectively adaptive" system and shows interesting behavior in a control-theoretic sense, such as being a band-pass filter.

Chapter 5 describes an attempt to understand the heterogeneity in cancers at the level of protein kinase signaling networks. A general model for signaling in networks was been developed, analogous to models of neural networks, with mutations modeled as changes in the topology of this network. Modeling cancer growth as a mutation-selection problem, we looking for patterns in secondary mutations during the directed evolution process. We found that lowest degree nodes along the shortest paths from the driver mutation to effector nodes were the most conserved, and the frequency of multiple optima depended mainly on the number of feedback loops present in the starting networks. A model of the effect of inhibitors and escape was also made, suggesting that inhibiting high-degree nodes had the greatest initial effect but was also the easiest to escape.

Finally, in Chapter 6, we look at the problem of constant activation thresholds

for activation of various types of T-Cells. We built a reaction-diffusion model for the network involved in the search process by which a pMHC-TCR finds a coreceptor-Lck, which enabled us to understand how the threshold for activation is determined by the parameters of a particular cell type. We also developed an approximate analytical solution for a simplified Markov Chain form of the model, which predicted that the activation rate scaled linearly with the fraction of coreceptors with Lck, increases (slowly) with diffusion and is independent of the number of coreceptors on the surface of the cell. These models suggest that this search process, rather than phosphorylation, is the rate-limiting step in the activation of TCR, and CD4s activate much faster than CD8s because of (1) the weaker coreceptor-MHC bonds in CD4s and (2) the fact that they have more Lck.

Bibliography

[1] Makio Iwashima, Bryan A Irving, NS Van Oers, Andrew C Chan, and Arthur Weiss. Sequential interactions of the tcr with two distinct cytoplasmic tyrosine kinases. *Science*, 263(5150):1136–1139, 1994.

[2] Ellen N Neumeister, Yuexin Zhu, Stephane Richard, Cox Terhorst, Andrew C Chan, and Andrey S Shaw. Binding of zap-70 to phosphorylated t-cell receptor zeta and eta enhances its autophosphorylation and generates specific binding sites for sh2 domain-containing proteins. *Molecular and cellular biology*, 15(6):3171–3178, 1995.

[3] Tomas Brdicka, Theresa A Kadlecek, Jeroen P Roose, Alexander W Pastuszak, and Arthur Weiss. Intramolecular regulatory switch in zap-70: analogy with receptor tyrosine kinases. *Molecular and cellular biology*, 25(12):4924–4933, 2005.

[4] Weiguo Zhang, Joanne Sloan-Lancaster, Jason Kitchen, Ronald P Trible, and Lawrence E Samelson. Lat: the zap-70 tyrosine kinase substrate that links t cell receptor to cellular activation. *Cell*, 92(1):83–92, 1998.

[5] Joseph Lin and Arthur Weiss. Identification of the minimal tyrosine residues required for linker for activation of t cell function. *Science Signaling*, 276(31):29588, 2001.

[6] Jon CD Houtman, Yuichiro Higashimoto, Nazzareno Dimasi, Sangwoo Cho, Hiroshi Yamaguchi, Brent Bowden, Carole Regan, Emilio L Malchiodi, Roy Mariuzza, Peter Schuck, et al. Binding specificity of multiprotein signaling

complexes is determined by both cooperative interactions and affinity preferences. *Biochemistry*, 43(14):4170–4178, 2004.

[7] Dvora Beach, Ronnie Gonen, Yaron Bogin, Ilona G Reischl, and Deborah Yablonski. Dual role of slp-76 in mediating t cell receptor-induced activation of phospholipase c-γ1. *Journal of Biological Chemistry*, 282(5):2937–2946, 2007.

[8] Yaron Bogin, Carmit Ainey, Dvora Beach, and Deborah Yablonski. Slp-76 mediates and maintains activation of the tec family kinase itk via the t cell antigen receptor-induced association between slp-76 and itk. *Science Signaling*, 104(16):6638, 2007.

[9] Bruce T Seet, Donna M Berry, Jonathan S Maltzman, Jacob Shabason, Monica Raina, Gary A Koretzky, C Jane McGlade, and Tony Pawson. Efficient t-cell receptor signaling requires a high-affinity interaction between the gads c-sh3 domain and the slp-76 rxxk motif. *The EMBO journal*, 26(3):678–689, 2007.

[10] JC Stone et al. Regulation of ras in lymphocytes: get a grp. *Biochemical Society Transactions*, 34(Pt 5):858–861, 2006.

[11] Elisabeth Genot and Doreen A Cantrell. Ras regulation and function in lymphocytes. *Current opinion in immunology*, 12(3):289–294, 2000.

[12] J. Lin and A. Weiss. T cell receptor signalling. *Journal of cell science*, 114(2):243–244, 2001.

[13] José Alberola-Ila and Gabriela Hernández-Hoyos. The ras/mapk cascade and the control of positive selection. *Immunological reviews*, 191(1):79–96, 2003.

[14] G. Altan-Bonnet and R.N. Germain. Modeling t cell antigen discrimination based on feedback control of digital erk responses. *PLoS biology*, 3(11):e356, 2005.

[15] Nigel E Wardle. *Guide to Signal Pathways in Immune Cells.* Humana Press, 2009.

[16] Kendall J Blumer and Gary L Johnson. Diversity in function and regulation of map kinase pathways. *Trends in biochemical sciences*, 19(6):236–240, 1994.

[17] AJ Whitmarsh and RJ Davis. Transcription factor ap-1 regulation by mitogen-activated protein kinase signal transduction pathways. *Journal of Molecular Medicine*, 74(10):589–607, 1996.

[18] Javier De Las Rivas and Celia Fontanillo. Protein–protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS computational biology*, 6(6):e1000807, 2010.

[19] Benno Schwikowski, Peter Uetz, and Stanley Fields. A network of protein–protein interactions in yeast. *Nature biotechnology*, 18(12):1257–1261, 2000.

[20] Stephen R Proulx, Daniel EL Promislow, and Patrick C Phillips. Network thinking in ecology and evolution. *Trends in Ecology & Evolution*, 20(6):345–353, 2005.

[21] Eric Davidson and Michael Levin. Gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102(14):4935–4935, 2005.

[22] Christof Francke, Roland J Siezen, and Bas Teusink. Reconstructing the metabolic network of a bacterium from its genome. *TRENDS in Microbiology*, 13(11):550–558, 2005.

[23] W B Kristan. *Information Processing in the Nervous System.* Raven, New York, 1980.

[24] A. Prasad, J. Zikherman, J. Das, J.P. Roose, A. Weiss, and A.K. Chakraborty. Origin of the sharp boundary that discriminates positive and negative selection of thymocytes. *Proceedings of the National Academy of Sciences*, 106(2):528–533, 2009.

[25] J. Das, M. Ho, J. Zikherman, C. Govern, M. Yang, A. Weiss, A.K. Chakraborty, and J.P. Roose. Digital signaling and hysteresis characterize ras activation in lymphoid cells. *Cell*, 136(2):337–351, 2009.

[26] Sašo Čemerski, Jayajit Das, Jason Locasale, Phoebe Arnold, Emanuele Giurisato, Mary A Markiewicz, Daved Fremont, Paul M Allen, Arup K Chakraborty, and Andrey S Shaw. The stimulatory potency of t cell antigens is influenced by the formation of the immunological synapse. *Immunity*, 26(3):345–355, 2007.

[27] Maxim N Artyomov, Jayajit Das, Mehran Kardar, and Arup K Chakraborty. Purely stochastic binary decisions in cell signaling models without underlying deterministic bistabilities. *Proceedings of the National Academy of Sciences*, 104(48):18958–18963, 2007.

[28] Georg Fritz, Nicolas E Buchler, Terence Hwa, and Ulrich Gerland. Designing sequential transcription logic: a simple genetic circuit for conditional memory. *Systems and synthetic biology*, 1(2):89–98, 2007.

[29] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361, 1977.

[30] Kyeong-Hee Lee, Aaron R Dinner, Chun Tu, Gabriele Campi, Subhadip Raychaudhuri, Rajat Varma, Tasha N Sims, W Richard Burack, Hui Wu, Julia Wang, et al. The immunological synapse balances t cell receptor signaling and degradation. *Science Signaling*, 302(5648):1218, 2003.

[31] Sašo Čemerski, Jayajit Das, Emanuele Giurisato, Mary A Markiewicz, Paul M Allen, Arup K Chakraborty, and Andrey S Shaw. The balance between t cell receptor signaling and degradation at the center of the immunological synapse is determined by antigen quality. *Immunity*, 29(3):414–422, 2008.

[32] Qi-Jing Li, Aaron R Dinner, Shuyan Qi, Darrell J Irvine, Johannes B Huppa, Mark M Davis, and Arup K Chakraborty. Cd4 enhances t cell sensitivity to

antigen by coordinating lck accumulation at the immunological synapse. *Nature immunology*, 5(8):791–799, 2004.

[33] Dennis C Wylie, Yuko Hori, Aaron R Dinner, and Arup K Chakraborty. A hybrid deterministic-stochastic algorithm for modeling cell signaling dynamics in spatially inhomogeneous environments and under the influence of external fields. *The Journal of Physical Chemistry B*, 110(25):12749–12765, 2006.

[34] M. Lis, M.N. Artyomov, S. Devadas, and A.K. Chakraborty. Efficient stochastic simulation of reaction–diffusion processes via direct compilation. *Bioinformatics*, 25(17):2289–2291, 2009.

[35] Albert-László Barabási and Zoltan N Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113, 2004.

[36] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science Signaling*, 298(5594):824, 2002.

[37] Shai S Shen-Orr, Ron Milo, Shmoolik Mangan, and Uri Alon. Network motifs in the transcriptional regulation network of escherichia coli. *Nature genetics*, 31(1):64–68, 2002.

[38] Onn Brandman and Tobias Meyer. Feedback loops shape cellular signals in space and time. *Science Signaling*, 322(5900):390, 2008.

[39] Leland H Hartwell, John J Hopfield, Stanislas Leibler, and Andrew W Murray. From molecular to modular cell biology. *Nature*, 402:C47–C52, 1999.

[40] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[41] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

[42] John J Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the national academy of sciences*, 81(10):3088–3092, 1984.

[43] Stuart Kauffman, Carsten Peterson, Björn Samuelsson, and Carl Troein. Random boolean network models and the yeast transcriptional network. *Proceedings of the National Academy of Sciences*, 100(25):14796–14799, 2003.

[44] Núria Domedel-Puig, Pau Rué, Antonio J Pons, and Jordi García-Ojalvo. Information routing driven by background chatter in a signaling network. *PLoS computational biology*, 7(12):e1002297, 2011.

[45] Dana Pe'er. Bayesian network analysis of signaling networks: a primer. *Science Signaling*, 2005(281):pl4, 2005.

[46] Nir Friedman, Michal Linial, Iftach Nachman, and Dana Pe'er. Using bayesian networks to analyze expression data. *Journal of computational biology*, 7(3-4):601–620, 2000.

[47] Karen Sachs, Omar Perez, Dana Pe'er, Douglas A Lauffenburger, and Garry P Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science Signaling*, 308(5721):523, 2005.

[48] Gezhi Weng, Upinder S Bhalla, and Ravi Iyengar. Complexity in biological signaling systems. *Science*, 284:92–96, 1999.

[49] Jeremy S Edwards, Rafael U Ibarra, and Bernhard O Palsson. In silico predictions of escherichia coli metabolic capabilities are consistent with experimental data. *Nature biotechnology*, 19(2):125–130, 2001.

[50] Rafael U Ibarra, Jeremy S Edwards, and Bernhard O Palsson. Escherichia coli k-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, 420(6912):186–189, 2002.

[51] Robert J Prill, Julio Saez-Rodriguez, Leonidas G Alexopoulos, Peter K Sorger, and Gustavo Stolovitzky. Crowdsourcing network inference: the dream predictive signaling network challenge. *Science signaling*, 4(189):mr7, 2011.

[52] Daniel Marbach, Robert J Prill, Thomas Schaffter, Claudio Mattiussi, Dario Floreano, and Gustavo Stolovitzky. Revealing strengths and weaknesses of methods for gene network inference. *Proceedings of the National Academy of Sciences*, 107(14):6286–6291, 2010.

[53] Nir Friedman. Inferring cellular networks using probabilistic graphical models. *Science*, 303(5659):799–805, 2004.

[54] John Jeremy Rice, Yuhai Tu, and Gustavo Stolovitzky. Reconstructing biological networks using conditional correlation analysis. *Bioinformatics*, 21(6):765–773, 2005.

[55] Timothy S Gardner, Diego Di Bernardo, David Lorenz, and James J Collins. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science Signaling*, 301(5629):102, 2003.

[56] Richard Bonneau, David J Reiss, Paul Shannon, Marc Facciotti, Leroy Hood, Nitin S Baliga, Vesteinn Thorsson, et al. The inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biol*, 7(5):R36, 2006.

[57] Alexandre Irrthum, Louis Wehenkel, Pierre Geurts, et al. Inferring regulatory networks from expression data using tree-based methods. *PLoS One*, 5(9):e12776, 2010.

[58] Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.

[59] Yong-Yeol Ahn, James P Bagrow, and Sune Lehmann. Link communities reveal multiscale complexity in networks. *Nature*, 466(7307):761–764, 2010.

[60] John David Crawford. Introduction to bifurcation theory. *Reviews of Modern Physics*, 63(4):991, 1991.

[61] Matthew Scott, Terence Hwa, and Brian Ingalls. Deterministic characterization of stochastic genetic circuits. *Proceedings of the National Academy of Sciences*, 104(18):7402–7407, 2007.

[62] Albert Goldbeter, Geneviève Dupont, and Michael J Berridge. Minimal model for signal-induced ca2+ oscillations and for their frequency encoding through protein phosphorylation. *Proceedings of the National Academy of Sciences*, 87(4):1461–1465, 1990.

[63] Steven H Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering.* Westview Press, 2001.

[64] B Alberts et al. *Molecular Biology of the Cell.* Garland Science, 2002.

[65] Lufen Chang and Michael Karin. Mammalian map kinase signalling cascades. *Nature*, 410(6824):37–40, 2001.

[66] Maosong Qi and Elaine A Elion. Map kinase pathways. *Journal of Cell Science*, 118(16):3569–3572, 2005.

[67] AnhCo Nguyen, W Richard Burack, Jeffrey L Stock, Robert Kortum, Oleg V Chaika, Maryam Afkarian, William J Muller, Kenneth M Murphy, Deborah K Morrison, Robert E Lewis, et al. Kinase suppressor of ras (ksr) is a scaffold which facilitates mitogen-activated protein kinase activation in vivo. *Molecular and cellular biology*, 22(9):3035–3045, 2002.

[68] Jason W Locasale, Andrey S Shaw, and Arup K Chakraborty. Scaffold proteins confer diverse regulatory properties to protein kinase cascades. *Proceedings of the National Academy of Sciences*, 104(33):13307–13312, 2007.

[69] Andre Levchenko, Jehoshua Bruck, and Paul W Sternberg. Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling

and reduce its threshold properties. *Proceedings of the National Academy of Sciences*, 97(11):5818–5823, 2000.

[70] Joseph Lin, Angus Harding, Emanuele Giurisato, and Andrey S Shaw. Ksr1 modulates the sensitivity of mitogen-activated protein kinase pathway activation in t cells without altering fundamental system outputs. *Molecular and cellular biology*, 29(8):2082–2091, 2009.

[71] DN Dhanasekaran, K Kashef, CM Lee, H Xu, and EP Reddy. Scaffold proteins of map-kinase modules. *Oncogene*, 26(22):3185–3202, 2007.

[72] A. Weiss. Kinases and phosphatases of the immune system. *Immunological reviews*, 228(1):5–8, 2009.

[73] S.B. Gauld and J.C. Cambier. Src-family kinases in b-cell development and signaling. *Oncogene*, 23(48):8001–8006, 2004.

[74] E. Ingley. Src family kinases: regulation of their activities, levels and identification of new pathways. *Biochimica et Biophysica Acta (BBA)-Proteins & Proteomics*, 1784(1):56–65, 2008.

[75] K. Nika, L. Tautz, Y. Arimura, T. Vang, S. Williams, and T. Mustelin. A weak lck tail bite is necessary for lck function in t cell antigen receptor signaling. *Journal of Biological Chemistry*, 282(49):36000–36009, 2007.

[76] G. Sun, A.K. Sharma, and RJ Budde. Autophosphorylation of src and yes blocks their inactivation by csk phosphorylation. *Oncogene*, 17(12):1587, 1998.

[77] K. Nika, C. Soldani, M. Salek, W. Paster, A. Gray, R. Etzensperger, L. Fugger, P. Polzella, V. Cerundolo, O. Dushek, et al. Constitutively active lck kinase in t cells drives antigen receptor signal transduction. *Immunity*, 32(6):766–777, 2010.

[78] D. Davidson, M. Bakinowski, M.L. Thomas, V. Horejsi, and A. Veillette. Phosphorylation-dependent regulation of t-cell activation by pag/cbp,

a lipid raft-associated transmembrane adaptor. *Molecular and cellular biology*, 23(6):2017–2028, 2003.

[79] M. Kawabuchi, Y. Satomi, T. Takao, Y. Shimonishi, S. Nada, K. Nagai, A. Tarakhovsky, and M. Okada. Transmembrane phosphoprotein cbp regulates the activities of src-family tyrosine kinases. *Nature*, 404(6781):999–1003, 2000.

[80] U. D'Oro and J.D. Ashwell. Cutting edge: the cd45 tyrosine phosphatase is an inhibitor of lck activity in thymocytes. *The Journal of Immunology*, 162(4):1879–1883, 1999.

[81] L. McNeill, R.J. Salmond, J.C. Cooper, C.K. Carret, R.L. Cassady-Cain, M. Roche-Molina, P. Tandon, N. Holmes, and D.R. Alexander. The differential regulation of lck kinase phosphorylation sites by cd45 is critical for t cell receptor signaling responses. *Immunity*, 27(3):425–437, 2007.

[82] M.L. Hermiston, J. Zikherman, and J.W. Zhu. Cd45, cd148, and lyp/pep: critical phosphatases regulating src family kinase signaling networks in immune cells. *Immunological reviews*, 228(1):288–311, 2009.

[83] G.G. Chiang and B.M. Sefton. Specific dephosphorylation of the lck tyrosine protein kinase at tyr-394 by the shp-1 protein-tyrosine phosphatase. *Journal of Biological Chemistry*, 276(25):23173–23178, 2001.

[84] N.P. Kaimachnikov and B.N. Kholodenko. Toggle switches, pulses and oscillations are intrinsic properties of the src activation/deactivation cycle. *FEBS Journal*, 276(15):4102–4118, 2009.

[85] H. Fufl, W. Dubitzky, C. Stephen Downes, and M.J. Kurth. Src family kinases and receptors: analysis of three activation mechanisms by dynamic systems modeling. *Biophysical journal*, 94(6):1995–2006, 2008.

[86] R. Falahati and D. Leitenberg. Selective regulation of tcr signaling pathways by the cd45 protein tyrosine phosphatase during thymocyte development. *The Journal of Immunology*, 181(9):6082–6091, 2008.

[87] J. Zikherman, K. Doan, R. Parameswaran, W. Raschke, and A. Weiss. Quantitative differences in cd45 expression unmask functions for cd45 in b-cell development, tolerance, and survival. *Proceedings of the National Academy of Sciences*, 109(1):E3–E12, 2012.

[88] J.R. Schoenborn, Y.X. Tan, C. Zhang, K.M. Shokat, and A. Weiss. Feedback circuits monitor and adjust basal lck-dependent events in t cell receptor signaling. *Science's STKE*, 4(190):ra59, 2011.

[89] M.L. Dustin and D. Depoil. New insights into the t cell synapse from single molecule techniques. *Nature Reviews Immunology*, 11(10):672–684, 2011.

[90] Charlotte Harrison. Kinase inhibitors: Analysing kinase inhibitor selectivity. *Nature Reviews Drug Discovery*, 11(1):21–21, 2012.

[91] Xi Chen, Haihong Ye, Rejji Kuruvilla, Narendrakumar Ramanan, Katherine W Scangos, Chao Zhang, Nicolas M Johnson, Pamela M England, Kevan M Shokat, and David D Ginty. A chemical-genetic approach to studying neurotrophin signaling. *Neuron*, 46(1):13–21, 2005.

[92] Jamie R Schoenborn, Chao Zhang, Kevan M Shokat, and Arthur Weiss. A chemical-genetic approach to inhibit csk activates t cells independently of the t cell receptor. *Journal of Immunology (Meeting Abstract Supplement)*, 182(35.4), 2009.

[93] Tanja Tamgüney, Chao Zhang, Dorothea Fiedler, Kevan Shokat, and David Stokoe. Analysis of 3-phosphoinositide-dependent kinase-1 signaling and function in es cells. *Experimental cell research*, 314(11):2299–2312, 2008.

[94] Christian Schmedt and Alexander Tarakhovsky. Autonomous maturation of α/βt lineage cells in the absence of cooh-terminal src kinase (csk). *The Journal of experimental medicine*, 193(7):815–826, 2001.

[95] Akira Imamoto and Philippe Soriano. Disruption of the¡ i¿ csk¡/i¿ gene, encoding a negative regulator of src family tyrosine kinases, leads to neural tube defects and embryonic lethality in mice. *Cell*, 73(6):1117–1124, 1993.

[96] Emil H Palacios and Arthur Weiss. Function of the src-family kinases, lck and fyn, in t-cell development and activation. *Oncogene*, 23(48):7990–8000, 2004.

[97] Salvatore Valitutti, Sabina Müller, Marina Cella, Elisabetta Padovan, and Antonio Lanzavecchia. Serial triggering of many t-cell receptors by a few peptide mhc complexes. *Nature*, 375(6527):148–151, 1995.

[98] Mayumi Naramura, Ihn-Kyung Jang, Hemanta Kole, Fang Huang, Diana Haines, and Hua Gu. c-cbl and cbl-b regulate t cell responsiveness by promoting ligand-induced tcr down-modulation. *Nature immunology*, 3(12):1192–1199, 2002.

[99] Michael S Krangel. Endocytosis and recycling of the t3-t cell receptor complex. the role of t3 phosphorylation. *The Journal of experimental medicine*, 165(4):1141–1159, 1987.

[100] Yasuhiro Minami, LE Samelson, and RD Klausner. Internalization and cycling of the t cell antigen receptor. role of protein kinase c. *Journal of Biological Chemistry*, 262(27):13342–13347, 1987.

[101] Salvatore Valitutti, Sabina Müller, Mariolina Salio, and Antonio Lanzavecchia. Degradation of t cell receptor (tcr)–cd3-ζ complexes after antigenic stimulation. *The Journal of experimental medicine*, 185(10):1859–1864, 1997.

[102] Damon Hou, Cristina Cenciarelli, Jane P Jensen, HB Nguygen, and Allan M Weissman. Activation-dependent ubiquitination of a t cell antigen receptor

subunit on multiple intracellular lysines. *Journal of Biological Chemistry*, 269(19):14244–14247, 1994.

[103] Trond Methi, Torunn Berge, Knut Martin Torgersen, and Kjetil Taskén. Reduced cbl phosphorylation and degradation of the ζ-chain of the t-cell receptor/cd3 complex in t cells with low lck levels. *European journal of immunology*, 38(9):2557–2563, 2008.

[104] Marina von Essen, Charlotte Menné Bonefeld, Volkert Siersma, Anette Bødker Rasmussen, Jens Peter H Lauritsen, Bodil L Nielsen, and Carsten Geisler. Constitutive and ligand-induced tcr degradation. *The Journal of Immunology*, 173(1):384–393, 2004.

[105] MATLAB. *7.11.0.584 (R2010b)*. The MathWorks Inc., Natick, Massachusetts, 2010.

[106] N Barkal and Stanislas Leibler. Robustness in simple biochemical networks. *Nature*, 387(6636):913–917, 1997.

[107] Andre Levchenko and Pablo A Iglesias. Models of eukaryotic gradient sensing: application to chemotaxis of amoebae and neutrophils. *Biophysical journal*, 82(1 Pt 1):50, 2002.

[108] Jerome T Mettetal, Dale Muzzey, Carlos Gómez-Uribe, and Alexander van Oudenaarden. The frequency dependence of osmo-adaptation in saccharomyces cerevisiae. *Science Signaling*, 319(5862):482, 2008.

[109] Hugh R Matthews and Johannes Reisert. Calcium, the two-faced messenger of olfactory transduction and adaptation. *Current opinion in neurobiology*, 13(4):469–475, 2003.

[110] H El-Samad, JP Goff, and M Khammash. Calcium homeostasis and parturient hypocalcemia: an integral feedback perspective. *Journal of theoretical biology*, 214(1):17–29, 2002.

[111] Tau-Mu Yi, Yun Huang, Melvin I Simon, and John Doyle. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proceedings of the National Academy of Sciences*, 97(9):4649–4653, 2000.

[112] George Stephanopoulos. *Chemical Process Control: An Introduction to Theory and Practice.* Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1984.

[113] Dominico Vigil, Jacqueline Cherfils, Kent L Rossman, and Channing J Der. Ras superfamily gefs and gaps: validated and tractable targets for cancer therapy? *Nature Reviews Cancer*, 10(12):842–857, 2010.

[114] Catherine Hartzell, Olga Ksionda, Ed Lemmens, Kristen Coakley, Ming Yang, Monique Dail, Richard C Harvey, Christopher Govern, Jeroen Bakker, Tineke L Lenstra, et al. Dysregulated rasgrp1 responds to cytokine receptor input in t cell leukemogenesis. *Science signaling*, 6(268):ra21, 2013.

[115] L.J. Van't Veer and R. Bernards. Enabling personalized cancer medicine through analysis of gene-expression patterns. *Nature*, 452(7187):564–570, 2008.

[116] L.M.F. Merlo, J.W. Pepper, B.J. Reid, and C.C. Maley. Cancer as an evolutionary and ecological process. *Nature Reviews Cancer*, 6(12):924–935, 2006.

[117] P. François and V. Hakim. Design of genetic networks with specified functions by evolution in silico. *Proceedings of the National Academy of Sciences of the United States of America*, 101(2):580, 2004.

[118] R. Albert, H. Jeong, and A.L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.

[119] E.O. Voit. *Computational Analysis of Biochemical Systems. A Practical Guide for Biochemists and Molecular Biologists.* Cambridge University Press, Cambridge, U.K., 2000.

[120] R. Albert. Scale-free networks in cell biology. *Journal of cell science*, 118(21):4947–4957, 2005.

[121] U. Stelzl, U. Worm, M. Lalowski, C. Haenig, F.H. Brembeck, H. Goehler, M. Stroedicke, M. Zenkner, A. Schoenherr, S. Koeppen, et al. A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, 122(6):957–968, 2005.

[122] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras. Generation of uncorrelated random scale-free networks. *Physical Review E*, 71(2):027103, 2005.

[123] A. Müller, R.M. MacCallum, and M.J.E. Sternberg. Structural characterization of the human proteome. *Genome research*, 12(11):1625–1641, 2002.

[124] J.M.G. Izarzugaza, O.C. Redfern, C.A. Orengo, and A. Valencia. Cancer-associated mutations are preferentially distributed in protein kinase functional sites. *Proteins: Structure, Function, and Bioinformatics*, 77(4):892–903, 2009.

[125] M.K. Morris, J. Saez-Rodriguez, P.K. Sorger, and D.A. Lauffenburger. Logic-based models for the analysis of cell signaling networks. *Biochemistry*, 49(15):3216–3224, 2010.

[126] C. Greenman, P. Stephens, R. Smith, G.L. Dalgliesh, C. Hunter, G. Bignell, H. Davies, J. Teague, A. Butler, C. Stevens, et al. Patterns of somatic mutation in human cancer genomes. *Nature*, 446(7132):153–158, 2007.

[127] Paul François, Vincent Hakim, and Eric D Siggia. Deriving structure from evolution: metazoan segmentation. *Molecular systems biology*, 3(1), 2007.

[128] Paul François and Eric D Siggia. A case study of evolutionary computation of biochemical adaptation. *Physical biology*, 5(2):026009, 2008.

[129] Kenneth Murphy. *Janeway's Immunobiology*. Garland Science, 2011.

[130] Christine Borowski, Colin Martin, Fotini Gounari, Loralee Haughn, Iannis Aifantis, Fabio Grassi, and Harald von Boehmer. On the brink of becoming a t cell. *Current opinion in immunology*, 14(2):200–206, 2002.

[131] Ken Shortman and Li Wu. Early t lymphocyte progenitors. *Annual review of immunology*, 14(1):29–47, 1996.

[132] Thomas J Kindt, Richard A Goldsby, Barbara Anne Osborne, and Janis Kuby. *Kuby's Immunology, 6th edition.* W.H. Freeman, New York, 2007.

[133] John J Hopfield. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proceedings of the National Academy of Sciences*, 71(10):4135–4139, 1974.

[134] Timothy W Mckeithan. Kinetic proofreading in t-cell receptor signal transduction. *Proceedings of the national academy of sciences*, 92(11):5042–5046, 1995.

[135] M.N. Artyomov, M. Lis, S. Devadas, M.M. Davis, and A.K. Chakraborty. Cd4 and cd8 binding to mhc molecules primarily acts to enhance lck delivery. *Proceedings of the National Academy of Sciences*, 107(39):16916–16921, 2010.

[136] Phillip D Holler and David M Kranz. Quantitative analysis of the contribution of tcr/pepmhc affinity and cd8 to t cell activation. *Immunity*, 18(2):255–264, 2003.

[137] Immanuel F Luescher, Eric Vivier, Andréas Layer, Jérôme Mahiou, François Godeau, Bernard Malissen, and Pedro Romero. Cd8 modulation of t-cell antigen receptor–ligand interactions on living cytotoxic t lymphocytes. *Nature*, 373:353–356, 1995.

[138] Mark A Daniels, Emma Teixeiro, Jason Gill, Barbara Hausmann, Dominique Roubaty, Kaisa Holmberg, Guy Werlen, Georg A Holländer, Nicholas RJ Gascoigne, et al. Thymic selection threshold defined by compartmentalization of ras/mapk signalling. *Nature*, 444(7120):724–729, 2006.

[139] Dieter Naeher, Mark A Daniels, Barbara Hausmann, Philippe Guillaume, Immanuel Luescher, and Ed Palmer. A constant affinity threshold for t cell tolerance. *The Journal of experimental medicine*, 204(11):2553–2559, 2007.

[140] Ed Palmer and Dieter Naeher. Affinity threshold for thymic selection through a t-cell receptor–co-receptor zipper. *Nature Reviews Immunology*, 9(3):207–213, 2009.

[141] Lee A Segel and Marshall Slemrod. The quasi-steady-state assumption: a case study in perturbation. *SIAM review*, 31(3):446–477, 1989.

[142] George F Gao, Zihe Rao, and John I Bell. Molecular coordination of $\alpha\beta$t-cell receptors and coreceptors cd8 and cd4 in their recognition of peptide-mhc ligands. *Trends in immunology*, 23(8):408–413, 2002.

Appendix A

Supplementary Material for the CD45 Model

Name	Initial Amount	Units
Lck	100.0	molecule
CD45	25.0	molecule
PAG	1000.0	molecule
Csk	10.0	molecule

Table A.1: List of Initial Concentrations in the Src Kinase Model

N.	**Name**	**Value**	**Units**
1	kLLon	0.1	1/(molecule*second)
2	kLLoff	1.0	1/second
3	kAcat	1.0	1/second
4	kBcat	0.1	1/second
5	kCcat	0.02	1/second
6	kDLon	1.1	1/(molecule*second)
7	kDLoff	0.1	1/second
8	kDLcat394	0.3	1/second
9	kDLcat505	0.8	1/second
10	kLSon	0.8	1/(molecule*second)
11	kLSoff	0.1	1/second
12	kLScat	0.1	1/second
13	kLPon	1.0	1/(molecule*second)
14	kLPoff	0.1	1/second
15	kPSon	1.0	1/(molecule*second)
16	kPSoff	0.1	1/second
17	kPDon	3.0	1/(molecule*second)
18	kPDoff	0.1	1/second
19	kPDcat	1.0	1/second
20	kLdephos	0.001	1/second

Table A.2: List of Rate Constants in the Src Kinase Model

N.	**Reaction**	**Rate Expression**
1	LckI + CD45 → [LckI:CD45]	kDLon*LckI*CD45
2	[LckI:CD45] → LckI + CD45	kDLoff*[LckI:CD45]
3	[LckI:CD45] → LckB + CD45	kDLcat505*[LckI:CD45]
4	LckB + [Csk:PAGp] → [LckB:Csk:PAGp]	kLSon*LckB*[Csk:PAGp]
5	[LckB:Csk:PAGp] → LckB + [Csk:PAGp]	kLSoff*[LckB:Csk:PAGp]
6	[LckB:Csk:PAGp] → LckI + [Csk:PAGp]	kLScat*[LckB:Csk:PAGp]
7	LckB + LckB → [LckB:LckB]	kLLon*LckB*LckB
8	[LckB:LckB] → LckB + LckB	kLLoff*[LckB:LckB]
9	[LckB:LckB] → LckA + LckB	kBcat*[LckB:LckB]
10	LckB + LckA → [LckB:LckA]	kLLon*LckB*LckA
11	[LckB:LckA] → LckB + LckA	kLLoff*[LckB:LckA]
12	[LckB:LckA] → LckA + LckA	kAcat*[LckB:LckA]
13	LckA + CD45 → [LckA:CD45]	kDLon*LckA*CD45
14	[LckA:CD45] → LckA + CD45	kDLoff*[LckA:CD45]
15	[LckA:CD45] → LckB + CD45	kDLcat394*[LckA:CD45]
16	LckI + LckB → [LckI:LckB]	kLLon*LckI*LckB
17	[LckI:LckB] → LckI + LckB	kLLoff*[LckI:LckB]
18	LckI + LckA → [LckI:LckA]	kLLon*LckI*LckA
19	[LckI:LckA] → LckI + LckA	kLLoff*[LckI:LckA]
20	LckC + CD45 → [LckC:CD45]	kDLon*LckC*CD45
21	[LckC:CD45] → LckC + CD45	kDLoff*[LckC:CD45]
22	[LckC:CD45] → LckA + CD45	kDLcat394*[LckC:CD45]
23	[LckC:CD45] → LckI + CD45	kDLcat505*[LckC:CD45]
24	LckB + LckC → [LckB:LckC]	kLLon*LckB*LckC
25	[LckB:LckC] → LckB + LckC	kLLoff*[LckB:LckC]
26	[LckB:LckC] → LckA + LckC	kCcat*[LckB:LckC]
27	LckI + LckC → [LckI:LckC]	kLLon*LckI*LckC
28	[LckI:LckC] → LckI + LckC	kLLoff*[LckI:LckC]
29	LckA + PAG → [LckA:PAG]	kLPon*LckA*PAG
30	LckB + PAG → [LckB:PAG]	kLPon*LckB*PAG
31	LckC + PAG → [LckC:PAG]	kLPon*LckC*PAG
32	[LckA:PAG] → LckA + PAG	kLPoff*[LckA:PAG]
33	[LckB:PAG] → LckB + PAG	kLPoff*[LckB:PAG]
34	[LckC:PAG] → LckC + PAG	kLPoff*[LckC:PAG]
35	[LckA:PAG] → LckA + PAGp	kAcat*[LckA:PAG]
36	[LckB:PAG] → LckB + PAGp	kBcat*[LckB:PAG]
37	[LckC:PAG] → LckC + PAGp	kCcat*[LckC:PAG]
38	Csk + PAGp → [Csk:PAGp]	kPSon*Csk*PAGp
39	[Csk:PAGp] → PAGp + Csk	kPSoff*[Csk:PAGp]
40	PAGp + CD45 → [PAGp:CD45]	kPDon*PAGp*CD45
41	[PAGp:CD45] → PAGp + CD45	kPDoff*[PAGp:CD45]
42	[PAGp:CD45] → PAG + CD45	kPDcat*[PAGp:CD45]
43	LckA → LckB	kLdephos*LckA
44	LckC → LckB	kLdephos*LckC
45	LckI → LckB	kLdephos*LckI
46	LckA + [Csk:PAGp] → [LckA:Csk:PAGp]	kLSon*LckA*[Csk:PAGp]
47	[LckA:Csk:PAGp] → LckA + [Csk:PAGp]	kLSoff*[LckA:Csk:PAGp]
48	[LckA:Csk:PAGp] → LckC + [Csk:PAGp]	kLScat*[LckA:Csk:PAGp]

Appendix B

Supplementary Material for the Csk Model

N.	**Reaction**	**Rate Expression**
1	cd45 + lcki → cd45lcki	kcd45lckon*cd45*lcki
2	cd45 + lckc → cd45lckc	kcd45lckon*cd45*lckc
3	cd45 + lcka → cd45lcka	kcd45lckon*cd45*lcka
4	cd45lcki → cd45 + lcki	kcd45lckoff*cd45lcki
5	cd45lckc → cd45 + lckc	kcd45lckoff*cd45lckc
6	cd45lcka → cd45 + lcka	kcd45lckoff*cd45lcka
7	cd45lcki → cd45 + lckb	klckib*cd45lcki
8	cd45lckc → cd45 + lckb	klckcb*cd45lckc
9	cd45lcka → cd45 + lckb	klckab*cd45lcka
10	2 lckb → lckbb	klckon*lckb*lckb
11	lckbb → 2 lckb	klckoff*lckbb
12	lckbb → 2 lcka	kautolckb*lckbb
13	pag + lcka → paglcka	klckpagon*pag*lcka
14	pag + lckb → paglckb	klckpagon*pag*lckb
15	pag + lckc → paglckc	klckpagon*pag*lckc

Continued on next page

Table B.1 – continued from previous page

N.	**Reaction**	**Rate Expression**
16	paglcka → pag + lcka	klckpagoff*paglcka
17	paglckb → pag + lckb	klckpagoff*paglckb
18	paglckc → pag + lckc	klckpagoff*paglckc
19	paglcka → pagp + lcka	klckapagphos*paglcka
20	paglckb → pagp + lckb	klckbpagphos*paglckb
21	paglckc → pagp + lckc	klckcpagphos*paglckc
22	pagp + csk → cskp	kcskpagon*pagp*csk
23	cskp → pagp + csk	kcskpagoff*cskp
24	cskp + lcka → cskplcka	kcsklckaon*cskp*lcka
25	cskp + lckb → cskplckb	klckcskon*cskp*lckb
26	cskplcka → cskp + lcka	klckcskoff*cskplcka
27	cskplckb → cskp + lckb	klckcskoff*cskplckb
28	cskplcka → cskp + lckc	klckcskphos*cskplcka
29	cskplckb → cskp + lcki	klckcskphos*cskplckb
30	cd45 + pagp → cd45pagp	kpagcd45on*cd45*pagp
31	cd45pagp → cd45 + pagp	kpagcd45off*cd45pagp
32	cd45pagp → cd45 + pag	kpagcd45cat*cd45pagp
33	lcka + lckb → lckalckb	klckon*lcka*lckb
34	lckalckb → lcka + lckb	klckoff*lckalckb
35	lckalckb → 2 lcka	kautolcka*lckalckb
36	lckc + lckb → lckclckb	klckon*lckc*lckb
37	lckclckb → lckc + lckb	klckoff*lckclckb
38	lckclckb → lckc + lcka	kautolckc*lckclckb
39	cd45 + cd45 → X	kcddon*cd45*cd45
40	X → 2 cd45	kcddoff*X
41	cskas + lckb → lckbas	kcskason*cskas*lckb

Continued on next page

Table B.1 – continued from previous page

N.	Reaction	Rate Expression
42	lckbas → cskas + lckb	kcskasoff*lckbas
43	lckbas → cskas + lcki	kcskascat*lckbas
44	cskas + lcka → lckaas	kcskason*cskas*lcka
45	lckaas → cskas + lcka	kcskasoff*lckaas
46	lckaas → cskas + lckc	kcskascat*lckaas
47	cskas + pp1 → Y	kpp1*cskas*pp1
48	lckb + cd3 → lckbcd3	klckcd3on*lckb*cd3
49	lckbcd3 → lckb + cd3	klckcd3off*lckbcd3
50	lckbcd3 → lckb + cd3p	klckbcat*lckbcd3
51	lcka + cd3 → lckacd3	klckcd3on*lcka*cd3
52	lckacd3 → lcka + cd3	klckcd3off*lckacd3
53	lckacd3 → lcka + cd3p	klckacat*lckacd3
54	cd3p → cd3	kcd3dphos*cd3p
55	cd3p → null	kcd3deg*cd3p
56	null → cd3	kcd3born
57	lckc + cd3 → lckccd3	klckcd3on*lckc*cd3
58	lckccd3 → lckc + cd3	klckcd3off*lckccd3
59	lckccd3 → lckc + cd3p	klckccat*lckccd3
60	lcka + sh → lckash	klckshon*lcka*sh
61	lckash → lcka + sh	klckshoff*lckash
62	lckash → lcka + shp	klckashp*lckash
63	lckb + sh → lckbsh	klckshon*lckb*sh
64	lckbsh → lckb + sh	klckshoff*lckbsh
65	lckbsh → lckb + shp	klckbshp*lckbsh
66	lckc + sh → lckcsh	klckshon*lckc*sh
67	lckcsh → lckc + sh	klckshoff*lckcsh

Continued on next page

Table B.1 – continued from previous page

N.	**Reaction**	**Rate Expression**
68	lckcsh → lckc + shp	klckcshp*lckcsh
69	shp → sh	kshpdeg*shp
70	lcka + shp → shplcka	kshplckon*lcka*shp
71	shplcka → lcka + shp	kshplckoff*shplcka
72	shplcka → lckb + shp	kshplckcat*shplcka
73	lckc + shp → shplckc	kshplckon*lckc*shp
74	shplckc → lckc + shp	kshplckoff*shplckc
75	shplckc → lcki + shp	kshplckcat*shplckc
76	T + lcka → Tlcka	kTlckaon*T*lcka
77	Tlcka → T + lcka	kTlckaoff*Tlcka
78	Tlcka → T + plcka	kTlckacat*Tlcka
79	plcka + cd3 → plckacd3	klckcd3on*plcka*cd3
80	plckacd3 → plcka + cd3	klckcd3off*plckacd3
81	plckacd3 → plcka + cd3p	klckacat*plckacd3
82	plcka → lcka	kplckadeg*plcka
83	cd3p → sos + cd3p	kcd3psos*cd3p
84	sos → null	ksosdeg*sos
85	sos + R → sosaR	ksosaRon*sos*R
86	sosaR → sos + R	ksosaRoff*sosaR
87	sos + R → soscR	ksoscRon*sos*R
88	soscR → sos + R	ksoscRoff*soscR
89	sos + T → sosaT	ksosaTon*sos*T
90	sosaT → sos + T	ksosaToff*sosaT
91	sosaR + R → sosaRcR	ksosaRcRon*sosaR*R
92	sosaRcR → sosaR + T	ksosaRcRcat*sosaRcR
93	sosaRcR → sosaR + R	ksosaRcRoff*sosaRcR

Continued on next page

Table B.1 – continued from previous page

N.	Reaction	Rate Expression
94	sosaRcR → soscR + R	ksosaRoff*sosaRcR
95	sosaT + R → sosaTcR	ksosaTcRon*sosaT*R
96	sosaTcR → sosaT + T	ksosaTcRcat*sosaTcR
97	sosaTcR → sosaT + R	ksosaTcRoff*sosaTcR
98	sosaTcR → soscR + T	ksosaToff*sosaTcR
99	soscR → sos + T	ksoscRcat*soscR
100	T + gap → Tgap	kTgapon*T*gap
101	Tgap → T + gap	kTgapoff*Tgap
102	Tgap → R + gap	kTgapcat*Tgap
103	plcka + pag → plckapag	klckpagon*plcka*pag
104	plckapag → plcka + pag	klckpagoff*plckapag
105	plckapag → plcka + pagp	klckapagphos*plckapag
106	R + grp → Rgrp	kRgrpon*R*grp
107	Rgrp → R + grp	kRgrpoff*Rgrp
108	Rgrp → T + grp	kRgrpcat*Rgrp
109	cd45 + plcki → cd45plcki	kcd45lckon*cd45*plcki
110	cd45 + plckc → cd45plckc	kcd45lckon*cd45*plckc
111	cd45plcki → cd45 + plcki	kcd45lckoff*cd45plcki
112	cd45plckc → cd45 + plckc	kcd45lckoff*cd45plckc
113	cd45plcki → cd45 + plckb	klckib*cd45plcki
114	cd45plckc → cd45 + plckb	klckcb*cd45plckc
115	cd45 + plcka → cd45plcka	kcd45lckon*cd45*plcka
116	cd45plcka → cd45 + plcka	kcd45lckoff*cd45plcka
117	cd45plcka → cd45 + plckb	klckab*cd45plcka
118	plckb + lckb → plckblckb	klckon*plckb*lckb
119	plckblckb → plckb + lckb	klckoff*plckblckb

Continued on next page

Table B.1 – continued from previous page

N.	**Reaction**	**Rate Expression**
120	plckblckb → plcka + lcka	kautolckb*plckblckb
121	plckb + plckb → plckbplckb	klckon*plckb*plckb
122	plckbplckb → plckb + plckb	klckoff*plckbplckb
123	plckbplckb → plcka + plcka	kautolckb*plckbplckb
124	plcka + lckb → plckalckb	klckon*plcka*lckb
125	lcka + plckb → lckaplckb	klckon*lcka*plckb
126	plcka + plckb → plckaplckb	klckon*plcka*plckb
127	plckalckb → plcka + lckb	klckoff*plckalckb
128	lckaplckb → lcka + plckb	klckoff*lckaplckb
129	plckaplckb → plcka + plckb	klckoff*plckaplckb
130	plckalckb → plcka + lcka	kautolcka*plckalckb
131	lckaplckb → lcka + plcka	kautolcka*lckaplckb
132	plckaplckb → plcka + plcka	kautolcka*plckaplckb
133	plckc + lckb → plckclckb	klckon*plckc*lckb
134	lckc + plckb → lckcplckb	klckon*lckc*plckb
135	plckc + plckb → plckcplckb	klckon*plckc*plckb
136	plckclckb → plckc + lckb	klckoff*plckclckb
137	plckclckb → plckc + lcka	kautolckc*plckclckb
138	lckcplckb → lckc + plckb	klckoff*lckcplckb
139	lckcplckb → lckc + plcka	kautolckc*lckcplckb
140	plckcplckb → plckc + plcka	kautolckc*plckcplckb
141	plckcplckb → plckc + plckb	klckoff*plckcplckb
142	pag + plckb → pagplckb	klckpagon*pag*plckb
143	pag + plckc → pagplckc	klckpagon*pag*plckc
144	pagplckb → pag + plckb	klckpagoff*pagplckb
145	pagplckc → pag + plckc	klckpagoff*pagplckc

Continued on next page

Table B.1 – continued from previous page

N.	**Reaction**	**Rate Expression**
146	pagplckb → pagp + plckb	klckbpagphos*pagplckb
147	pagplckc → pagp + plckc	klckcpagphos*pagplckc
148	cskp + plcka → cskpplcka	kcsklckaon*cskp*plcka
149	cskp + plckb → cskpplckb	klckcskon*cskp*plckb
150	cskpplcka → cskp + plcka	klckcskoff*cskpplcka
151	cskpplckb → cskp + plckb	klckcskoff*cskpplckb
152	cskpplcka → cskp + plckc	klckcskphos*cskpplcka
153	cskpplckb → cskp + plcki	klckcskphos*cskpplckb
154	cskas + plcka → cskasplcka	kcsklckaon*cskas*plcka
155	cskas + plckb → cskasplckb	klckcskon*cskas*plckb
156	cskasplcka → cskas + plcka	klckcskoff*cskasplcka
157	cskasplckb → cskas + plckb	klckcskoff*cskasplckb
158	cskasplcka → cskas + plckc	kcskascat*cskasplcka
159	cskasplckb → cskas + plcki	kcskascat*cskasplckb
160	plckb + cd3 → plckbcd3	klckcd3on*plckb*cd3
161	plckc + cd3 → plckccd3	klckcd3on*plckc*cd3
162	plckbcd3 → plckb + cd3	klckcd3off*plckbcd3
163	plckccd3 → plckc + cd3	klckcd3off*plckccd3
164	plckbcd3 → plckb + cd3p	klckbcat*plckbcd3
165	plckccd3 → plckc + cd3p	klckccat*plckccd3
167	plckb → lckb	kplckadeg*plckb
168	plckc → lckc	kplckadeg*plckc
169	plcka + sh → plckash2	klckshon*plcka*sh
170	plckb + sh → plckbsh2	klckshon*plckb*sh
171	plckc + sh → plckcsh2	klckshon*plckc*sh
172	plckash2 → plcka + sh	klckshoff*plckash2

Continued on next page

Table B.1 – continued from previous page

N.	**Reaction**	**Rate Expression**
173	plckbsh2 → plckb + sh	klckshoff*plckbsh2
174	plckcsh2 → plckc + sh	klckshoff*plckcsh2
175	plckash2 → plcka + shp	klckashp*plckash2
176	plckbsh2 → plckb + shp	klckbshp*plckbsh2
177	plckcsh2 → plckc + shp	klckcshp*plckcsh2
178	plcki → lcki	kplckadeg*plcki

Table B.1: List of Reactions in the Csk Model

N.	Name	Value	Units
1	kcd45lckon	1.0e-03	1/(molecule*second)
2	kcd45lckoff	1	1/second
3	klckib	1.0e-01	1/second
4	klckcb	5.0e-01	1/second
5	klckab	1	1/second
6	klckon	1.0e-03	1/(molecule*second)
7	klckoff	1	1/second
8	kautolckb	1	1/second
9	kcskpagon	1.0e-02	1/(molecule*second)
10	kcskpagoff	1	1/second
11	klckpagon	1.0e-03	1/(molecule*second)
12	klckpagoff	4.0e-01	1/second
13	klckapagphos	5.0e-02	1/second
14	kpagcd45on	1.0e-02	1/(molecule*second)
15	kpagcd45off	1	1/second
16	kpagcd45cat	1	1/second
17	klckbpagphos	1.0e-02	1/second
18	klckcpagphos	2.0e-02	1/second
19	klckcskon	1.0e-01	1/(molecule*second)
20	klckcskoff	1.0e-01	1/second
21	klckcskphos	1.0e-01	1/second
22	kautolcka	10	1/second
23	kautolckc	1	1/second
24	kcddon	1.0e-03	1/(molecule*second)
25	kpp1	1.0e-01	1/second
26	kcddoff	1.0e-02	1/second
27	kcskason	4.0e-02	1/(molecule*second)

Continued on next page

Table B.2 – continued from previous page

N.	Name	Value	Units
28	kcskasoff	1.0e-01	1/second
29	kcskascat	9.20e-01	1/second
30	klckcd3on	1.0e-02	1/(molecule*second)
31	klckcd3off	1.0e-02	1/second
32	klckbcat	1.0e-01	1/second
33	klckacat	2	1/second
34	kcd3dphos	2.50e-02	1/second
35	kcd3deg	1.0e-02	1/second
36	kcd3born	5.0e-01	molecule/second
37	klckccat	4.0e-01	1/second
38	kcsklckaon	1.0e-04	1/(molecule*second)
39	klckshon	1.0e-03	1/(molecule*second)
40	klckshoff	1	1/second
41	klckashp	1	1/second
42	klckbshp	1.0e-01	1/second
43	klckcshp	2.0e-01	1/second
44	kshpdeg	5	1/second
45	kshplckon	5.0e-01	1/(molecule*second)
46	kshplckoff	1	1/second
47	kshplckcat	1.0e-01	1/second
48	kTlckaon	1.0e-02	1/(molecule*second)
49	kTlckaoff	1	1/second
50	kTlckacat	1	1/second
51	kplckadeg	2.0e-01	
52	ksosaRon	2.50e-02	1/(molecule*second)
53	ksosaRoff	3	1/second

Continued on next page

Table B.2 – continued from previous page

N.	Name	Value	Units
54	ksosaTon	2.50e-02	1/(molecule*second)
55	ksosaToff	4.0e-01	1/second
56	ksoscRon	1.40e-04	1/(molecule*second)
57	ksoscRoff	1	1/second
58	ksoscRcat	6.0e-04	1/second
59	ksosaRcRon	1.0e-02	1/(molecule*second)
60	ksosaRcRoff	1	1/second
61	ksosaRcRcat	3.0e-03	1/second
62	ksosaTcRon	1.50e-02	1/(molecule*second)
63	ksosaTcRoff	1.0e-01	1/second
64	ksosaTcRcat	4.0e-02	1/second
65	kTgapon	5.0e-02	1/(molecule*second)
66	kTgapoff	1	1/second
67	kTgapcat	1.0e-01	1/second
68	kcd3psos	4.0e-03	1/second
69	ksosdeg	1.0e-01	1/second
70	kRgrpon	3.20e-01	1/(molecule*second)
71	kRgrpoff	1	1/second
72	kRgrpcat	1.0e-02	1/second

Table B.2: List of Rate Constants in the Csk Model

Name	Initial Amount	Units
CD45	200	molecule
Lck (Total)	1200	molecule
PAG	200	molecule
Csk	500	molecule
CskAS	1000	molecule
PP1	1000 (Introduced)	molecule
CD3	Created Zeroth-Order	molecule
Shp	1000	molecule
Ras	1000	molecule
RasGAP	15	molecule

Table B.3: List of Initial Concentrations in the Csk Model

Appendix C

Supplementary Material for the Cancer Model

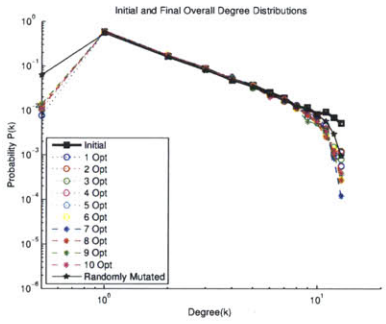


Figure C-1: Coupled Growth Model: Overall Degree distributions

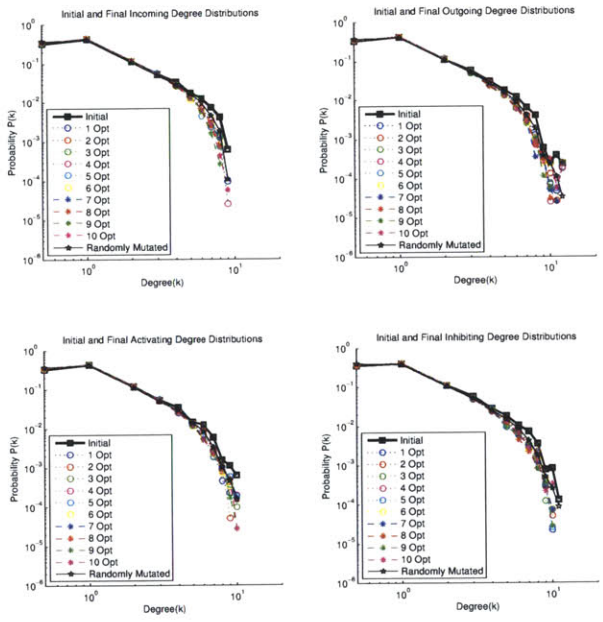


Figure C-2: Coupled Growth Model: all degree distributions

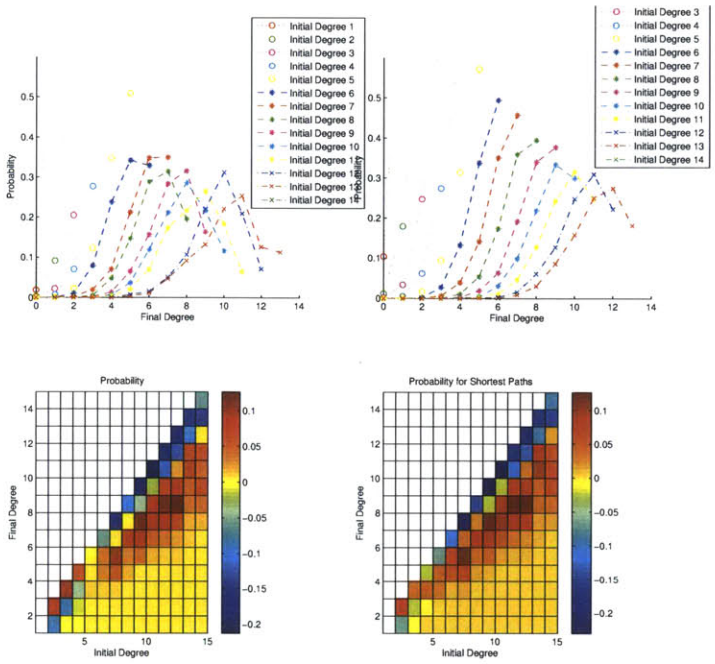


Figure C-3: Coupled Growth Model: Degree Change surface plots

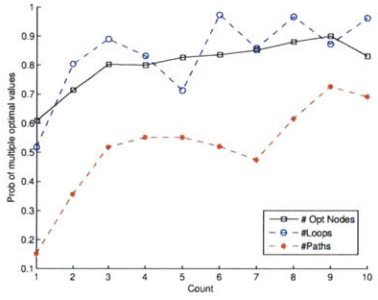


Figure C-4: Coupled Growth Model: Frequency of Multiple optima

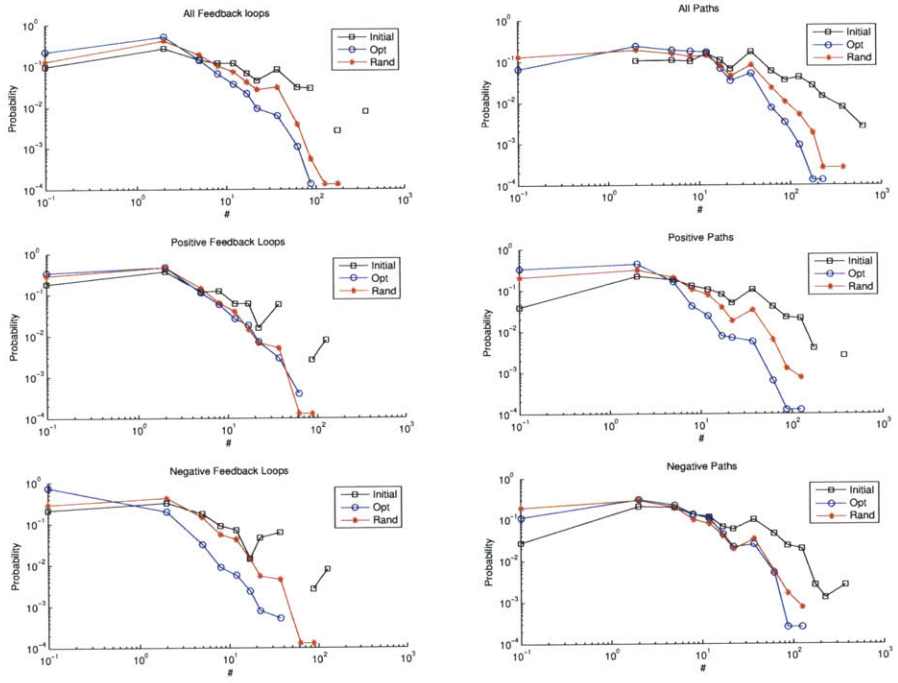


Figure C-5: Coupled Growth Model: Number of feedback loops and paths

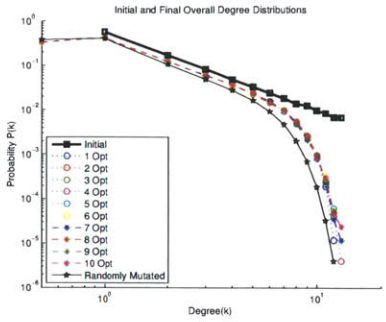


Figure C-6: Potts-Like Model: Overall Degree distributions

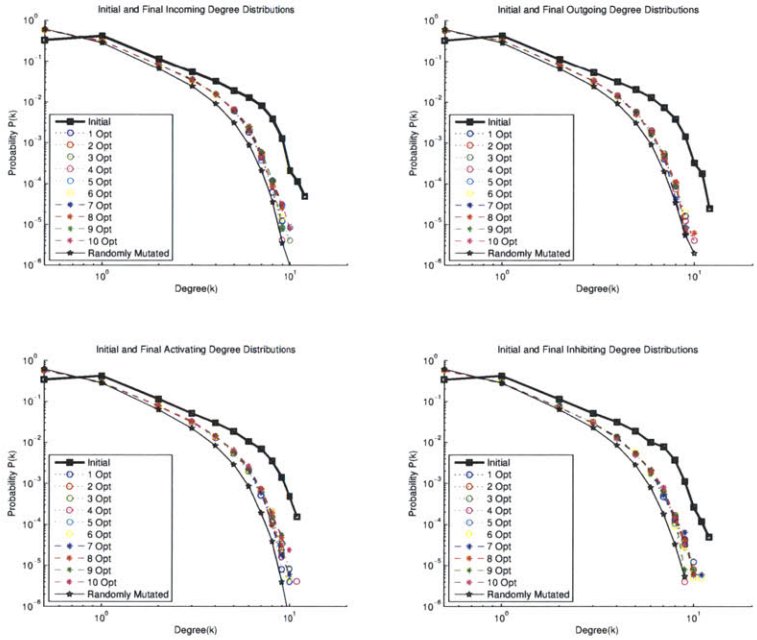


Figure C-7: Potts-Like Model: all degree distributions

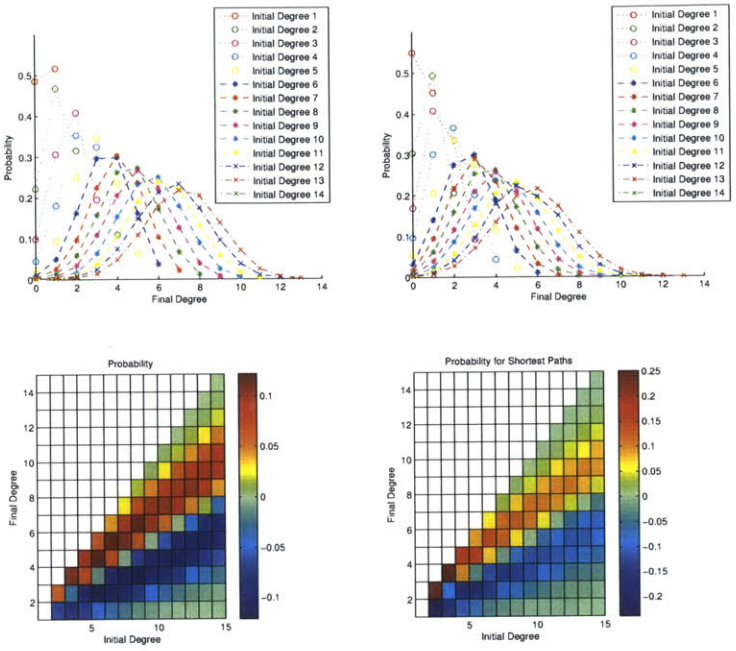


Figure C-8: Potts-Like Model: Degree Change surface plots

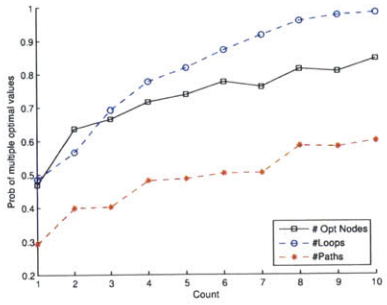


Figure C-9: Potts-Like Model: Frequency of Multiple optima

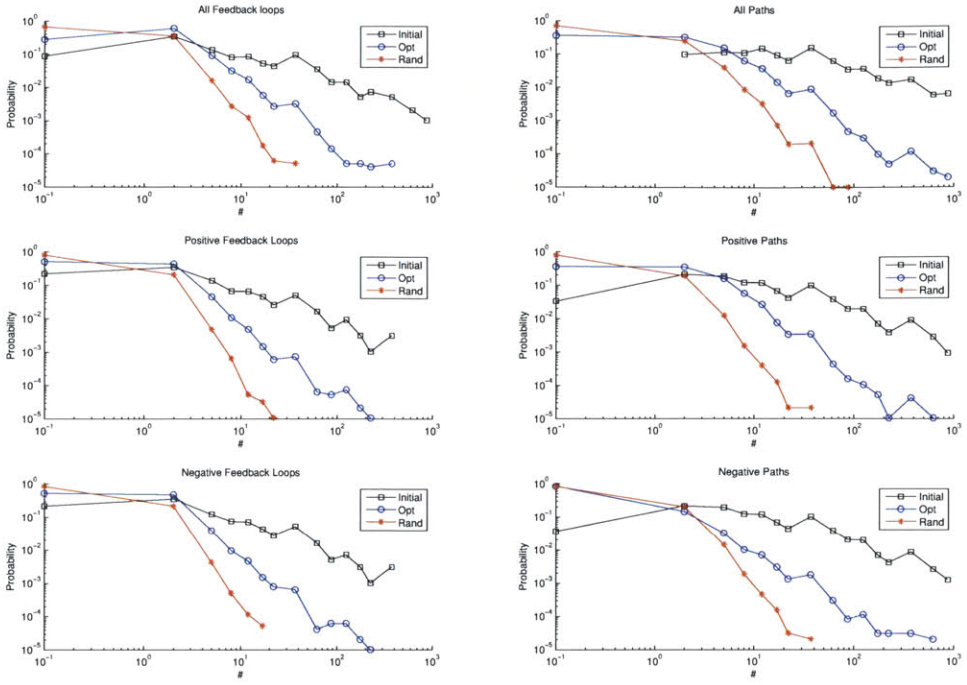


Figure C-10: Potts-Like Model: Number of feedback loops and paths

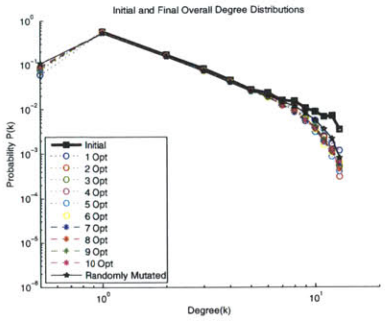


Figure C-11: Correlated Model: Overall Degree distributions

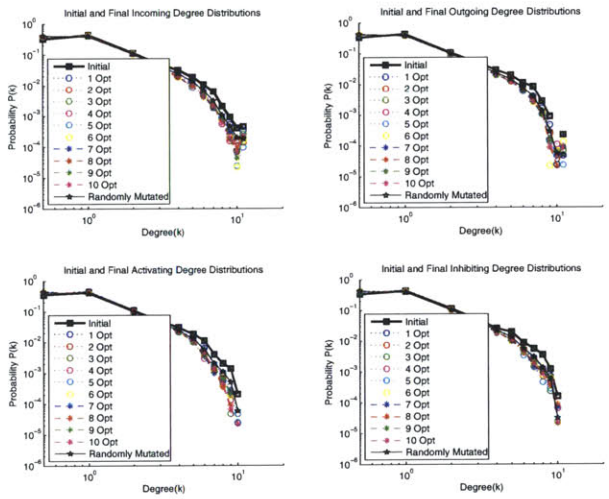


Figure C-12: Correlated model: all degree distributions

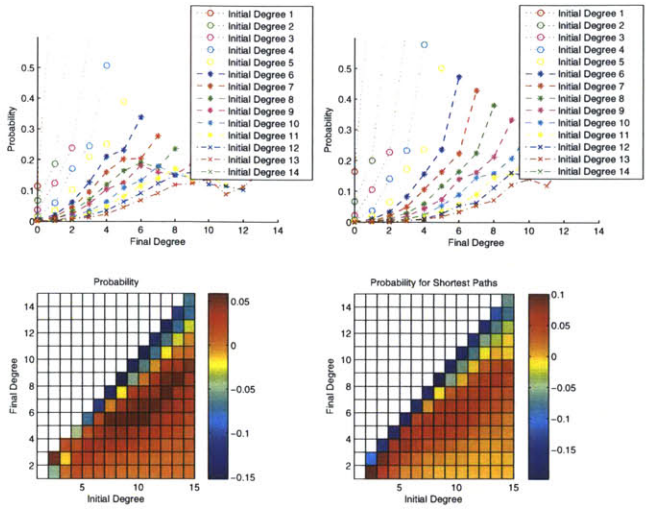


Figure C-13: Correlated model: Degree Change surface plots

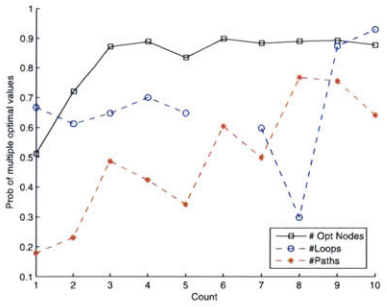


Figure C-14: Correlated model: Frequency of Multiple optima

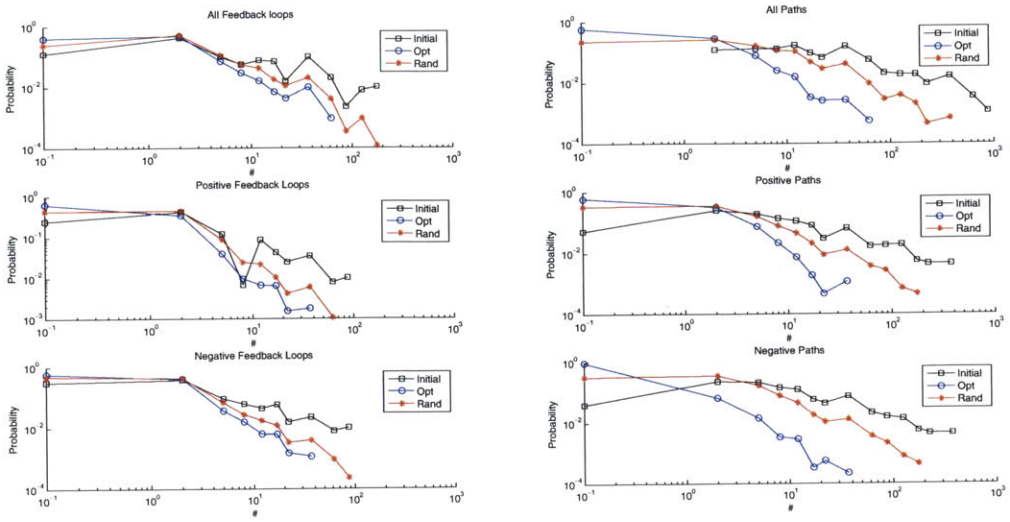


Figure C-15: Correlated model: Number of feedback loops and paths