Stochastic Processes in Biological Systems:
Selected Problems

by

Maksym Artomov

B.Sc. Chemistry, Moscow State University, 2004
M.Sc. Chemistry, University of Chicago, 2005

Submitted to the Department of Chemistry
in partial fulfillment of the requirements for the degree of
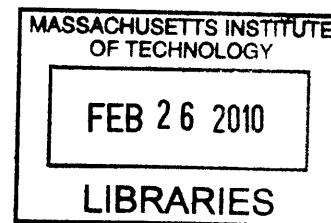
Doctor of Philosophy

at the

Massachusetts Institute of Technology

November 2009

Signature of Author: _____
Department of Chemistry, November 2009

Certified by: _____
Arup K. Chakraborty, Ph.D.
Professor of Biological Engineering
Robert T. Haslam Professor of Chemical Engineering
Professor of Chemistry
Thesis Supervisor

Accepted by: _____
Robert W. Field, Ph.D.
Professor of Chemistry
Chairman, Departmental Committee on Graduate Students

This thesis has been accepted by a committee of the Chemistry Department as follows:

Professor Robert J. Silbey
     Class of 1942 Professor of Chemistry
     Thesis Committee Chairman

Professor Arup K. Chakraborty
     Professor of Biological Engineering,
     Robert T.Haslam Professor of Chemical Engineering,
     Professor of Chemistry
     Thesis Supervisor

Professor Mehran H. Kardar
     Professor of Physics

Stochastic Processes in Biological Systems:

Selected Problems

by

Maksym Artomov

Submitted to the Department of Chemistry
on Dec 20, 2009, in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

Abstract

Majority of biological processes can not be described deterministically. Multple levels of regulation contribute to the noise in the observable properties of the cells: fluctuations are ubiquitous in biological networks and in their spatial organization. In this thesis we consider several examples from three broad categories. Firstly, we study two problems that highlight connection between network topologies and manifestations of stochastic fluctuation in networks of chemical reactions that are meant to represent biological networks in the coarse-grained way. We show that specific network structure can have profound consequences on the steady-state probability distribution function of corresponding chemical system. Secondly, we study effects of spatial organization of the proteins on the membrane surface of T-cells on the initialization of signal propagation. We show that coordinated diffusion of proteins is critical for signal-enhancing properties of co-receptors CD4 and CD8. In third part of the thesis we attempt to reconstruct network topology based on incomplete information about specific interactions between the network nodes and some information about "macroscopic" behavior of the system governed by the network in question. The matter of the Part III, however, is one scale larger than the corresponding objects considered in Part II and I. Specifically, we consider transformations of cells between different cell types and molecular origins that underlie cell transformations (such as differentiation/de-differentiation). Our model suggests specific structure of the master-regulatory network of genes and makes testable predictions.

Thesis Supervisor: Arup K. Chakraborty

Title: Professor of Biological Engineering
        Robert T. Haslam Professor of Chemical Engineering
        Professor of Chemistry

Acknowledgements

This work was only possible due to constant support and encouragement of my beloved wife Yulia. Although completion of thesis writing is a remarkable source of joy, I consider my greatest achievement this year (and my life so far) to be my son Luká coming to the world on Oct 10, 2009.

This dissertation is devoted to Yulia and Luká; they make my life meaningful.

Support of my family throughout these years far away from the Motherland was central for my ability to focus on the scientific problems. My mother Olena, father Nikolai and brother Nikita are filling my heart with the happiness and calm even while being an ocean away.

I would like to thank my Thesis Supervisor Arup K. Charkaborty for two things above all. Firstly, for introducing me into thriving field of biological sciences while maintaining a deep anchor in physical sciences. Secondly, for gathering a group of intelligent and devoted people from whom I learned a great deal of "skill in the trade". Although the list is bound to be incomplete, I would like to give my thanks to Jayajit Das, Bo Jin, Jason Locasale and Abhishek Jha.

My great thanks are for the experimentalists who patiently responded to my often silly questions and, thus, taught me a great deal of understanding about experimental immunology: Herman Eisen, Jennifer Stone, David Kranz, Saso Cemerski.

I would like to take this opportunity to thank people who have affected my scientific thinking in the most vivid way (in chronological order): persistent interactions with Anatoly Kolomeisky allowed me to get infected with his passion for puzzles and for science as a perpetual source of puzzles; the scientific and personal style of Karl Freed will always be one of the most important guide-marks in my professional life; lastly, I would like to join the crowd of the admirers of statistical mechanics class taught by Mehran Kardar at MIT, this class laid the deep appreciation of the way physical sciences should be done into my uninitiated mind.

I am grateful to my friends from Chemistry Department Steve Presse and Maksym Kryvohuz (α-Max): many scientific and not so scientific discussions have created special atmosphere of camaraderie in a "Zoo".

Last but not least, I would like to mention with the deepest warmth my friends from "Chicago cohort", who have now spread all over US (pending the further spread all over the globe): Joji (Hayashida), Mridu (Saikia), Praket (Jha), Abhishek (Jha), thanks for being there!

Table of Content

Chapter 1

Introduction

1.1 Background

Several levels of regulation govern life of biological cells. Signals from the environment are received and processed by networks of proteins that undergo different post-translational modifications, such as phosphorylation, dephosphorylation, ubiquitination *etc*. Appropriately processed signals activate and induce translocation of transcription factors into the nucleus, and promote the action of factors (e.g. chromatin modifying factors) that could cause proliferation and differentiation. The resulting alterations of the transcriptional program of the cell enable it to properly respond to the external stimuli. Often, response involves significant changes in cell appearance, function, and cell numbers.

Enormous numbers of proteins and nucleic acids regulate every aspect of cellular existence. This complexity is accentuated by noise that is present at all levels of regulation. Protein numbers differ from cell to cell and they fluctuate with time in a single cell. Proteins are not distributed homogeneously or in any particular order inside the cell, and are subject to stochastic forces and diffusive transport. Also, often very small numbers of proteins or chemicals are involved in intracellular biochemical reactions. For example, the typical acidity (pH=1) inside the phagosome of macrophages implies that there are only 50 protons present in this closed space. This generates a source of extrinsic noise because, when there are small numbers of reactant molecules, one must account for the intrinsic stochasticity of chemical reactions (McQuarrie, 1967) that is often neglected in classical chemical kinetics.

A biologically relevant situation where stochasticity of biochemical reactions plays a critical role is recognition of ligands by cell surface receptors when stimulatory ligands are limiting. One very important example of this situation is T-cell signaling. T-cells are one of the most numerous cell types in an organism. T-cells represent 1 to 5 % of all cells in humans, which reflects their importance as orchestrators of adaptive the immune response to infectious pathogens that have evaded the defense mechanisms of the innate immune system. Multiple cell surface molecules mediate T-cell interactions with the environment in order to ensure appropriate responses and prevent spurious responses to proteins of the host (which would lead to autoimmunity) (Janeway et al., 2008).

T cells have evolved to combat pathogens that have invaded host cells. Proteins transcribed by these intracellular pathogens (bacteria or viruses) are chopped up in to short peptide fragments. These peptide fragments are then loaded on to proteins coded for by the major histocompatibility (MHC) gene complex. These peptide-MHC complexes are transported to the cell surface, and serve as molecular flags of the pathogen. The most important interactions of a T-cell with its surroundings take place through engagement of the T-cell receptor with these peptide-MHC molecules, which are most prominently displayed on the surface of antigen-presenting cells (typically macrophages, B-cells or dendritic cells). Outcome of such an engagement is critically dependent on the pathogen-derived peptide.

In the healthy state of the organism, only peptides derived from the self-proteins are presented to the T-cells for there are only self-proteins present in the system. Naïve T-cells are generally unresponsive to the self-antigens (due to thymic selection), thus, avoiding autoimmune responses. When organism gets infected with a microbe or virus, non-self proteins start circulating in the system. Pathogen-derived proteins give rise to peptides that are different from the self-derived peptides and T-cell receptors sense this difference which sometime can be as small as single aminoacid substitution (Fig.1.1).

From the point of view of biophysical characterization, interactions between ligand and receptors are described by an association-dissociation reaction. Sufficiently strong binding is known to result in biochemical transformations of a myriad proteins inside the T cell (signaling). These downstream events are represented by a chemical

network, either detailed or coarse-grained. Thus, understanding the influence of topology of chemical networks on the signal processing properties of T cells is an important research frontier. Part I of this thesis is devoted to the study of two coarse-grained models of signaling networks (inspired by T cell biology) where topological effects enable peculiar deviations from the mean-field behavior predicted by classical chemical kinetics.

Ligand quality (i.e. whether it is stimulating or non-stimulating ligand) is typically reflected by the affinity (dissociation constant, Keq) and/or lifetime in the bound state (dissociation rate, koff). In terms of T-cell ligands there is no clear understanding as to which one is the critical parameter distinguishing between the antigen and non-antigen. In our work we are using fixed kon since it has less variablility and vary koff values, thus, making no distinction between Keq and koff.

T-cell sensory apparatus must be able to distinguish ligands with rather small differences in k¬off. Three major membrane proteins play important roles in this discrimination: T-cell receptor (TCR), coreceptor (CD4 or CD8) on the T-cell surface and peptide-MHC complex on the surface of antigen-presenting cell.

At the level of the membrane proximal events, stimulatory ligand sensing should be translated into increased phosphorylation of intracellular part of TCR, which is then used to propagate the signal further. This phosphorylation is carried out by Lck, membrane associated kinase. Lck is present at the inner membrane in two forms: as a free protein or as a coreceptor-bound form. Upon TCR-peptide-MHC engagement, Lck phosphorylates intracellular domains of TCR thus mediating the signal input into the cell. Coreceptor aids the phosphorylation process because it spans the membrane and is capable of binding MHC with its extracellular part while being constitutively associated with Lck inside the T-cell (Fig.1.2). Part II of this thesis presents detailed studies of the earliest signaling events during T-cell activation. A new computational method for efficient simulation of cell signaling processes is described, and some results that may alter the textbook descriptions of coreceptor function are highlighted.

The complexity and robustness of post-translational modifications plays essential role in propagating the signal down to the nuclei where it starts transcriptional programs. The latter can change the cell in the most dramatic ways. For example, it can force the cell to change its identity through differentiation. Although controlled by mechanisms of

colossal size and complexity, cell identity (T-cell, B-cell, red blood cell *etc*) is reasonably stable and well-defined. Yet, it is far from being a deterministic concept. Cell differentiation is viewed as probabilistic event, in that a progenitor can differentiate into progeny 1 with some probability and into progeny 2 with some other probability.

Celular differentiation is usually encountered in the forward direction. For example, an embryonic stem cell differentiates in to a hematopoietic stem which, upon the receipt of appropriate cues, differentiates in to blood cells (e.g., T cells), *etc..* Interestingly, current experiments (Jaenisch and Young, 2008) showed that cells can change their identities in the opposite direction too, albeit with very low probabilities. Although some information is available with respect to genes that maintain cellular identities and mechanisms of transition between the cell states, our knowledge is far from complete at this moment. Part III of this thesis focuses on understanding the general topology of the network of interacting master regulatory genes that is compatible with data on both cellular differentiation and de-differentiation. Specifically, the first theory describing exciting new developments in reprogramming of differentiated cells to a pluripotent state that can be used for patient-specific therapy is presented.

I would like to conclude this introductory chapter by noting that biology offers a broad spectrum of problems ranging from fundamental studies of the stochastic behavior of chemical networks important for signal processing all the way to the application of coarse-grained statistical mechanical models for phenomenological description of cellular behavior. Examples of such problems that span a spectrum of scales and levels of theoretical description have been explored in this thesis.

1.2 Thesis outline

The thesis consists of three major parts which encompass, although not exhaustively, the extent of my work as a graduate student in the Department of Chemistry at MIT.

In Part I we study the appearance of bimodal steady-state probability distribution in two non-typical model networks of chemical reactions. It is customary to observe bimodal steady-state probability distribution in the situations when underlying system has two stable fixed points. In these settings, and in the presence of noise, transitions between

stable states occur, and complete probability distribution has two peaks (or more, in the case of multistability). In Chapters 2 and 3 we have considered situations when the systems in question do not have any underlying multistabilities and, thus, are not expected to exhibit multipeaked probability distribution. We show that in the regime where intrinsic stochasticity of chemical networks must be taken into account bimodal distribution can emerge as a consequence of combination of stochastic fluctuations in numbers of particles and particular network topologies.

We have studied conditions on network topologies and kinetic parameters of the networks that enable this effect in two classes of systems. Firstly, Chapter 2 illustrates how a strong feedback loop can lead to the appearance of bimodal distribution in *irreversible* systems. Secondly, Chapter 3 reports a novel mechanism, termed network topology reduction, that leads to appearance of bimodality in *monostable reversible* systems. In fact, biologically relevant systems discussed in the literature (Miller and Beard, 2008; Samoilov et al., 2005a) exhibit bimodality due to network topology reduction, which was not recognized previously.

Although material of Part I was motivated by biological systems, Chapters 2 and 3 do not address directly important biological questions. They consider general properties of networks emerging in the biological applications. Part II, however, focuses on biological problem where underlying stochasticity plays critical role.

T-cell signaling networks are the main subject of interest in Part II. Early signaling events that take place in the membrane proximal region during the T-cell activation serve as unique filtering module discriminating between noise and signal. By natural design of the human (or mouse, as these are main objects of study in clinical immunology) body, T-cell receptors are constantly interacting with multiple non-stimulating ligands derived from self-proteins of the organism (i.e. self-peptides presented in context of MHC complexes). Yet T-cell receptors have to be able to deliver activation signal in rare cases when pathogenic-derived, stimulating ligand is presented to them. T-cell machinery can distinguish these two classes of ligands, although they are often very similar, sometimes differing by a single aminoacid substitution.

This remarkable resolution of T-cell sensory apparatus leads to the following general question: what are the principles guiding the topology of signaling networks in T-

cell that allow it to achieve such discriminatory capabilities? In this thesis we only consider a very limited part of T-cell signaling pathway which, nonetheless, amounts to approximately 1,000 reaction chemical network. (Even though the number of interacting proteins that we consider is less than ten, see below about the combinatorial expansion.) The general idea that we adhere to in Part II is to deduce the topology of the molecular network based on "microscopic" information about protein-protein interactions and "macroscopic" information about T-cell response to different perturbations.

We face two major obstacles when studying realistic problems in signal propagation. The most commonly recognized one is combinatorial expansion of number of reactions and reactants in the typical biological systems. Unlike ordinary chemicals, proteins that participate in signaling network do not loose their identity as a result of reaction. Rather, they are modified in some way; for example, phosphorylated, dephosphorylated, ubiquitinated *etc*. The origin of the combinatorial expansion problem lies in the multi-domain structure of the proteins. One recognizes that the same protein can have multiple modification states. Thus, presense of just two phosphorylation sites implies that there are 4 different states this protein can be at. In the very direct manner this also affects the number of chemical reactions in the network under consideration: one has to consider four explicit binding reaction if two reacting proteins have one phosphorylation site each. This problem has been recognized previously and several solutions have been proposed, most notably based on rule-based modeling or κ-calculus.(Danos and Laneve, 2004; Faeder et al., 2009).

Second major problem has to do with spatial organization of the proteins participating in the signaling network. Although this aspect is often neglected when considering intracellular signaling cascades that take place in cytoplasm, it is critical to consider interplay between the interactions and spatial motion of proteins when action the interactions take place in two-dimensions, e.g. in cell membrane. The reaction-diffusion view is especially appropriate in the problem of T-cell membrane-proximal signaling because timescales of chemical reactions and diffusion processes of proteins in the membrane are very similar, as will become apparent in Chapter 5. However, by the year 2009, there was only one simulation software that allow spatially resolved simulation of

stochastic chemical networks (MesoRD) that was created by Johan Elf and collaborators (Hattne et al., 2005).

However, there was no computational solution that would provide the capabilities to address both of the aforementioned problems. We were able to address this question through the collaboration with Mieszko Lis, graduate student at MIT Computer Science Department. He created the software (Lis et al., 2009), named SSC for Stochastic Simulation Compiler, that combined the rule-based approach to solving the problem of combinatorial expansion of chemical network and next-subvolume algorithm (Hattne et al., 2005) for simulating spatial motion of particles. Appendix to Chapter 4, which is due to Mieszko Lis and is not a part of my thesis, describes the details of the SSC software and provides the basic examples of its use. Chapters 5 and 6 make extensive use of SSC software and, thus, I feel that inclusion of rather large Appendix to Chapter 4 is justified.

Having clarified our simulation methods in Chapter 4 we move on to consider the question of the coreceptor involvement in T-cell early signaling in Chapter 5. We find that observed experimental characteristics of coreceptor-MHC interactions are inconsistent with one of the commonly accepted (even in the textbooks) mechanisms of coreceptor-mediated signal enhancement. By carrying out explicit molecular simulations with the help of SSC software we were able to accurately describe the origins of the positive effect of coreceptor involvement in signal initiation.

Next chapter, Chapter 6, deals with fascinating phenomenon of synergy between two classes of TCR ligands. It was found relatively recently (Krogsgaard et al., 2007a) that peptides derived from self-proteins can enhance the signal that originates from stimulation with pathogenic peptides (Fig.1.3), even though self-peptides do not deliver any activation signal by themselves. Although conceptual models treating involvement of self-peptides have been already published (Li et al., 2004b; Wylie et al., 2007b), coherent description that would employ all of the available biophysical information in a consistent manner has been lacking. In Chapter 6 we describe three explicit molecular models that are compatible with current biological knowledge about protein-protein interactions and corresponding biophysical parameters. This explicit description allows us to address an important debate about origins of difference of self-peptide involvement in CD4 and CD8 systems (Ebert et al., 2009; Lo et al., 2009; Yachi et al., 2007). All the different models

that we consider point to the importance of the positive selection threshold for identification of range of co-enhancing self-peptides. Our conclusions are corroborated by recently published results in CD4 systems where only positively selecting peptides were capable of synergizing the antigen-derived signal (Ebert et al., 2009; Lo et al., 2009; Yachi et al., 2007).

Part III of this thesis is in line with general philosophy of Part II. Here too, we attempt to reconstruct network topology based on incomplete information about specific interactions between the network nodes and some information about "macroscopic" behavior of the system governed by the network in question. The matter of the Part III, however, is one scale larger than the corresponding objects considered in Part II. Specifically, we consider transformations of cells between different cell types and molecular origins that underlie cell transformations (such as differentiation/de-differentiation).

Cell differentiation is ubiquitous phenomenon when different kinds of cells are arising upon division of the older cells. One of the classical examples of differentiation is formation of an organism starting from a single fertilized egg. It is commonly recognized that all cells in an organism have the same DNA (in fact, only majority of cells have the same DNA). Yet, the cells often appear as differently as red blood cells and T-cells and skin cells. They express different proteins and carry out different functions. This is because of epigenetic differences; i.e., DNA in different cell types is packaged distinctly, making it hard to express certain genes while facilitating the expression of others. This additional above-(epi)-genetic level of regulations insures that diverse cell types can arise based on the exact same DNA sequence. During development, upon receipt of appropriate cues, pluripotent embryonic stem cells differentiate in to diverse cell types that make up the organism (e.g., a human). There has long been an effort to make this process go backward – i.e., reprogram a differentiated cell (e.g., a skin cell) to pluripotent status. Recently, this has been achieved by transfecting certain transcription factors in to differentiated cells (Jaenisch and Young, 2008). This method does not use embryonic material and promises the development of patient-specific regenerative medicine, but it is inefficient. The mechanisms that make reprogramming rare, or even possible, are poorly understood. In Chapter 7 we report the first computational model of transcription factor-

induced reprogramming. Results obtained from the model are consistent with diverse observations, and identify the rare pathways that allow reprogramming to occur. If validated by further experiments, our model could be further developed to design optimal strategies for reprogramming and shed light on basic questions in biology.

1.3 References to published work and work outside the thesis scope

The work presented in Chapter 2 has been published in Proceedings of the National Academy of Sciences (Artyomov et al., 2007a) and was featured in Physics Today (Fleming and Ratner, 2008). The work described in Chapter 3 is currently in press at Journal of Chemical Physics. Collaborative work on creation of SSC software that is described in Chapter 4 has been published in BioInformatics (Lis et al., 2009). The materials of Chapters 5 and 6 are in the final stages of preparation for submission. Chapter 7 is currently under review in the PLoS Computational Biology.

Although the Chapters of the thesis correspond to the most important research projects of my PhD career, there have been many others that came along due to discussions with experimental and theoretical scientists.

I was a part of three major experimental collaborative efforts during my PhD career: the most current collaboration with Michel Nussenzweig lab at Rockefeller University; collaboration with David Kranz and Jennifer Stone at UIUC; and collaboration with Uli von Andrian lab from Harvard Medical School (HMS). In the first (Rockefeller) project we were lucky to have contributed to understanding the phenomenon of heteroligation of antibodies on the viral particles (one paper submitted to Nature, and the other in final stages of preparation for submission to Journal of Immunological Methods). As a result of this collaboration two manuscripts are now in preparation. In the second (UIUC) project, we had a fascinating opportunity to dwell deeper into the details and caveats of famous MHC tetramer assay. I believe that our findings will improve the level of understanding of this widespread experimental tool. The manuscript that reports the details of our work is currently in preparation. In the last (HMS) project we were able to reconcile dynamics of peptides dissociation off the MHC complexes of dendritic cells in vivo and in vitro, which appeared to be inconsistent on a first glance. This work has been published in Nature Immunology (Henrickson et al.,

2008). In all of these collaborative projects, I want to be very clear about it, the most difficult scientific work was done by our experimental colleagues; but I would like to believe that we, too, contributed critical pieces of understanding.

During my MIT years I have been fortunate to continue interactions with Professor Anatoly Kolomeisky from Rice University. Multiple discussions with him have yielded interesting and analytically tractable questions in the theory of molecular motors. We have addressed these questions in several publications (Artyomov, 2009; Artyomov et al., 2008; Artyomov et al., 2007b; Morozov et al., 2007) with some containing exact analytical results for burnt-bridge model of motor proteins (Artyomov et al., 2007b).

Please note that I use different spelling of my name in my publications. This originates from the conflict between direct official transliteration of my family name from Cyrillic to Latin alphabet (Maksym Artomov) versus the spelling that most appropriately corresponds to phonetic pronunciation of my last name (Maxim N. Artyomov).

1.4 Bibliography

Artyomov, M.N. (2009). Comment on "reciprocal relations for nonlinear coupled transport". Phys Rev Lett *102*, 149701; discussion 149702.

Artyomov, M.N., Das, J., Kardar, M., and Chakraborty, A.K. (2007a). Purely stochastic binary decisions in cell signaling models without underlying deterministic bistabilities. Proceedings of the National Academy of Sciences of the United States of America *104*, 18958-18963.

Artyomov, M.N., Morozov, A.Y., and Kolomeisky, A.B. (2008). Molecular motors interacting with their own tracks. Physical review *77*, 040901.

Artyomov, M.N., Morozov, A.Y., Pronina, E., and Kolomeisky, A.B. (2007b). Dynamic properties of molecular motors in burnt-bridge models. Journal of Statistical Mechanics: Theory and Experiment *2007*, P08002-P08002.

Danos, V., and Laneve, C. (2004). Formal molecular biology. Theoretical Computer Science *325*, 69-110.

Ebert, P.J., Jiang, S., Xie, J., Li, Q.J., and Davis, M.M. (2009). An endogenous positively selecting peptide enhances mature T cell responses and becomes an autoantigen in the absence of microRNA miR-181a. Nat Immunol *10*, 1162-1169.

Faeder, J.R., Blinov, M.L., and Hlavacek, W.S. (2009). Rule-Based Modeling of Biochemical Systems with BioNetGen. Methods in Molecular Biology, 113-167.

Fleming, G.R., and Ratner, M.A. (2008). Grand challenges in basic energy sciences. Physics Today *61*, 28-33.

Hattne, J., Fange, D., and Elf, J. (2005). Stochastic reaction-diffusion simulation with MesoRD. Bioinformatics (Oxford, England) *21*, 2923-2924.

Henrickson, S.E., Mempel, T.R., Mazo, I.B., Liu, B., Artyomov, M.N., Zheng, H., Peixoto, A., Flynn, M.P., Senman, B., Junt, T., *et al.* (2008). T cell sensing of antigen dose governs interactive behavior with dendritic cells and sets a threshold for T cell activation. Nature Immunology *9*, 282-291.

Jaenisch, R., and Young, R. (2008). Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell *132*, 567-582.

Janeway, C., Murphy, K.P., Travers, P., Walport, M., and Janeway, C. (2008). Janeway's immuno biology (New York: Garland Science).

Krogsgaard, M., Juang, J., and Davis, M.M. (2007). A role for "self" in T-cell activation. Seminars in immunology *19*, 236-244.

Li, Q.J., Dinner, A.R., Qi, S.Y., Irvine, D.J., Huppa, J.B., Davis, M.M., and Chakraborty, A.K. (2004). CD4 enhances T cell sensitivity to antigen by coordinating Lck accumulation at the immunological synapse. Nature Immunology *5*, 791-799.

Lis, M., Artyomov, M.N., Devadas, S., and Chakraborty, A.K. (2009). Efficient stochastic simulation of reaction-diffusion processes via direct compilation. Bioinformatics (Oxford, England) *25*, 2289-2291.

Lo, W.L., Felix, N.J., Walters, J.J., Rohrs, H., Gross, M.L., and Allen, P.M. (2009). An endogenous peptide positively selects and augments the activation and survival of peripheral CD4+ T cells. Nat Immunol *10*, 1155-1161.

McQuarrie, D.A. (1967). Stochastic Approach to Chemical Kinetics. Journal of Applied Probability *4*, 413-478.

Miller, C.A., and Beard, D.A. (2008). The effects of reversibility and noise on stochastic phosphorylation cycles and cascades. Biophysical Journal *95*, 2183-2192.

Morozov, A.Y., Pronina, E., Kolomeisky, A.B., and Artyomov, M.N. (2007). Solutions of burnt-bridge models for molecular motor transport. Physical review *75*, 031910.

Samoilov, M., Plyasunov, S., and Arkin, A.P. (2005). Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations. Proceedings of the National Academy of Sciences of the United States of America *102*, 2310-2315.

Wylie, D.C., Das, J., and Chakraborty, A.K. (2007). Sensitivity of T cells to antigen and antagonism emerges from differential regulation of the same molecular signaling module. Proceedings of the National Academy of Sciences of the United States of America *104*, 5533-5538.

Yachi, P.P., Lotz, C., Ampudia, J., and Gascoigne, N.R. (2007). T cell activation enhancement by endogenous pMHC acts for both weak and strong agonists but varies with differentiation state. The Journal of experimental medicine *204*, 2747-2757.

1.5 Figures for Chapter 1

Foreign Microb

APC processing

Fragments of proteins (8-14 aa)
bound to MHC protein on the
APC surface
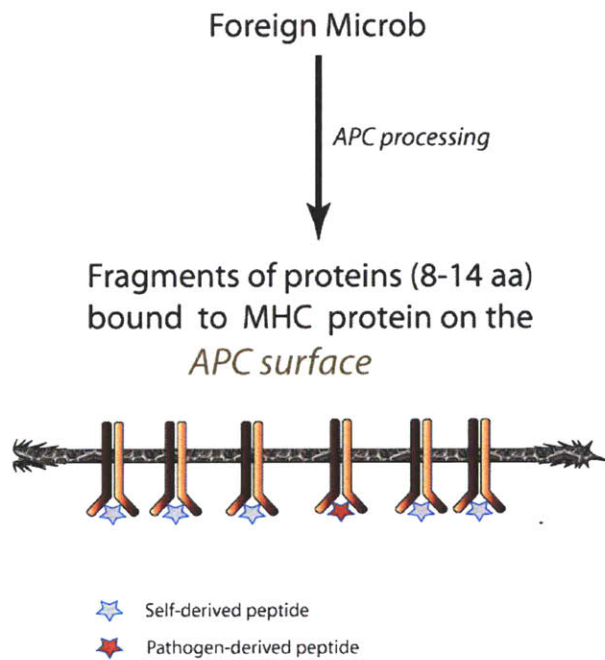
Self-derived peptide

Pathogen-derived peptide

Fig.1.1. Schematic representation of MHC loading with foreign derived peptide. After MHC has been loaded with peptide, it is targeted to the cell surface where it presents peptide to the T-cell.
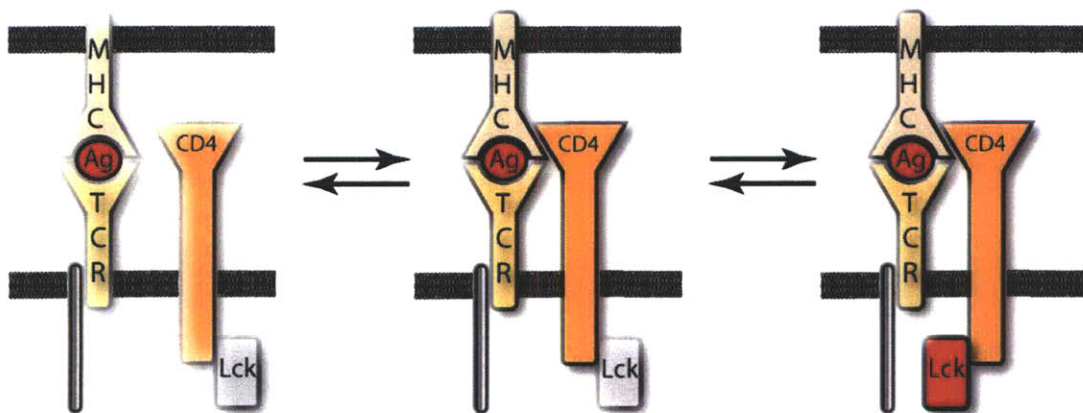
Fig.1.2. Schematic representation of coreceptor (CD4 in this case) involvement into TCR-pepMHC interactions.
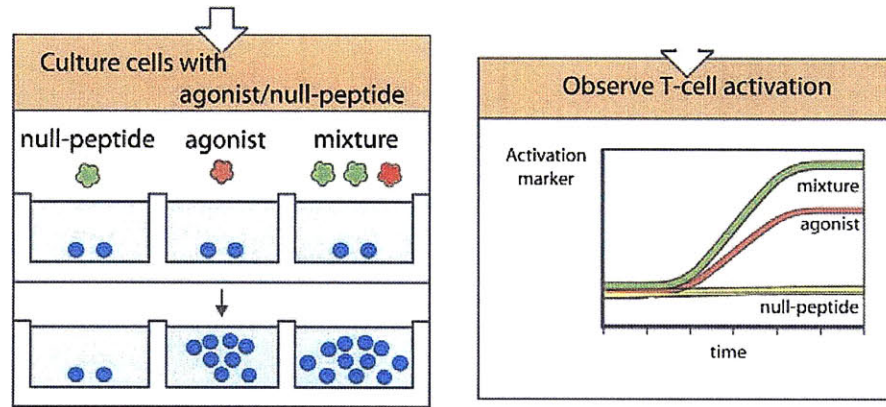
Fig.1.3. Biological manifestation of the self-peptide enhancement phenomenon. Self-peptides (designated null-peptides on the picture) do not stimulate T-cell activation and proliferation (yellow curve on the right panel). However, when mixed with stimulatory peptides (designated agonist on the picture), it provokes higher degree of activation as measured by proliferation or, for instance, cytokine production.

PART I

Chapter 2

Purely stochastic binary decisions in cell signaling models without underlying deterministic bistabilities

2.1 Introduction

The detection of external stimuli by receptors on a cell membrane followed by intracellular signaling, gene transcription, and effector functions is ubiquitous, and necessary for life. The regulatory processes involved in gene transcription are often mediated by small numbers of molecules. This makes stochastic effects important and, in recent years, many interesting consequences of such fluctuations have been elucidated theoretically and observed in experiments (e.g., (Acar et al., 2005; Elowitz et al., 2002; McAdams and Arkin, 1997; Weinberger et al., 2005)). The importance of stochastic effects on enzymatic reactions in the zero order ultrasensitivity regime has also been described (Berg et al., 2000; Samoilov et al., 2005a). Less attention has been devoted to the effects of stochastic fluctuations on cell signaling dynamics. Yet, many such processes involve small numbers of molecules. One important example is provided by T lymphocytes (T cells), the orchestrators of the adaptive immune response. T cell signaling and activation can be stimulated by as few as 3 molecules that represent signatures of pathogens (called agonists) (Brower et al., 1994; Davis et al., 2007; Irvine et al., 2002; Li et al., 2004a; Purbhoo et al., 2004; Sykulev et al., 1996). The small numbers of molecules involved can make stochastic effects important for membrane-proximal

signaling in T cells. Here, we study simple and general models inspired by recent descriptions of membrane-proximal signaling in T cells, and find an interesting consequence of stochastic fluctuations. An essential feature of the model, dueling positive and negative feedback loops, is ubiquitous, and so our findings may be of broad relevance in cell biology.

Many examples (particularly models of gene regulation) have been studied previously wherein a deterministic treatment of the kinetic scheme describing the relevant processes has two stable steady states in a certain parameter regime (Acar et al., 2005; Elowitz et al., 2002; Karmakar and Bose, 2007; Kepler and Elston, 2001b; McAdams and Arkin, 1997). In such systems, stochastic effects can lead to bimodality (e.g., populated "on" and "off" states) in the parameter range where bistability is predicted by the deterministic equations as well as outside this range where there is a single stable steady state (Acar et al., 2005; Elowitz et al., 2002; Karmakar and Bose, 2007; Kepler and Elston, 2001b; McAdams and Arkin, 1997). The latter phenomenon is a consequence of stochastic fluctuations enabling the system to sample parameters (e.g., rate constants) that effectively fall within the range where two deterministically stable fixed points are present. In these examples, the existence of bistability in the deterministic analysis in some parameter range underlies the observation of bimodal behavior in the stochastic treatment.

The model we study exhibits a different feature. The deterministic dynamical equations yield a single steady state in *all* parameter ranges; i.e., there is no bistability. Yet, stochastic fluctuations result in a bimodal long-time response with neither mode corresponding to the steady state obtained deterministically. Upon increasing the copy numbers of molecules, the stochastic description ultimately converges to the deterministic behavior. Thus, we find a purely stochastically driven instability when none exists in the deterministic treatment in any parameter range. When fluctuations are important, we find that average quantities scale with parameters "anomalously" compared to the corresponding mean-field behavior. Our analyses suggest that the necessary and sufficient conditions for this phenomenon to occur are quite common.

2.2 Signaling model

Our simple ("toy") model is inspired by ideas proposed recently to describe T cell responses to diverse stimuli (Altan-Bonnet and Germain, 2005; Davis et al., 2007; Irvine et al., 2002; Li et al., 2004a; Stefanova et al., 2003; Wylie et al., 2007a). T cell receptor (TCR) molecules expressed on the surface of T cells can bind complexes of peptides (p) bound to major histocompatibility (MHC) proteins on the surface of antigen presenting cells (APCs). TCR can potentially bind strongly to pMHC molecules where the peptide is derived from a pathogen's proteins (agonists). In contrast, thymic selection ensures that TCR bind weakly to "self" or endogenous pMHC molecules that are also expressed on APCs (Starr et al., 2003). The binding of TCRs to pMHC molecules can initiate signaling cascades that result in T cell activation and an immune response. T cells are as good a sensory apparatus as any in biology, and can detect as few as three agonists in a sea of tens of thousands of endogenous pMHC molecules, and it has been suggested that this extraordinary sensitivity is mediated by cooperative interactions between self pMHC and agonists(Davis et al., 2007; Irvine et al., 2002; Li et al., 2004a; Purbhoo et al., 2004; Yachi et al., 2005b).

Another interesting response of T cells to pMHC molecules is called antagonism (Evavold et al., 1994; Stefanova et al., 2003). Antagonists are pMHC molecules obtained by mutating agonist peptide residues. When present on APC surfaces in sufficient numbers, they can shut down intracellular signaling stimulated in response to agonists. Recent experimental results (Stefanova et al., 2003) have suggested that this phenomenon may be mediated by dueling positive and negative feedback loops (Fig. 2.1). One of the earliest steps in downstream signaling initiated by the binding of the TCR to pMHC molecules is the phosphorylation of cytoplasmic domains of the TCR complex by a kinase called Lck. It has been proposed that Lck also activates its own inhibitor, a phosphatase called Shp (negative feedback). This inhibitory interaction is prevented by a product (ERK) of signaling downstream of phosphorylation of the TCR complex that protects Lck by phosphorylating one of its sites (positive feedback). It has been proposed, and detailed calculations support this (Altan-Bonnet and Germain, 2005; Wylie et al., 2007a), that the positive feedback is dominant when T cells are stimulated by

agonists (and synergistic endogenous ligands), and negative feedback shuts down signaling when sufficient numbers of antagonists are present.

While the specific molecular identity of positive and negative regulators involved in T cell signaling is still debated (Li et al., 2007), the idea that dueling positive and negative feedback loops play a role in determining whether signaling is shut off (antagonism) or sustained/amplified (agonism) is of general significance to cellular decisions that lead to distinct outcomes. Furthermore, such processes are often mediated by small numbers of molecules. Therefore, we set out to study the effects of stochastic fluctuations on the following simple and general model with dueling positive and negative feedback regulation:

$$A_1 \xrightarrow{k_1} E + A_1 \ (1); \qquad\qquad A_2 \xrightarrow{k_2} S + A_2 \qquad (2)$$

$$E + A_1 \xrightarrow{k_3} E + A_1^{PROT} \ (3); \qquad\qquad A_1^{PROT} \xrightarrow{k_4} E + A_1^{PROT} \qquad (4)$$

$$S + A_1 \xrightarrow{k_5} S + A_1^{INACTIV} \quad (5); \qquad\qquad S \xrightarrow{k_D} \phi, \quad E \xrightarrow{k_D} \phi \qquad (6)$$

While this model is general, seeing how it relates to T cell signaling makes clear that it is relevant to situations where cells make distinct decisions (e.g., agonism and antagonism in Fig. 2.1). The first reaction mimics the production of the positive regulator ERK (E) upon agonist (A_1) binding to TCR. Thus, it subsumes a large number of steps in the actual signaling cascade into one. Of course, agonists also lead to production of the negative regulator Shp (S), but this is ignored in this general model. Similarly, some production of E by antagonist (A_2) binding to the receptor is ignored, and reaction 2 mimics the production of the negative regulator. Reaction 3 represents positive feedback and mimics protection of Lck from the action of Shp, in that the interaction of E with A_1 protects it (by forming A_1^{PROT}) from the inhibitory action of S (reaction 5). Protected A_1 species can generate positive regulators E (reaction 4), and both positive and negative regulators can be inactivated (reaction 6).

2.3 Results

The mean-field deterministic equations corresponding to the model described by Eqs. 1-6 can be written down following mass action kinetics (web supplement), and yield the following solution for the steady state:

$$A_1^{(SS)} = 0, A_1^{PROT\,(SS)} + A_{1,INACTIV}^{(SS)} = A_{1,initial}, \quad E^{(SS)} = \frac{k_4}{k_D} A_1^{PROT\,(SS)}, \quad S^{(SS)} = \frac{k_2}{k_D} A_2 \qquad (7)$$

At steady state, the number of A_1 molecules equals zero, the number of S molecules is a function of the number of A_2 molecules, the number of E molecules depends upon the number of protected A_1 molecules, and all solutions which satisfy the constraint that the sum of the number of A_1^{PROT} species and A_1^{INACT} species sum to the initial number of A_1 are allowed. Thus, unique steady states cannot be obtained from Eqs. 7 without knowledge of the initial conditions. Rather, there is a line of possible steady states. Stability analysis shows that all, but one, eigenvalues of the Jacobian matrix are negative. The only non-negative eigenvalue is zero, and corresponds to sliding along the line of possible steady states, $A_1^{PROT\,(SS)} + A_{1,INACTIV}^{(SS)} = A_{1,initial}$, with corresponding change in the steady-state value of E. Solving the dynamical equations with specific initial conditions and taking the long-time limit obtains a unique point on this fixed line. Thus, the deterministic solutions of the model are a set of unique steady-states for all parameter values.

While we have studied different parameter ranges for a stochastic description of this model (web supplement), let us first consider situations that are inspired by T cell signaling. Reactions (1), (2), and (4) represent multi-step processes (Lin and Weiss, 2001). Reactions (3) and (5), the dueling feedback loops, are thought to represent one step phosphorylation or deactivation steps (Stefanova et al., 2003). So, we study situations where k_3 and k_5 are much larger than k_1, k_2 and k_4; i.e., both positive and negative feedback loops are strong. Recent studies (Li et al., 2004a; Wylie et al., 2007a) with detailed models of membrane-proximal signaling in T cells suggests that k_4 could be larger than k_1, but we have taken them to be equal ($k_4 > k_1$ is considered in the web supplement). Changing the relative values of k_1 and k_2 would simply modify the specific value of the ratio of initial numbers of A_1 and A_2 molecules that would result in a transition from "agonism" to "antagonism".

Fig. 2.2 shows results of spatially homogeneous stochastic simulations with discrete number of molecules (using the Gillespie algorithm (Gillespie, 1977)) of the model represented by Eqs. 1-6. When there are only a few molecules of A_1 and A_2, essentially all the stochastic trajectories commit to one of two final states: all the A_1

molecules are converted to the protected species, A_1^{PROT}, or are annihilated and signaling stops. This bimodality is in striking contrast to the mean-field solution that does not exhibit bistability for any parameter values. The qualitative phenomenon of finding a bimodal stochastic solution when the deterministic solution is unique for all parameter values is preserved as long as the positive and negative feedback loops are sufficiently strong (web supplement).

The mechanism underlying this result is as follows. The species A_1 is converted to either A_1^{PROT} or A_1^{INACT}. The effective rates of production of these species can be obtained from the deterministic equations. Both rates equal zero initially and at long times, and exhibit a maximum (Fig. S2.4). The initial rise and amplitudes of the maxima depend upon the values of the initial number of A_1 and A_2 molecules, and are very different if one of these quantities is much larger than the other. In these circumstances, either agonism or antagonism dominates in the deterministic and stochastic solutions. The more interesting cases are ones where the generation of positive and negative regulations is roughly balanced (Fig. 2.2) as it could result in a transition from agonism to antagonism. Now, the rates at short times and amplitudes of the maxima for the production of A_1^{PROT} and A_1^{INACT} are comparable in the mean-field sense, and the deterministic equations yield a single steady state solution with an intermediate value of A_1^{PROT}. However, stochastically, one of two reactions (1) and (2) occurs first. There is a stochastic delay, τ, before the other reaction occurs, and for this duration, the reaction propensities are effectively as in cases where $A_1 >> A_2$, or *vice versa*. For small numbers of A_1 and A_2 molecules, τ can be long. If τ is longer than the intrinsic time scale associated with the feedback reaction corresponding to the reaction that occurred first (e.g., reaction (3) if (1) occurred first), then the small number of A_1 molecules will all be converted to either A_1^{PROT} or be annihilated, depending upon whether reaction (1) or (2) occurred first. So, the stochastic trajectories partition into two classes (those that end with all A_1 molecules annihilated or protected), and the stochastic solution is bimodal.

The time delay (τ) becomes smaller as the number of molecules of A_1 and A_2 increases. This suggests that, for a sufficiently large number of particles, it will not be longer than the intrinsic time scale associated with the feedback loops and the stochastic solution will not be bimodal. Rather, it will be distributed around the mean-field solution.

Fig. 2.3 shows results of simulations that demonstrate this unequivocally. Thus, for the same parameter values, as the number of molecules decreases past a threshold, the stochastic solution exhibits an instability from one solution to bimodality. *This transition from unimodal to bimodal solutions is driven by stochastic effects, and occurs in the absence of any underlying deterministic bistability.*

The qualitative differences between the stochastic and deterministic descriptions due to the dominance of fluctuation effects suggests that the manner in which the response scales with different control parameters may be different. For example, we expect the steady state amount of A_1^{PROT} to scale with $\dfrac{k_1 A_1}{k_2 A_2}$ for the stochastic simulations. This is because the probability of conversion to A_1^{PROT} is essentially equal to the probability that reaction (1) occurs first, which is given by $\dfrac{k_1 A_1}{k_1 A_1 + k_2 A_2}$. Conversely, probability of annihilating all A_1 molecules is equal to $\dfrac{k_2 A_2}{k_1 A_1 + k_2 A_2}$ (equal to probability that reaction (2) occurs first). Both expressions depend only on the combination $\dfrac{k_1 A_1}{k_2 A_2}$.

This implies, for example, that the amount of A_1^{PROT} scales linearly with k_1 (a measure of how effective the agonist is in stimulating signaling). The deterministic solution, on the other hand, is not expected to obey this linear scaling. Indeed, numerical solutions support these expectations (Fig. S2.3).

The complexity of the model described by Eqs. 1-6, however, makes it difficult to explore these differences in scaling behavior precisely. The complexity also prevents us from analyzing the necessary and sufficient conditions for purely stochastic instabilities (results in Figs. 2.2, 2.3) in cell signaling dynamics. Therefore, we formulated a simpler model that enabled exploration of these issues.

This minimal model, which can be solved *exactly*, includes the following features: irreversibility, branching, and feedback. The model is described in terms of the three coupled reactions shown below:

$$Z + Y \xrightarrow{\ k_1\ } X + Y, \quad Z + X + Y \xrightarrow{\ k_2\ } 2X + Y, \quad Y \xrightarrow{\ k_3\ } \varnothing \qquad (8)$$
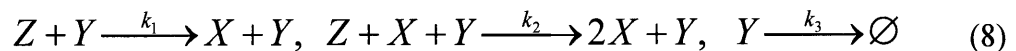
The deterministic equations corresponding to these reactions can be written down (web supplement) following the mass action kinetics. Let us denote the numbers of x, y and z species at time t by $N_x(t)$, $N_y(t)$ and $N_z(t)$, respectively. At $t = 0$, only Z and Y species are present; i.e., $N_x(0) = 0$, $N_y(0) = N$, and $N_z(0) = M$. As for the more complex model, the steady state values of the numbers of each species cannot be determined by setting the right sides of the above rate equations to zero; i.e., only a line of possible steady states can be obtained. Linear stability analysis of the steady state solutions shows that there is a neutral mode (with an eigenvalue 0) corresponding to sliding along the line of possible steady states, and stable modes along the directions $\delta N_x + (k_2 N_x^s + k_1)(M - N_x^s)/k_3 \delta N_y$ and δN_y respectively, which span the plane of the steady states. It is easy to solve the time dependent equations and take the $t \rightarrow \infty$ limit to obtain the unique steady-state solution for given initial conditions. The time-dependent solution to the deterministic equations describing system (8) is:

$$N_x(t) = \frac{k_1 M(F(t)-1)}{Mk_2 + k_1 F(t)}, \quad N_y(t) = Ne^{-k_3 t}, \quad N_z(t) = M - N_x(t) \tag{9}$$

where $F(t) = \exp\left[\frac{(Mk_2 + k_1)N(1 - e^{-k_3 t})}{k_3}\right]$. At long times ($t \gg k_3^{-1}$), the steady state particle numbers are,

$$N_x(t \rightarrow \infty) = N_x^s = \frac{k_1 M(\exp[N(Mk_2 + k_1)/k_3] - 1)}{Mk_2 + k_1 \exp[N(Mk_2 + k_1)/k_3]} \tag{10}$$

$$N_y(t \rightarrow \infty) = N_y^s = 0, \quad N_z(t \rightarrow \infty) = N_z^s = M - N_x(t \rightarrow \infty). \tag{11}$$

Given initial conditions, these equations determine a unique steady state, a behavior identical to that exhibited by the model described by Eqs. 1-6. Unlike the more complex model, the deterministic scaling behavior can be determined, and is given by,

$N_x^s(k_1, k_2, k_3, N, M) = Mf(Mk_2 k_1^{-1}, Nk_1 k_3^{-1})$.

The following Master Equation describes the stochastic time evolution of the reactions shown in (8):

$$\frac{\partial P(n_x, n_y, n_z, t)}{\partial t} = [k_2(n_x - 1)n_y(n_z + 1) + k_1 n_y(n_z + 1)]P(n_x - 1, n_y, n_z + 1, t)$$

$$+ k_3(n_y + 1)P(n_x, n_y + 1, n_z, t) - (k_2 n_x n_y n_z + k_1 n_y n_z + k_3 n_y)P(n_x, n_y, n_z, t) \tag{12}$$

$P(n_x, n_y, n_z, t)$ denotes the probability of having n_x, n_y and n_z particles at time t. The probability distribution at $t = 0$ is given by $P(n_x, n_y, n_z, t = 0) = \delta_{n_x,0}\delta_{n_y,N}\delta_{n_z,M}$. Note that at steady state (or in the limit, $t \to \infty$) there will be no y species present, and therefore, $P(n_x, n_y, n_z, t \to \infty) = \phi(n_x, n_z)\delta_{n_y,0}$. However, any form of $\phi(n_x, n_z)$ will make the right hand side of Eq. 12 vanish. Therefore, as for the deterministic equations, irreversibility makes it necessary to solve the time dependent Master equation for a particular initial condition in order to obtain the steady state solution.

Using the method of generating functions (Gardiner, 2004), Eq. 12 can be solved exactly (web supplement) to obtain:

$$P(n_x, n_y, n_z, t) = \delta_{n_x+n_z,M} \sum_{r=n_z}^{M} \lambda_r p_{n_z r} \, {}^N C_{n_y} \left[\frac{k_3}{A_r + k_3}(1 - \exp(-(A_r + k_3)t)) \right]^{N-n_y} \exp\left(-n_y(A_r + k_3)t\right)$$

(13)

where $A_r = r((M-r)k_2 + k_1)$ and $p_{n_z r} = {}^r C_{n_z}(-1)^{n_z} \dfrac{\Gamma(M + k_1/k_2 + 1 - r)\Gamma(M + k_1/k_2 - n_z)}{\Gamma(M + k_1/k_2 + 1 - n_z - r)\Gamma(M + k_1/k_2)}$.

$\{\lambda_r\}$ are determined from the equations:

$$\sum_{r=n}^{M} \lambda_r p_{nr} = 0 \quad \text{for} \quad n < M$$

$$= 1 \quad \text{for} \quad n = M$$

(14)

At long times ($t \to \infty$), the above probability distribution takes the form,

$$P(n_x, n_y, n_z, t \to \infty) = \delta_{n_x+n_z,M}\delta_{n_y,0} \sum_{r=n_z}^{M} \lambda_r p_{n_z r} \left[\frac{k_3}{(r(M-r)k_2 + rk_1) + k_3} \right]^N$$

(15)

Note that this solution to the Master equation indicates the appearance of a spectrum of time scales (indexed by r and n_y), which is presumably related to stochastic delays.

Eq. 15 results in a steady state probability distribution that is bimodal for small numbers of molecules (Fig. S2.5a) when the deterministic solution does not exhibit bistability in any parameter range. In the more complex model that we studied (Eqs. 1-6), mean-field behavior was obtained as the numbers of A_1 and A_2 molecules increased past a threshold value even though their relative numbers were kept constant. The corresponding limit for the minimal model is $k_3 \to \infty, N \to \infty$, with the ratio N/k_3 (or the dimensionless, Nk_1/k_3) remaining constant. This is because a large value of N corresponds to a large amount of the source of a positive regulator (A_1 in Eqs. 1-6) and a

33

large value of k_3 corresponds to greater annihilation or a big source (A_2 in Eqs. 1-6) of negative regulation. Fig. S2.5b shows that, like the more complex model, there is a purely stochastic transition as the stochastic solution is unimodal and distributed around the deterministic solution above a threshold value of N and k_3. So, these results establish that the sufficient conditions for the phenomena we report are: irreversibility, branching, and feedback loops. But, are these also necessary conditions?

The possibility of two different outcomes is obviously necessary, and branching is ubiquitous in cell signaling processes that lead to functional decisions. We have also found that removing irreversibility abolishes the phenomenon (data not shown). Ultimately, all reactions are, in principle, reversible. However, in the time scales of interest to signal propagation in cells, many steps are effectively irreversible.

Feedback regulation is also necessary as the bimodal stochastic solution does not exist if k_2 in the minimal model tends to zero (Fig. S2.9). Insight into the kind of feedback regulation that is necessary can be obtained by contrasting our studies of dueling feedback loops in cell signaling to a model for binary drift in population genetics (Gillespie, 2004; Rice, 2004). Consider a population of heterozygote individuals with two forms, B_1 and B_2, for a particular allele. In the absence of mutations, the number of each type of allele can change from generation to generation, even in a population of fixed size, due to mating. The effects of binary selection on the numbers of B_1 and B_2 forms can be roughly represented as follows (Gillespie, 2004; Rice, 2004):

$$B_1 + B_2 \xrightarrow{\ k\ } B_1 + B_1, \qquad B_1 + B_2 \xrightarrow{\ k'\ } B_2 + B_2 \qquad (16)$$
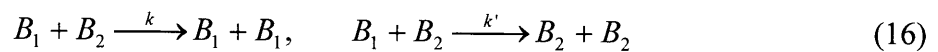
with k and k' related to the relative fitness of each phenotype.

The model described by Eq. 16 also contains branching and irreversibility. There is also an effective feedback, but unlike Eqs. 1-6 or Eq. 8, there is no separate intrinsic time scale associated with the feedback loops. A special case of this model (with no selection), k = k', shares some features with the systems we are considering. The deterministic changes in this limit are trivially zero, and any initial condition (along the fixed line of $B_1+B_2 = $ population size) remains fixed. These deterministic steady states are unique, but the stochastic solutions yield a bimodal distribution. This is because the stochastic trajectories are divided into two classes: ones which terminate when the

number of B_1 particles vanishes and those which terminate when the number of B_2 reaches zero.

There is an important difference, however, between the model for binary drift with $k = k'$ and the class of cell signaling models we have been considering. The stochastic solution of the model represented by Eq. 16 does not converge to the deterministic solution when the number of particles becomes large. The stochastic solution at $t \rightarrow \infty$ is always bimodal! The stochastic trajectories cease to evolve when either B_1 or B_2 become zero because only then is the effective rate of conversion between these species equal to zero. The deterministic rates of formation of B_1 and B_2 equal the same constant for all times. Increasing the numbers of molecules does not eliminate this difference between the deterministic and stochastic cases. As the number of particles increases, the stochastically determined time (τ') required for B_1 or B_2 to equal zero increases, but ultimately it always happens. There is no separate intrinsic time scale that can compete with increasing values of τ' as the number of particles increases and prevent this from happening (i.e., a bimodal solution). Recall that, for the signaling models that we focused on, the relative values of the stochastic delay, τ, and the separate time scale associated with feedback loops determined the stochastically driven transition when the number of molecules was lowered (Figs. 2.3 and S2.5b). The absence of such an interplay prevents a purely stochastic instability in the binary drift model as the number of particles decreases. Correspondingly, if the rate coefficients in the model represented by Eq. 16 were time dependent with an intrinsic time scale, the phenomenon of a purely stochastic instability would be recovered.

The analyses presented above suggest that the necessary and sufficient conditions for a purely stochastic bimodality in the absence of any deterministic bistabilities are: 1] irreversibility 2] branching and 3] feedback regulation with an associated distinct and fast time scale.

The analytical solution for the probability distribution (Eq. 15) obtained for the minimal model of cell signaling that satisfies these conditions enables us to calculate average properties, such as the average number of molecules of the product, $<x>$. This allows us to examine whether $<x>$ scales with parameter values in the same way as N_x determined from the mean-field equations (see above). The average value, $<x>$, is:

$$\langle x \rangle = \sum_{n=0}^{M} \sum_{r=n}^{M} a_r p_{nr} (M-n) \left(\frac{k_3}{r((M-r)k_2 + k_1) + k_3} \right)^N. \tag{17}$$

So, in general, there is no simple scaling law, such as N_x scaling with Nk_1/k_3, as in the deterministic limit. Does this "anomalous" scaling, originating from the importance of stochastic fluctuations, revert to mean-field scaling behavior in the limit corresponding to a large numbers of particles?

In order to answer this question, as shown above, we need to consider the value of <x> in the limit of large values of N, M, and k_3. Consider first the limit of large values of N and k_3 for a fixed value of Nk_1/k_3. Simple algebra yields the value of <x> in this limit to be:

$$Lt_{\substack{N \to \infty \\ k_3 \to \infty \\ k_3/N \text{ fixed}}} \langle x \rangle = \sum_{n=0}^{M} \sum_{r=n}^{M} a_r p_{nr} (M-n) \exp(-r((M-r)k_2/k_1 + 1)Nk_1/k_3). \tag{18}$$

So, the deterministic scaling with Nk_1/k_3 (Eq. 10) is recovered in the appropriate limit. Similarly, mean-field scaling is recovered in the limit of large values of N and M (web supplement).

The general solution (Eq. 18) for <x> does not allow us to explicate the non-mean-field scaling when fluctuations are important. This can be obtained analytically only in special limits. For example, consider the limit of infinitely strong feedback ($k_2 \to \infty$). In this limit, $\langle x \rangle$ takes the following form (web supplement):

$$\langle x \rangle = M \left(1 - \left(\frac{k_3}{Mk_1 + k_d} \right)^N \right) = M(1 - e^{-N \ln(1 + Mk_1/k_3)}) \tag{19}$$

Fig. 2.4 shows that $\langle x \rangle$ obtained from numerical solutions of the Master equation (Eq. 8) for different values of N and k_3 collapse to one master curve when scaled according to Eq. 19, a scaling that is distinctly different from the mean-field scaling with Nk_1/k_3. We have not been able to determine whether these specific differences in scaling laws between the deterministic and biologically relevant stochastic solutions are universal to all models which satisfy the necessary and sufficient conditions (identified earlier) for a purely stochastic instability.

2.4 Discussion

Dueling positive and negative feedback loops are ubiquitous in biology. In many instances, these processes involve small numbers of the pertinent molecules, and hence stochastic fluctuations can be important. We report a striking result for such systems. The models we have studied correspond to unique deterministic steady states for all parameter values, and do not exhibit bistability. Yet, when there are a small number of molecules, stochastic effects result in a bimodal solution with neither solution corresponding to the mean-field result. Our analyses suggest that the necessary and sufficient conditions for this phenomenon are irreversibility, branching, and the existence of an intrinsic and relatively fast time scale associated with feedback regulation. Our studies show that for specific examples of such systems, near the transition from one phenotype to another (e.g., agonism to antagonism), mean-field scaling does not apply to the stochastic solutions. Whether or not the specific differences in scaling between mean-field and stochastic solutions that we report are universal for the class of models which exhibit the phenomenon revealed by our studies remains an open question.

There is a key difference between models of gene regulation and cell signaling where bimodality has been observed in stochastic limits under conditions where the deterministic equations yield monostable solutions and our results. In the former examples (double negative feedback, dimer mediated gene regulation, *etc. e.g.* (Allen et al., 2006; Bhalla et al., 2002; Lai et al., 2004; McAdams and Arkin, 1997; Ozbudak et al., 2004; Sasai and Wolynes, 2003; Xiong and Ferrell, 2003)), bistable deterministic solutions exist in some other parameter regime. Stochastic bimodality displayed by binary drift models in population genetics are also different from the phenomena we report in that the stochastic solutions are always bimodal, regardless of the number of particles; i.e., there is no stochastically driven transition from a single solution to bistable solutions.

The necessary and sufficient conditions for the phenomenon that we report (branching, irreversibility, and feedback loops with distinct time scales) are quite common in cell biology. Our results suggest that these features, when combined with stochastic fluctuations, can enable cells to make binary decisions while this would not be possible in a deterministic world. For instance, if gene transcription and effector function

required greater than a threshold value of a downstream signaling product, in a mean-field world, cells would be unable to make decisions with a distinct functional outcome (Fig. 2.5). Under the same conditions, stochastic effects would result in cells being either "on" or "off" (Fig. 2.5), as observed in experimental studies in diverse contexts.

For example, a recent study of HIV latency by Weinberger *et al* (Weinberger et al., 2005) showed a 'temporary' bimodal cell population in a time window when there is no instability in the set of rate equations used to describe the signaling events. The main difference between this study and the results we have discussed is that in (Weinberger et al., 2005) the observed bimodality disappears at long times. Another example is provided by T cell signaling. It has been proposed that dueling feedback regulation could underlie how antagonists shut off signaling in T cells. Experiments show a bimodal response for a downstream signaling product (Erk), with the proportion of "off" cells increasing as the number of antagonists becomes larger (Altan-Bonnet and Germain, 2005; Stefanova et al., 2003). Stochastic simulations of a model of the T cell signaling network are in accord with these experimental observations (web supplement); i.e., bimodal distributions are the norm because of fluctuations, while the deterministic equations do not exhibit bistability in any parameter regime. We emphasize, however, that a bimodal or "digital" ERK response in T cells could also result from important contributions from other molecular mechanisms (Roose et al., 2007).

We hope that the possibility of purely stochastic instabilities which lead to distinct cellular decisions will be broadly explored in the context of cell signaling processes by carrying out single cell assays for systems where the necessary and sufficient conditions we have described are naturally present or are engineered.

2.5 Appendix to Chapter 2

1. Deterministic Equations for model described in section 2

$$\frac{dA_1}{dt} = -k_3 A_1 E - k_5 A_1 S$$

$$\frac{dA_1^{PROT}}{dt} = k_3 A_1 E$$

$$\frac{dE}{dt} = k_1 A_1 + k_4 A_1^{PROT} - k_D E$$

$$\frac{dS}{dt} = k_2 A_2 - k_D S$$

2. Exploring different ranges of parameters for model described in section 2

Our computational studies show that any combination of parameters that preserves strong feedbacks leads to a stochastic bimodal response. In addition to the case discussed in the text, bimodal behavior is observed when k_4, $k_5 >> k_1$, k_2, k_3. Biologically, this can correspond to a situation where upon action of the positive regulator E, A_1 gets further activated or it could result from cooperativity (Li et al., 2004a) with other molecules, resulting in faster rate of production of the positive regulator. Figures S2.1 and S2.2 are drawn in complete analogy to Fig. 2.2 and 2.3 of the main text of the Chapter and show stochastic bistability. This "all-or-none" behavior is explained by exactly the same arguments as those described in the text.

3. Stochastic and mean-field scaling in the model described in section 2

To illustrate the differences in scaling behavior between the stochastic and deterministic descriptions of the system, we calculated the amount of protected A_1 as a function of the rate constant k_1 for fixed values of k_2 and the amount of A_1. This could be considered to be analogous to computing the cellular response as the nature of agonist is changed. In Fig. S2.3, stochastic and deterministic dose response curves for the system with 10 initial A_1 molecules are plotted. The stochastic behavior (Fig. S2.3a) is manifested in linear *scaling* of the amount of A_1^{PROT} with k_1 (agonist quality): all curves coincide when the amount of A_1^{PROT} in the steady state is plotted against A_2 / k_1, which

represents the scaling variable $\dfrac{k_1 A_1}{k_2 A_2}$ (see main text) with k_2 and A_1 fixed. As can be seen in Fig. S2.3b, the deterministic solution does not obey this linear scaling. Moreover, the value of k_4 naturally affects the deterministic steady state value of A_1^{PROT}, while it does

not have any influence in the stochastic case, since feedback regulation occurs before reaction 4.

4. Rates of protection and inactivation as functions of time

On fig. S2.4, the time dependence of the rate of production of A_1^{PROT} (red curve) and of A_1^{INACT} (blue curve) are shown for excess of antagonist (S2.4a) and equal amounts of agonist and antagonist (S2.4b). The parameters of the model are the same as those used in th Figs 2 and 3 in the main text: $k_1=1$, $k_2=1$, $k_3=100$, $k_4=1$, $k_5=100$, $k_D=1$.

5. Mathematical Details of Solutions to the simpler model (eq 12 of the main text of this chapter).

This section is due to Jayajit Das

5a. Solution of the Meanfield Rate Equations

Here we describe the details of the calculations for the meanfield rate equations shown in Eqs. 13-15 in the main text.
The mean field equations are,

$$\frac{dN_x}{dt} = k_2 N_x N_y N_z + k_1 N_z N_y \qquad (A1)$$

$$\frac{dN_z}{dt} = -k_2 N_x N_y N_z - k_1 N_z N_y \qquad (A2)$$

$$\frac{dN_y}{dt} = -k_3 N_y \qquad (A3)$$

and the initial conditions are, $N_x(0)=0, N_y(0)=N$ and $N_z(0)=M$. Since the total number of x and z species are conserved at all times, we need to solve only two equations,

$$\frac{dN_x}{dt} = k_2 N_x N_y (M-N_x) + k_1 (M-N_x) N_y \text{ and, } \frac{dN_y}{dt} = -k_3 N_y.$$

The equation for N_y can be readily solved to get $N_y(t) = Ne^{-k_3 t}$. Substituting this form of $N_y(t)$ in the equation for N_x we get,

$$\frac{dN_x}{dt} = (k_2 N_x + k_1)(M - N_x)Ne^{-k_3 t}$$

$$\Rightarrow \frac{dN_x}{(k_2 N_x + k_1)(M - N_x)} = Ne^{-k_3 t}dt$$

$$\Rightarrow \frac{1}{Mk_2 + k_1}\left[\frac{k_2 dN_x}{(k_2 N_x + k_1)} + \frac{dN_x}{(M - N_x)}\right] = Ne^{-k_3 t}$$

$$\Rightarrow N_x(t) = \frac{k_1 M(F(t)-1)}{Mk_2 + k_1 F(t)}$$

$$\text{where, } F(t) = \exp\left[(Mk_2 + k_1)\frac{N}{k_3}(1 - e^{-k_3 t})\right]$$

Therefore, the solutions to Eq. (A1-A3) are,

$$N_x(t) = \frac{k_1 M(F(t)-1)}{Mk_2 + k_1 F(t)} \qquad (A4)$$

$$N_y(t) = Ne^{-k_3 t} \qquad (A5)$$

$$N_z(t) = M - N_x(t) \qquad (A6)$$

Fig. S2.6 shows the variation of the steady state value of N_x with k_3 for various initial numbers of the y species. The number of x species produced at the steady state decreases exponentially as $k_3 > N(Mk_2 + k_1)$.

5b. Large particle number limit ($M \to \infty$ and $N \to \infty$) from the mean-field solution

From Eq. A4, $N_x(t) = \dfrac{M(e^{Nk_1(Mk_2/k_1+1)k_3^{-1}}-1)}{Mk_2/k_1 + e^{Nk_1(Mk_2/k_1+1)k_3^{-1}}} = \dfrac{M(1 - e^{-Nk_1(Mk_2/k_1+1)k_3^{-1}})}{1 + (Mk_2/k_1)e^{-Nk_1(Mk_2/k_1+1)k_3^{-1}}}.$ As,

$M \to \infty$ and $N \to \infty$, $Nk_1(Mk_2/k_1+1)/k_3 \gg 1$, therefore,

$$\underset{\substack{N \to \infty \\ M \to \infty}}{Lt} N_x(t) = M(1 - O(e^{-Nk_1(Mk_2/k_1+1)k_3^{-1}})) \qquad (A7)$$

In the next section, we will show how the average particle number of species x, calculated from the stochastic solution of the Master Equation corresponds to Eq. A7 in large particle number limit.

5c. Exact Solution of the Master Equation

We describe the details of calculations for the solution of the Master Equation in Eq. 19. The Master Equation is given by,

$$\frac{\partial P(n_x, n_y, n_z, t)}{\partial t} = [k_2(n_x - 1)n_y(n_z + 1) + k_1 n_y(n_z + 1)]P(n_x - 1, n_y, n_z + 1, t) + k_3(n_y + 1)P(n_x, n_y + 1, n_z, t)$$

$$- (k_2 n_x n_y n_z + k_1 n_y n_z + k_3 n_y)P(n_x, n_y, n_z, t)$$

$$(B1)$$

We define a generating function, $G(s_1,s_2,s_3,t) = \sum_{n_x=0}^{M}\sum_{n_y=0}^{N}\sum_{n_z=0}^{M} s_1^{n_x} s_2^{n_y} s_3^{n_z} P(n_x,n_y,n_z,t)$

(Gardiner, 2004). The time evolution of the generating function determined by the above Master Equation is given by,

$$\frac{\partial G}{\partial t} = k_2 s_1 s_2 (s_1 - s_3)\partial_{s_1}\partial_{s_2}\partial_{s_3}G + k_1(s_1 - s_3)\partial_{s_2}\partial_{s_3}G - k_3(s_2 - 1)\partial_{s_2}G \tag{B2}$$

At $t = 0$, $G(s_1,s_2,s_3,t=0) = s_2^N s_3^M$, in addition to that, it should satisfy $G(1,1,1,t) = 1$ at all times, which is a condition for the conservation of the sum of the probabilities for all possible particle configurations.

If we look for a solution in terms of the reduced variables, s_1, s_2 and $\xi = (s_1 - s_3)/s_1$ then $G(s_1,s_2,\xi,t)$ satisfies the following equation:

$$\frac{\partial G}{\partial t} = -k_2 s_1 s_2 \xi \partial_{s_1}\partial_{s_2}\partial_\xi G - k_2\xi(\xi-1)s_2\partial_{s_2}\partial_\xi^2 G + (k_2 - k_1)s_2\xi\partial_{s_2}\partial_\xi G - k_3(s_2 - 1)\partial_{s_2}G \tag{B3}$$

We define, $G(s_1,s_2,\xi,t) = s_1^{\alpha_l}G'(s_2,\xi,t)$,

$$\frac{\partial G'}{\partial t} = -k_2 s_2 \xi(1-\xi)\partial_{s_2}\partial_\xi^2 G' - (k_2\alpha_l - k_2 + k_1)s_2\xi\partial_{s_2}\partial_\xi G' - k_3(s_2 - 1)\partial_{s_2}G' \tag{B4}$$

If, $G'(s_2,\xi,t) = e^{E_m t}\phi(s_2,\xi)$, then,

$$E_m = -k_2 s_2 \xi(1-\xi)\partial_{s_2}\partial_\xi^2\phi - (k_2\alpha_l - k_2 + k_1)s_2\xi\partial_{s_2}\partial_\xi\phi - k_3(s_2 - 1)\partial_{s_2}\phi \tag{B5}$$

Introducing a separation of variables, $\phi(s_2,\xi) = S(s_2)\Xi(\xi)$, Eq. B5 becomes,

$$\frac{k_2\xi(1-\xi)}{S\Xi}\frac{d^2\Xi}{d\xi^2}\frac{dS}{ds_2} + \frac{\xi(k_2\alpha_l - k_2 + k_1)}{S\Xi}\frac{d\Xi}{d\xi}\frac{dS}{ds_2} + \frac{k_3(s_2 - 1)}{s_2 S}\frac{dS}{ds_2} + \frac{E_m}{s_2} = 0 \tag{B6}$$

The above equation will be satisfied if,

$$\frac{k_2\xi(1-\xi)}{S\Xi}\frac{d^2\Xi}{d\xi^2}\frac{dS}{ds_2} + \frac{\xi(k_2\alpha_l - k_2 + k_1)}{S\Xi}\frac{d\Xi}{d\xi}\frac{dS}{ds_2} = A_n f(s_2) \tag{B7}$$

and,

$$\frac{k_3(s_2 - 1)}{s_2 S}\frac{dS}{ds_2} + \frac{E_m}{s_2} = -A_n f(s_2). \tag{B8}$$

However, from Eq. (B7) we get,

$$\frac{1}{S}\frac{dS}{ds_2} = f(s_2) \tag{B9}$$

Therefore, in order to make Eq. B8 consistent with Eq. B9 we have to choose the following form for $f(s_2)$: $f(s_2) = -\dfrac{E_m}{(k_3 + A_n)s_2 - k_3}$.

Using the above form we get the solution for $S(s_2)$ as,

$$S(s_2) = \left(s_2 - \frac{k_d}{k_3 + A_n}\right)^{-\frac{E_m}{k_3 + A_n}}.$$

$\Xi(\xi)$ satisfies the equation below,

$$k_2\xi(1-\xi)\frac{d^2\Xi}{d\xi^2} + \xi(k_2\alpha_l - k_2 + k_1)\frac{d\Xi}{d\xi} - A_n\Xi = 0, \text{ By changing } \xi \text{ to } \eta = 1 - \xi = s_3/s_1, \text{ we get,}$$

$$k_1\eta(1-\eta)\frac{d^2\Xi}{d\eta^2} - (1-\eta)\beta_l\frac{d\Xi}{d\eta} - A_n\Xi = 0 \tag{B10}$$

where, $\beta_l = k_2\alpha_l - k_2 + k_1$

This equation has the form of the ODE which yields hypergeometric functions (Bateman Manuscript Project. et al., 1953) as its solutions. The following ODE has hypergeometric functions as its solutions.

$$z(1-z)\frac{d^2u}{dz^2} + (c - (a+b+1)z)\frac{du}{dz} - abu = 0 \tag{B11}$$

The hypergeometric function is defined as,

$$_2F_1(a,b,c;z) = \sum_{n=0}^{\infty}\frac{(a)_n(b)_n}{(c)_n}z^n \tag{B12}$$

, where, $(a)_n = \dfrac{\Gamma(a+n)}{\Gamma(a)}$. The series in Eq. B12 is convergent for a positive c and $|z|<1$. If a is a negative integer, i.e., $a = -n$, the series terminates after z^n. Comparing Eq. B10, to Eq. B11 we get, if, $c = -\dfrac{\beta_l}{k_2} = -\alpha_l + 1 - \dfrac{k_1}{k_2}$, and $a = -n$, then $b = -\alpha_l + n - \dfrac{k_1}{k_2}$. In that case, $A_n = n(-\alpha_l + \dfrac{k_1}{k_2} - n)k_2$.

Therefore, the general solution to Eq. B2 is,

$$G(s_1,s_2,s_3,t) = \sum_{m=0}^{\infty}\sum_{l=0}^{\infty}\sum_{n=0}^{\infty}\lambda_{lmn}s_1^{\alpha_l}\left(s_2 - \frac{k_3}{A_n+k_3}\right)^{-\frac{E_m}{A_n+k_3}}e^{E_m t}\,_2F_1\left(a,b,c;\frac{s_3}{s_1}\right). \tag{B13}$$

The constants are chosen in such a way that the solution satisfies the initial condition and also the sum of the probabilities is conserved at all times.

At t = 0, Eq.B13 takes the form,

$$G(s_1,s_2,s_3,t=0) = \sum_{m=0}^{\infty}\sum_{l=0}^{\infty}\sum_{n=0}^{\infty}\lambda_{lmn}s_1^{\alpha_l}\left(s_2 - \frac{k_3}{A_n+k_3}\right)^{-\frac{E_m}{A_n+k_3}}\,_2F_1\left(a,b,c;\frac{s_3}{s_1}\right) \tag{B14}$$

Now, $s_2^N = \left[\left(\dfrac{k_3}{A_n+k_3}\right) + \left(s_2 - \dfrac{k_3}{A_n+k_3}\right)\right]^N = \sum_{m=0}^{N}{}^N C_m\left(\dfrac{k_3}{A_n+k_3}\right)^{N-m}\left(s_2 - \dfrac{k_3}{A_n+k_3}\right)^m$, therefore, if we choose,

$$E_m = -m(k_3 + A_n) \text{ and } \lambda_{lmn} = {}^N C_m\left(\frac{k_3}{A_n+k_3}\right)^{N-m}\lambda_{ln} \text{ for } m \leq N$$

$$= 0 \qquad\qquad \text{for } m > N$$

Then Eq. B14 satisfies the initial condition for s_2. If we choose, $\alpha_M = M$ and $\lambda_{ln} = \delta_{lM}\lambda_n$ then Eq.B14 assumes the form below,

$$G(s_1, s_2, s_3, t=0) = s_2^{N} s_1^{M} \sum_{n=0}^{\infty} \lambda_n \, _2F_1\left(a, b, c; \frac{s_3}{s_1}\right)$$

$$= s_2^{N} s_1^{M} \sum_{n=0}^{\infty} \lambda_n \sum_{r=0}^{n} p_{rn}\left(\frac{s_3}{s_1}\right)^{r}$$

$$= s_2^{N} s_1^{M} \sum_{n=0}^{M} \sum_{r=n}^{M} \lambda_r p_{nr}\left(\frac{s_3}{s_1}\right)^{n}$$

$$= s_2^{N} s_1^{M} \sum_{n=0}^{M} q_n\left(\frac{s_3}{s_1}\right)^{n}$$

where, $p_{rn} = \dfrac{(a_n)_r (b_n)_r}{(c_n)_r \, r!}$, $q_n = \sum_{r=n}^{M} \lambda_r p_{nr}$ and $a_n = -n$, $b_n = -M + n - \dfrac{k_1}{k_2}$ and

$c_n = -M + 1 - \dfrac{k_1}{k_2}$. If $\begin{aligned} q_n &= 0 \text{ for } n < M \\ &= 1 \text{ for } n = M \end{aligned}$ then, Eq.B14 satisfies the initial condition,

$G(s_1, s_2, s_3, t=0) = s_2^{N} s_3^{M}$.

Therefore, the time dependent generating function is given by,

$$G(s_1, s_2, s_3, t) = \sum_{n=0}^{M} \sum_{r=n}^{M} \lambda_r p_{nr} s_1^{M-n} f(r, s_2, t) s_3^{n} \tag{B15}$$

where, $f(s_2, r, t) = \left[\dfrac{k_3}{A_r + k_3} + \left(s_2 - \dfrac{k_3}{A_r + k_3}\right)e^{-(k_3 + A_r)t}\right]^{N}$ and

$$\begin{aligned} \sum_{r=n}^{M} \lambda_r p_{nr} &= 0 \text{ for } n < M \\ \lambda_M p_{MM} &= 1 \text{ for } n = M \end{aligned} \tag{B16}$$

It is straightforward to show, $G(s_1 = 1, s_2 = 1, s_3 = 1, t) = 1$ for the above choices of the coefficients. The proof follows below. From Eq. B15,

$$G(s_1 = 1, s_2 = 1, s_3 = 1, t) = \sum_{n=0}^{M} \sum_{r=n}^{M} \lambda_r p_{nr} f(r, s_2 = 1, t) = \sum_{n=0}^{M} \sum_{r=0}^{M} \lambda_n p_{rn} f(n, s_2 = 1, t).$$

$\therefore G(1,1,1,t) = 1$, if, $\sum_{r=0}^{n} p_{rn} = 0$ for $n > 0$ and $\lambda_0 = 1$. This is because, $f(0,1,t) = 1$ and $p_{00} = 1$.

<u>Proof:</u> $\sum_{r=0}^{n} p_{rn} = 0$, when, $n > 0$.

$$\sum_{r=0}^{n} p_{rn} = \frac{\Gamma(M + k_1/k_2 + 1 - n)}{\Gamma(M + k_1/k_2)} \left[\sum_{r=0}^{n} {}^nC_r(-1)^r \frac{\Gamma(M + k_1/k_2 - r)}{\Gamma(M + k_1/k_2 + 1 - n - r)} \right]$$

$$= \frac{\Gamma(M + k_1/k_2 + 1 - n)}{\Gamma(M + k_1/k_2)} \left[\sum_{r=0}^{n} {}^nC_r(-1)^r \sum_{p=0}^{n-1} S_{n-1}^{(p)}(a - r)^p \right] \text{ where, } a = M + k_1/k_2 - 1$$

$$= \frac{\Gamma(M + k_1/k_2 + 1 - n)}{\Gamma(M + k_1/k_2)} \left[\sum_{p=0}^{n-1} \sum_{q=0}^{p} S_{n-1}^{(p)} {}^pC_q a^{p-q} \left(\sum_{r=0}^{n} r^q {}^nC_r(-1)^r \right) \right]$$

In the above, expressions, $S_n^{(m)}$ denotes the Stirling number of the first kind (Gradshtein et al., 2000). However, $\left(x\frac{\partial}{\partial x} \right)^q (1-x)^n = \sum_{r=0}^{n} {}^nC_r(-1)^r r^q x^r$, therefore,

$$\sum_{r=0}^{n} {}^nC_r(-1)^r r^q = \left(x\frac{\partial}{\partial x} \right)^q (1-x)^n \bigg|_{x=1}. \text{ For, } 0 \le q \le n-1, \left(x\frac{\partial}{\partial x} \right)^q (1-x)^n \bigg|_{x=1} = 0, \text{ hence,}$$

$$\sum_{r=0}^{n} p_{rn} = 0.$$

<u>Proof:</u> $\lambda_0 = 1$

Using, Eq. B16,

$$\sum_{n=0}^{M} \sum_{r=n}^{M} \lambda_r p_{nr} = 1, \Rightarrow \sum_{n=0}^{M} \sum_{r=0}^{n} \lambda_n p_{rn} = 1, \Rightarrow \lambda_0 p_{00} + \sum_{n=1}^{M} \sum_{r=0}^{n} \lambda_n p_{rn} = 1.$$

Since, $\sum_{r=0}^{n} p_{rn} = 0$ for n > 0 and $p_{00} = 1$; $\lambda_0 = 1$.

Therefore, $G(1,1,1,t) = 1$. QED.

Expanding the polynomials one can easily get the probability distribution

$$P(n_x, n_y, n_z, t) = \delta_{n_x + n_z, M} \sum_{r=n_z}^{M} \lambda_r p_{n_z r} {}^NC_{n_y} \left[\frac{k_3}{A_r + k_3} (1 - \exp(-(A_r + k_3)t)) \right]^{N - n_y} \exp\left(-n_y(A_r + k_3)t\right)$$

(B17)

where, $A_r = r((M - r)k_2 + k_1)$ and $p_{n_z r} = {}^rC_{n_z}(-1)^{n_z} \frac{\Gamma(M + k_1/k_2 + 1 - r)\Gamma(M + k_1/k_2 - n_z)}{\Gamma(M + k_1/k_2 + 1 - n_z - r)\Gamma(M + k_1/k_2)}$.

$\{\lambda_r\}$ are determined from the equations,

$$\sum_{r=n}^{M} \lambda_r p_{nr} = 0 \quad \text{for} \quad n < M$$

$$= 1 \quad \text{for} \quad n = M$$

At large times, $t \to \infty$, $f(s_2, r, t) = \left[\frac{k_3}{A_r + k_3} \right]^N$, because,

$A_r = r((M - r)k_2 + k_1) > 0$ for $r \le M$. Therefore, the generating function at the steady state has the following form,

$$G(s_1, s_2, s_3, t \to \infty) = \sum_{n=0}^{M} \sum_{r=n}^{M} \lambda_r p_{nr} s_1^{M-n} \left(\frac{k_3}{k_3 + A_r} \right)^N s_3^n \tag{B18}$$

Hence, the probability distribution function at $t \to \infty$ is given by,

$$P(n_x, n_y, n_z, t \to \infty) = \delta_{n_x + n_z, M} \delta_{n_y, 0} \sum_{r=n_z}^{M} a_r p_{n_z r} \left[\frac{k_3}{(r(M-r)k_2 + rk_1) + k_3} \right]^N. \qquad (B19)$$

For fixed k_1, k_2, and k_d, the factor $u(r) = \left[\dfrac{k_d}{(r(M-r)k_1 + rk_2) + k_d} \right]^N$ is peaked at $r = 0$ and r

$= M$ which corresponds to the cases at $n_x = M$ and $n_x = 0$ respectively. The values of the

peaks are $u(r = 0) = 1$ and $u(r = M) = \left(\dfrac{1}{Mk_1/k_3 + 1} \right)^N$. Therefore, the peak at $r = M$, will

have significant contribution when, $k_3 \gg Mk_1$, i.e., the y particles decay at a much faster rate than it generates particles of the x species. Furthermore, if the initial number of the y particles increases, the value of the peak at $r = M$ goes down. Therefore, we can expect to see a bimodal behavior in the distribution function for $k_3 \gg Mk_1$ and small N. Fig. S2.7 displays the above characteristics in the distribution function.

5d. Calculation of the average particle number

The average particle number of any species can be easily calculated from the generating function, $G(s_1, s_2, s_3, t)$. For example, the average number of x species is given by,

$$\langle x(t) \rangle = \partial_{s_1} G(s_1, s_2, s_3, t) \big|_{s_1 = 1, s_2 = 1, s_3 = 1} = \sum_{n=0}^{M} \sum_{r=n}^{M} \lambda_r p_{nr} f(r, s_2 = 1, t)(M - n) \qquad (B20)$$

At, $t \to \infty$, the above average takes the following form,

$$\langle x(t \to \infty) \rangle = \sum_{n=0}^{M} \sum_{r=n}^{M} \lambda_r p_{nr} \left(\frac{k_3}{r((M-r)k_2 + k_1) + k_3} \right)^N (M - n) \qquad (B21)$$

5e. Limit of large N and large k_1/k_3, keeping the ratio Nk_1/k_3 fixed:
Let us, write,

$$\left(\frac{k_3}{r((M-r)k_2 + k_1) + k_3} \right)^N = \left(\frac{1}{a/N + 1} \right)^N, \text{ where, } a = r((M-r)k_2/k_1 + 1)Nk_1/k_3.$$

Therefore, in the limit, $N \to \infty$ and $Nk_1/k_3 = const$,

$$Lt_{N \to \infty} \left(\frac{1}{a/N + 1} \right)^N = Lt_{y \to 0} \left(\frac{1}{ay + 1} \right)^{1/y} = g$$

when $y = 1/N$.

Now, $\ln(g) = -1/y \ln(ay + 1) \to -a + O(y)$ as $y \to 0$.

$$\therefore \ Lt_{N \to \infty} \left(\frac{1}{a/N + 1} \right) = \exp(-a)$$

Thus,

$$Lt_{\substack{N \to \infty \\ k_3 \to \infty \\ k_3/N \text{ fixed}}} \langle x \rangle = \sum_{n=0}^{M} \sum_{r=n}^{M} a_r p_{nr} (M - n) \exp(-r((M-r)k_2/k_1 + 1)Nk_1/k_3). \qquad (B22)$$

5f. Large M limit:

Rewriting, Eq. B22 as,

$$Lt_{\substack{N\to\infty\\k_3\to\infty\\k_3/N\,\text{fixed}}} \quad \langle x\rangle = \sum_{n=0}^{M}\sum_{r=0}^{n} a_n p_{rn}(M-r)\exp(-n((M-n)k_2/k_1+1)Nk_1/k_3)$$

$$= M - \sum_{n=1}^{M} a_n \exp(-n((M-n)k_2/k_1+1)Nk_1/k_3)\sum_{r=0}^{n} rp_{rn}$$

(B23)

In the limit, $M\to\infty$, each term in the sums of Eq. B23 decay exponentially with M, thus, we can write, $Lt_{\substack{N\to\infty\\k_3\to\infty\\k_3/N\,\text{fixed}\\M\to\infty}} \quad \langle x\rangle = M(1-O(e^{-Nk_1/k_3(Mk_2/k_1+1)}))$.

This form is consistent with the large particle limit of the solutions of the meanfield rate equations in Eq. A7.

5g. Strong feedback limit ($k_2\to\infty$ and $k_1/k_2=\varepsilon\to 0$)
In this limit, $k_2\to\infty$ and $k_1/k_2=\varepsilon\to 0$.

Particle distribution function
The probability of having no x species in the steady state is given by,

$$P(n_x=0,n_y=0,n_z=M)=\lambda_M p_{MM}\left(\frac{k_3}{Mk_1+k_3}\right)^N = \left(\frac{k_3}{Mk_1+k_3}\right)^N, \text{ because, } \lambda_M p_{MM}=1 \text{ from}$$

Eq.B16. Note, the distribution does not depend on k_2 and this form of the distribution function holds good for any value of k_2.
Now, in the limit, $k_1/k_2=\varepsilon\to 0$,

$$p_{MM}={}^{M}C_M(-1)^M\frac{\Gamma(1)\Gamma(\varepsilon)}{\Gamma(-M+1+\varepsilon)\Gamma(M)}=-1, \text{ using (Abramowitz and Stegun, 1972),}$$

$$\Gamma(-n+\varepsilon)=\frac{(-1)^n}{n!}\left(\frac{1}{\varepsilon}-\gamma\right) \text{ as } \varepsilon\to 0.$$

(B22)

where, $\gamma\approx 0.5772$ is the Euler Mascheroni constant.
We need to evaluate, other p_{nr} for $r=n....M$ ($0\le n\le M$) and $\lambda_1.....\lambda_M$ in order to compute the probability distribution function for all particle numbers. Now,

$$p_{nr}={}^{r}C_n(-1)^n\frac{\Gamma(M+\varepsilon+1-r)\Gamma(M+\varepsilon-n)}{\Gamma(M+\varepsilon+1-n-r)\Gamma(M+\varepsilon)}\to 0 \qquad \text{for} \quad n+r\ge M+1 \text{ and } n\ne r. \text{ For}$$

$n+r\ge M+1$ and $n=r$; $p_{nn}=-1$. Therefore, $\lambda_r=0$ for $M/2\le r<M$ which can be easily shown from Eq. B16. For, $n+r<M+1$ $p_{nr}\to{}^{r}C_n(-1)^n\frac{\Gamma(M+1-r)\Gamma(M-n)}{\Gamma(M+1-n-r)\Gamma(M)}$, however, from Eq. B16 it can be shown that, $\lambda_r=0$ for $0<r<M/2$ and $\lambda_0=1$. Therefore, the probability distribution is,

$$P(n_x,n_y=0,n_z)=\left(\frac{k_3}{Mk_1+k_3}\right)^N\delta_{n_x,0}\delta_{n_z,M}+\left[1-\left(\frac{k_3}{Mk_1+k_3}\right)^N\right]\delta_{n_x,M}\delta_{n_z,0}$$

(B23)

Note, that the distribution is strictly bimodal with peaks at $n_x = 0$ and $n_x = M$. The magnitudes of the peaks depend on Mk_1, N and k_3. The probability of having no particles of x species is easy to guess from following observation: Starting with N particles of y species at $t = 0$, the probability of having N successive y annihilation events is $\left(\dfrac{1}{Mk_1/k_3+1}\right)^N$, which is also the probability of having no particles of x species in the steady state. Now intuitively on can think that when a single reaction for the creation of the x species occurs, the strong positive feedback will convert all the z species into the x species. Fig. S2.8 shows the comparison of Eq. B23 with the Gillespie simulation with a very large positive feedback.

5h. Calculation of the average particle number
Using the same properties of the coefficients, $\{\lambda_r\}$ and $\{p_{nr}\}$ it can be shown that,

$$\langle x(t \to \infty)\rangle = \lambda_0 M + \lambda_M M\left(\frac{k_3}{k_3 + Mk_1}\right)^N = M\left(1 - \left(\frac{k_3}{k_3 + Mk_1}\right)^N\right) \qquad \text{(B24)}$$

5i. No Feedback Limit ($k_2 \to 0$)
The limit, $k_2 \to 0$, is tricky to take directly from Eq. 20 because k_2 multiplies the highest derivative in Eq. B3, therefore, analyzing the limit $k_2 \to 0$ becomes a case of singular perturbation theory. A simpler approach would be to analyze the case with $k_2 = 0$ from the Master Equation and solve it directly. In that case, the Master Equation will be given by,

$$\frac{\partial P(n_x,n_y,n_z,t)}{\partial t} = k_1 n_y(n_z + 1)P(n_x - 1, n_y, n_z + 1, t) + k_3(n_y + 1)P(n_x, n_y + 1, n_z, t) \qquad \text{(B25)}$$
$$- (k_1 n_y n_z + k_3 n_y)P(n_x, n_y, n_z, t)$$

The equation followed by the generating function,
$$G(s_1, s_2, s_3, t) = \sum_{n_x=0}^{M}\sum_{n_y=0}^{N}\sum_{n_z=0}^{M} s_1^{n_x} s_2^{n_y} s_3^{n_z} P(n_x, n_y, n_z, t) \text{ is,}$$

$$\frac{\partial G}{\partial t} = k_1 s_2 (s_1 - s_3)\partial_{s_2}\partial_{s_3} G - k_3(s_2 - 1)\partial_{s_2} G \qquad \text{(B26)}$$

This equation can be solved in a similar way by changing to variables, s_1, s_2 and $\xi = (s_1 - s_3)/s_1$ and performing separation of variable on the ensuing equation. The general solution of Eq. B26 is

$$G(s_1, s_2, s_3, t) = s_1^M \sum_{m=0}^{N}\sum_{n=0}^{M}(-1)^n\, {}^M C_n\, {}^N C_m \left(1 - \frac{s_3}{s_1}\right)^n \left(s_2 - \frac{k_3}{k_3 + nk_1}\right)^m \left(\frac{k_3}{k_3 + nk_1}\right)^{N-m} \exp(-m(k_3 + nk_1)t) \qquad \text{(B27)}$$

In the steady state ($t \to \infty$), the probability distribution can be easily obtained from Eq. B27, which is given by,

$$P(n_x, n_y, n_z, t \to \infty) = \delta_{n_x+n,M}\,\delta_{n_y,0} \sum_{r=n_z}^{M} {}^M C_r\, {}^r C_{n_z} (-1)^{r+n_z}\left(\frac{k_3}{rk_1 + k_3}\right)^N. \qquad \text{(B28)}$$

This form always gives a unimodal distribution (Fig. S2.9, also see Fig. S2.8(b)). Therefore, the nonlinear feedback is essential in order to realize a bimodal distribution.

2.6 Bibliography

Abramowitz, M., and Stegun, I.A. (1972). Handbook of mathematical functions with formulas, graphs, and mathematical tables, 10th printing, with corrections. edn (Washington: U.S. Govt. Print. Off.).

Acar, M., Becskei, A., and van Oudenaarden, A. (2005). Enhancement of cellular memory by reducing stochastic transitions. Nature *435*, 228-232.

Allen, R.J., Frenkel, D., and ten Wolde, P.R. (2006). Simulating rare events in equilibrium or nonequilibrium stochastic systems. Journal of Chemical Physics *124*, 024102.

Altan-Bonnet, G., and Germain, R.N. (2005). Modeling T cell antigen discrimination based on feedback control of digital ERK responses. PLoS Biol *3*, e356.

Bateman Manuscript Project., Bateman, H., Erdélyi, A., and United States. Office of Naval Research. (1953). Higher transcendental functions (New York,: McGraw-Hill).

Berg, O.G., Paulsson, J., and Ehrenberg, M. (2000). Fluctuations and quality of control in biological cells: Zero-order ultrasensitivity reinvestigated. Biophysical Journal *79*, 1228-1236.

Bhalla, U.S., Ram, P.T., and Iyengar, R. (2002). MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. Science *297*, 1018-1023.

Brower, R.C., England, R., Takeshita, T., Kozlowski, S., Margulies, D.H., Berzofsky, J.A., and Delisi, C. (1994). Minimal Requirements for Peptide Mediated Activation of Cd8(+) Ctl. Molecular Immunology *31*, 1285-1293.

Davis, M.M., Krogsgaard, M., Huse, M., Huppa, J., Lillemeier, B.F., and Li, Q.J. (2007). T cells as a self-referential, sensory organ. Annu Rev Immunol *25*, 681-695.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science *297*, 1183-1186.

Evavold, B.D., Sloan-Lancaster, J., and Allen, P.M. (1994). Antagonism of superantigen-stimulated helper T-cell clones and hybridomas by altered peptide ligand. Proc Natl Acad Sci U S A *91*, 2300-2304.

Gardiner, C.W. (2004). Handbook of stochastic methods : for physics, chemistry, and the natural sciences, 3rd edn (Berlin ; New York: Springer).

Gillespie, D.T. (1977). Exact Stochastic Simulation of Coupled Chemical-Reactions. Journal of Physical Chemistry *81*, 2340-2361.

Gillespie, J.H. (2004). Population genetics : a concise guide, 2nd edn (Baltimore, Md.: Johns Hopkins University Press).

Gradshtein, I.S., Ryzhik, I.M., and Jeffrey, A. (2000). Table of integrals, series, and products, 6th edn (San Diego: Academic Press).

Irvine, D.J., Purbhoo, M.A., Krogsgaard, M., and Davis, M.M. (2002). Direct observation of ligand recognition by T cells. Nature *419*, 845-849.

Karmakar, R., and Bose, I. (2007). Positive feedback, stochasticity and genetic competence. Phys Biol *4*, 29-37.

Kepler, T.B., and Elston, T.C. (2001). Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. Biophys J *81*, 3116-3136.

Lai, K., Robertson, M.J., and Schaffer, D.V. (2004). The Sonic hedgehog signaling system as a bistable genetic switch. Biophysical Journal *86*, 2748-2757.

Li, Q.J., Chau, J., Ebert, P.J., Sylvester, G., Min, H., Liu, G., Braich, R., Manoharan, M., Soutschek, J., Skare, P., *et al.* (2007). miR-181a is an intrinsic modulator of T cell sensitivity and selection. Cell *129*, 147-161.

Li, Q.J., Dinner, A.R., Qi, S., Irvine, D.J., Huppa, J.B., Davis, M.M., and Chakraborty, A.K. (2004). CD4 enhances T cell sensitivity to antigen by coordinating Lck accumulation at the immunological synapse. Nat Immunol *5*, 791-799.

Lin, J., and Weiss, A. (2001). T cell receptor signalling. J Cell Sci *114*, 243-244.

McAdams, H.H., and Arkin, A. (1997). Stochastic mechanisms in gene expression. Proc Natl Acad Sci U S A *94*, 814-819.

Ozbudak, E.M., Thattai, M., Lim, H.N., Shraiman, B.I., and Van Oudenaarden, A. (2004). Multistability in the lactose utilization network of Escherichia coli. Nature *427*, 737-740.

Purbhoo, M.A., Irvine, D.J., Huppa, J.B., and Davis, M.M. (2004). T cell killing does not require the formation of a stable mature immunological synapse. Nat Immunol *5*, 524-530.

Rice, S.H. (2004). Evolutionary theory : mathematical and conceptual foundations (Sunderland, Mass., USA: Sinauer Associates).

Roose, J.P., Mollenauer, M., Ho, M., Kurosaki, T., and Weiss, A. (2007). Unusual interplay of two types of Ras activators, RasGRP and SOS, establishes sensitive and robust Ras activation in lymphocytes. Mol Cell Biol *27*, 2732-2745.

Samoilov, M., Plyasunov, S., and Arkin, A.P. (2005). Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations. Proceedings of the National Academy of Sciences of the United States of America *102*, 2310-2315.

Sasai, M., and Wolynes, P.G. (2003). Stochastic gene expression as a many-body problem. Proceedings of the National Academy of Sciences of the United States of America *100*, 2374-2379.

Starr, T.K., Jameson, S.C., and Hogquist, K.A. (2003). Positive and negative selection of T cells. Annu Rev Immunol *21*, 139-176.

Stefanova, I., Hemmer, B., Vergelli, M., Martin, R., Biddison, W.E., and Germain, R.N. (2003). TCR ligand discrimination is enforced by competing ERK positive and SHP-1 negative feedback pathways. Nat Immunol *4*, 248-254.

Sykulev, Y., Joo, M., Vturina, I., Tsomides, T.J., and Eisen, H.N. (1996). Evidence that a single peptide-MHC complex on a target cell can elicit a cytolytic T cell response. Immunity *4*, 565-571.

Weinberger, L.S., Burnett, J.C., Toettcher, J.E., Arkin, A.P., and Schaffer, D.V. (2005). Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. Cell *122*, 169-182.

Wylie, D.C., Das, J., and Chakraborty, A.K. (2007). Sensitivity of T cells to antigen and antagonism emerges from differential regulation of the same molecular signaling module. Proc Natl Acad Sci U S A *104*, 5533-5538.

Xiong, W., and Ferrell, J.E., Jr. (2003). A positive-feedback-based bistable 'memory module' that governs a cell fate decision. Nature *426*, 460-465.

Yachi, P.P., Ampudia, J., Gascoigne, N.R.J., and Zal, T. (2005). Nonstimulatory peptides contribute to antigen-induced CD8-T cell receptor interaction at the immunological synapse. Nature Immunology *6*, 785-792.

2.7 Figures for Chapter 2

No signal

Activating Stimuli
Inhibiting Stimuli
Positive Regulator
Negative Regulator
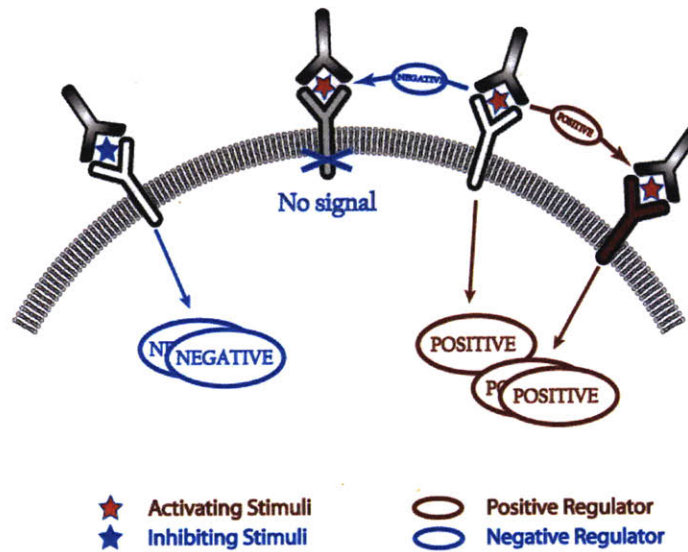
NEGATIVE

POSITIVE

POSITIVE

Figure 2.1 A schematic representation of dueling positive and negative feedback loops stimulated upon receptor binding to stimulatory or inhibitory ligands. The negative regulator can shut off signaling by inactivating the receptor-associated signaling complex (negative feedback), while the positive regulator could prevent this inhibitory interaction and increase or continue production of downstream signaling products (positive feedback).
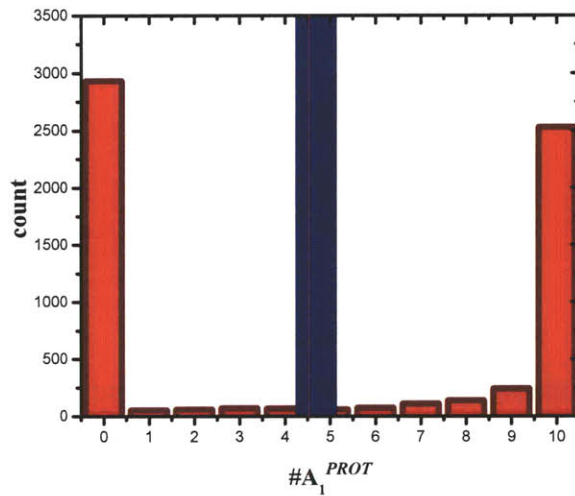
Figure 2.2 Bimodal stochastic solutions distinct from the unique deterministic solution. Histogram showing the bimodal distribution of the protected agonists at steady state (red) for a situation when there are 10 agonist (A1) and 10 antagonists (A2). The corresponding single steady state solution of the deterministic ODEs (blue) is also shown. The other parameter values are: k1 =1, k2 =1, k3 =100, k4 =1, k5 =100, kD =1 (all s-1), and statistics were collected over 5000 trajectories obtained using the Gillespie algorithm. The result is robust to variations in the parameter values as long as there is strong feedback.
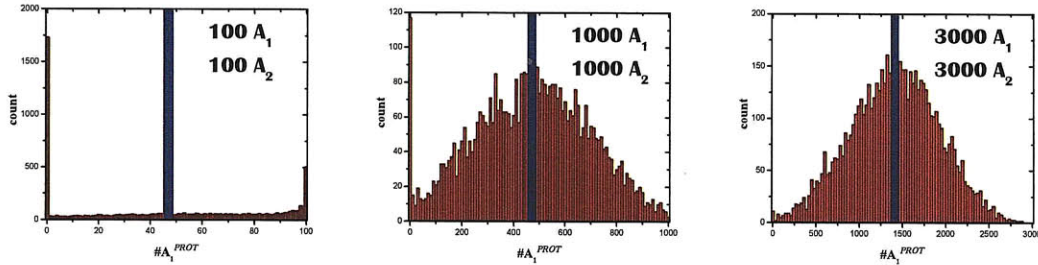
Figure 2.3 A purely stochastically driven transition. Histograms showing the distribution of the protected agonists at steady state (red) and corresponding steady state solution of deterministic ODEs (blue) for different amounts of agonist and antagonist. All other parameters are identical to that in Fig. 2.
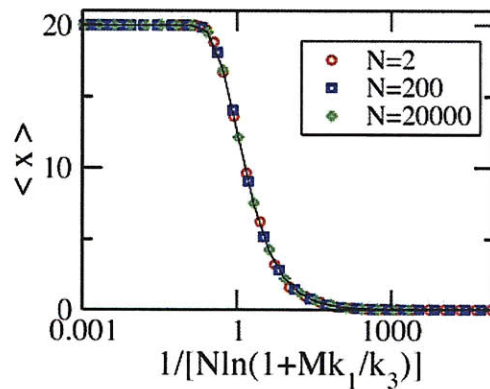

Figure 2.4 Results from the Minimal Model. Non meanfield scaling in the limit of large positive feedback ($k_2 \rightarrow \infty$):The average values of X species, <x>, at steady state obtained from Gillespie simulations scale with $(1/\log(1 + Mk_1/k_3))/N$ instead of the mean field scaling variable Nk1/k3. The values of the parameter k2 is 100 s-1 (i.e. a large value). k1=0.0012 s-1 and M=20 are held fixed as k3 and N are varied. The solid line is a plot of the scaling function shown in Eq. 19.
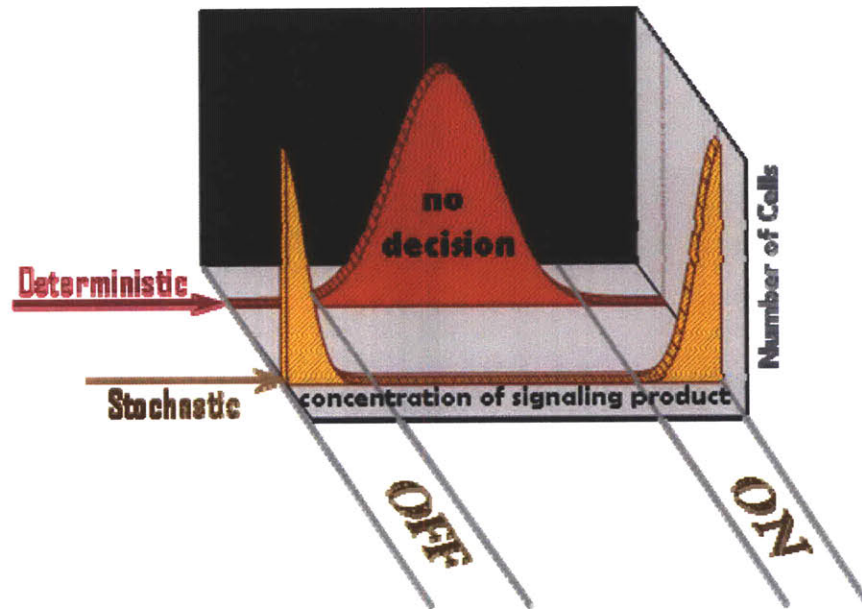
Figure 2.5 Stochastic fluctuations can enable cellular decisions. Schematic representation showing that irreversibility, branching, and dueling feedback loops associated with intrinsic time scales, when combined with stochastic effects, can result in distinct functional decisions for each cell. A deterministic treatment would mask this ability of cells to make decisions.
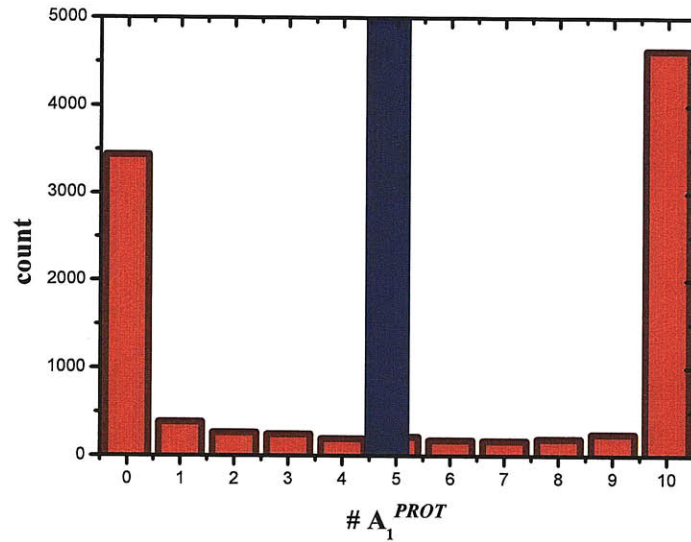
Fig. S2.1: Bimodal stochastic solutions distinct from the monostable deterministic solution. Histogram showing the bimodal distribution of the protected agonists at steady state (red) for a situation when there are 10 agonist (A_1) and 10 antagonists (A_2). The corresponding single steady state solution of the deterministic ODEs (blue) is also shown. The other parameter values are: $k_1=3$, $k_2=0.7$, $k_3=1$, $k_4=1000$, $k_5=50$, $k_D=1$, and statistics were collected over 5000 trajectories obtained using the Gillespie algorithm. The robustness of this result to variations in the parameter values is discussed in the main text and section 1 of the web supplement.
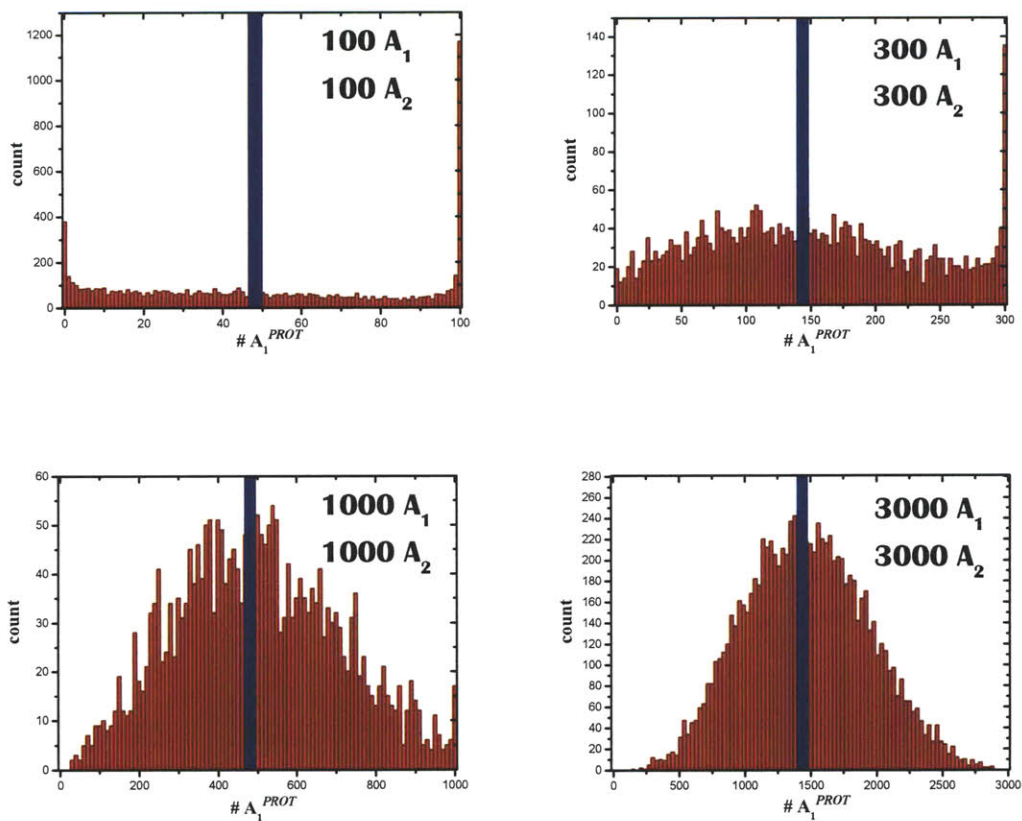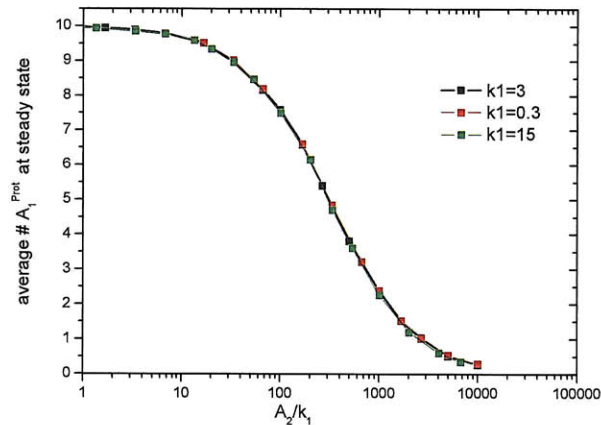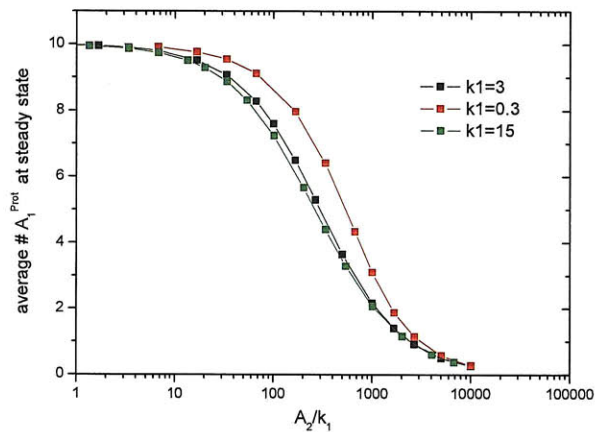
Fig. S2.2: A purely stochastic instability. Histograms showing the distribution of the protected agonists at steady state (red) and corresponding steady state solution of deterministic ODEs (blue) for different amounts of A_1 and A_2 molecules. All other parameters are identical to that in Fig. S1.

(a)

(b)

Fig. S2.3: The variation of the average number of A_1^{PROT} species as a function of the number of A_2 molecules and the parameter k_1 (which could be considered to reflect the quality of the agonist).

a) Results of the stochastic simulations collapse to one master curve when scaled with $\dfrac{A_2}{k_1}$

b) The deterministic results do not follow this scaling. All results are for cases where k_3, $k_5 \gg k_1$, k_2, k_4.
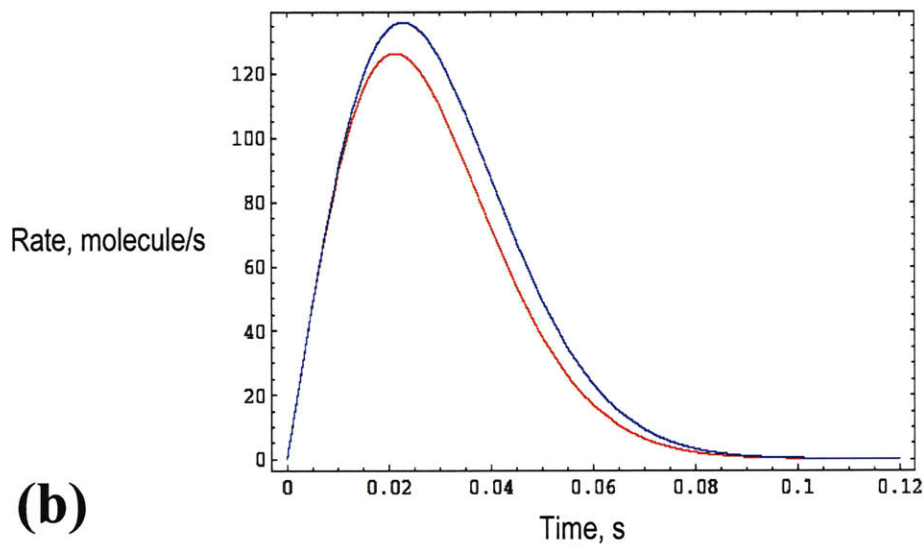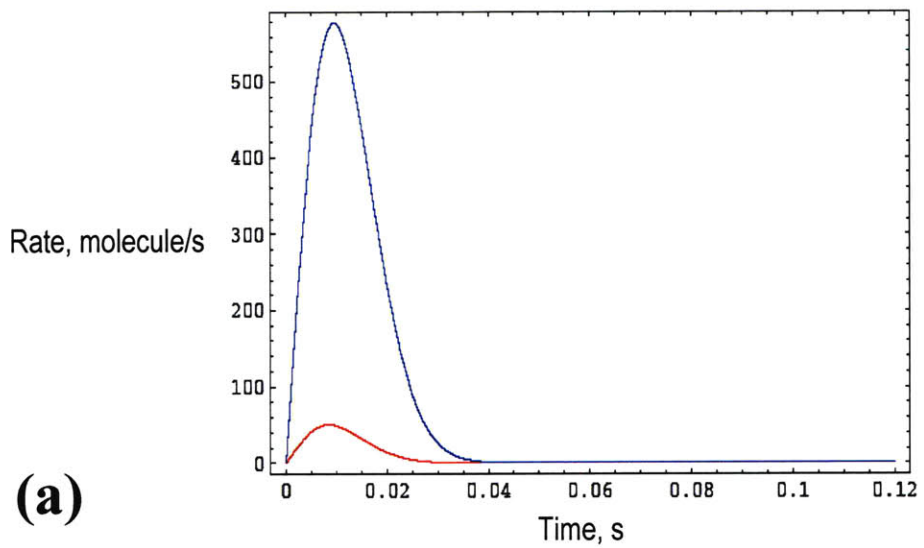
(a)

(b)

Fig. S2.4: Time dependence of the deterministic rate of production of protected species (red) and inactivated species (blue) as a function of time.

    a) For 10 molecules A_1 and 100 molecules A_2

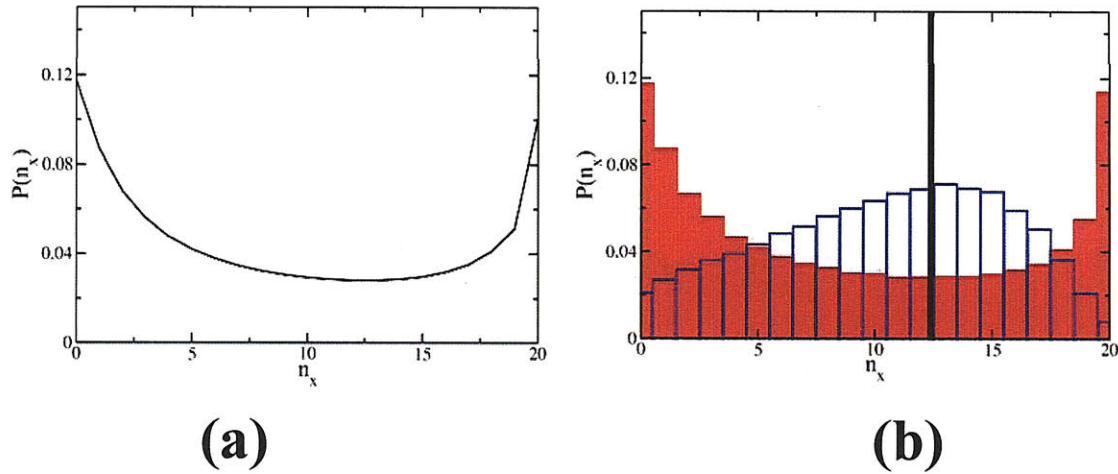   b) For 10 molecules A_1 and 10 molecules A_2

(a)

(b)

Fig. S2.5: Results from the Minimal Model. (a) *Exact solution of the Master Equation:* The exact solution (Eq. 20) shows a bimodal distribution for low values of initial numbers of Y (N=2) and Z (M=20) species. The reaction rates are taken to be, k_1=0.0012 s^{-1} , k_2 = 0.0010 s^{-1}, and k_3 =0.0125 s^{-1} (b) *Minimal model captures the essential characteristics of the larger model:* Distribution of the number of X species at steady state is calculated from Gillespie simulations as both N (related to agonist number) and k_3 (related to antagonist number) are increased keeping the ratio N/k_3 fixed. All the other parameters, as well as the ratio, N/k_3, are the same as in (a). The distribution is markedly bimodal for a low value (shown in red) of $N = 2$. The distribution becomes unimodal (shown in blue), peaked at the mean field value (shown with the black bar) as both N and k_3 are increased 1000 fold keeping all other parameters unchanged. We use Gillespie simulations instead of the exact solution (Eq. 20) for the above cases because numerical evaluation of the Gamma functions for large arguments (required to evaluate Eq. 20) is computationally more expensive than carrying out Gillespie simulations. The Gillespie simulations agree with the exact solution.
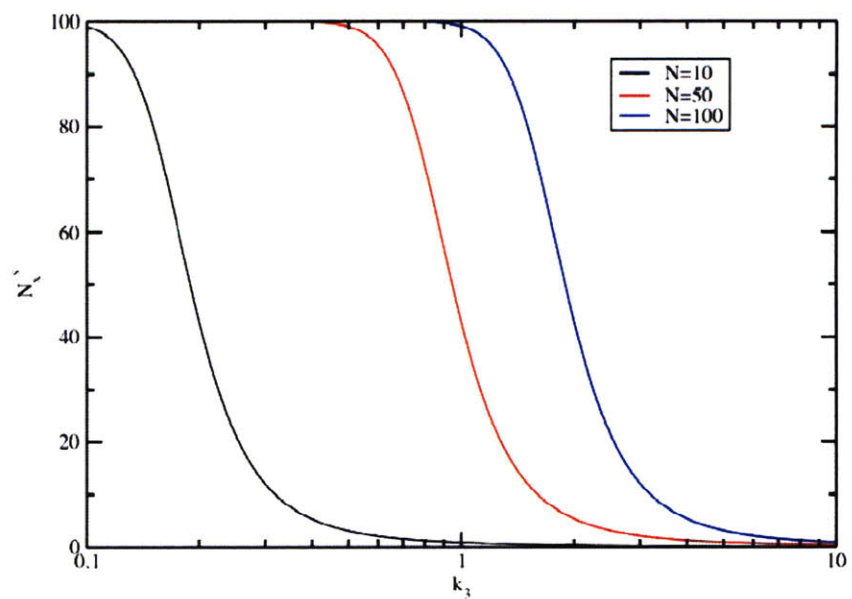
Fig. S2.6: The steady values of the number (N_x^s) of the x species as the decay rate of the y species (k_3) is varied. Three cases, corresponding to different values of the initial numbers (N) of species Z are shown.
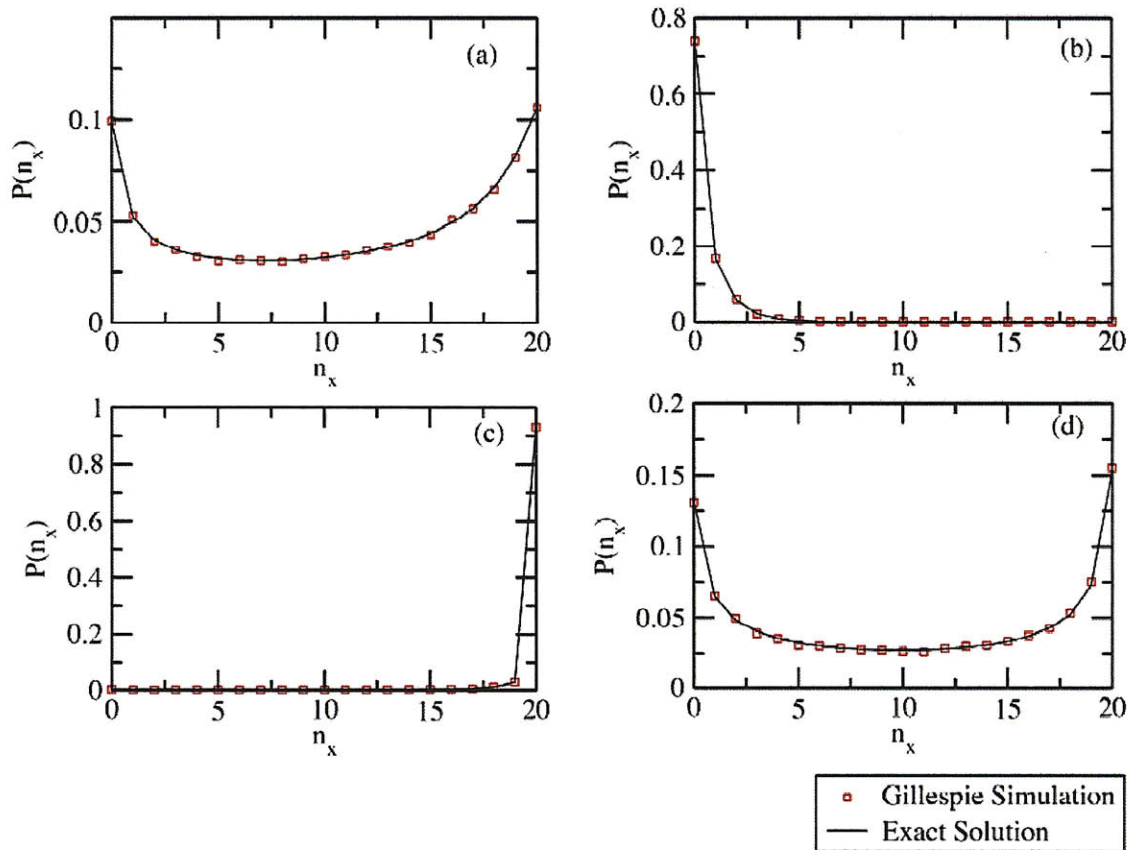
Fig. S2.7 The distribution of the number of x particles at the steady state. Results obtained from the analytic solutions (given by Eq. 20) are compared with the results from Gillespie simulations. All the cases have a fixed $k_2=0.001$ s^{-1}, and are started with the same number ($M = 20$) of particles for the Z species. (a) A bimodal distribution is obtained for $k_1=0.0005$ s^{-1}, $k_3=0.06$ s^{-1} and $N = 15$, where N is the number of Y species at t=0. (b) The bimodal distribution in (a) turns in a unimodal distribution as k_3 is inceased to 0.5 keeping other parameters fixed. (c) The distribution becomes unimodal when k_1 is increased to 0.005 s^{-1}. (d) The bimodal distribution in (a) becomes sharper as N is reduced to 5 with other parameters held fixed.
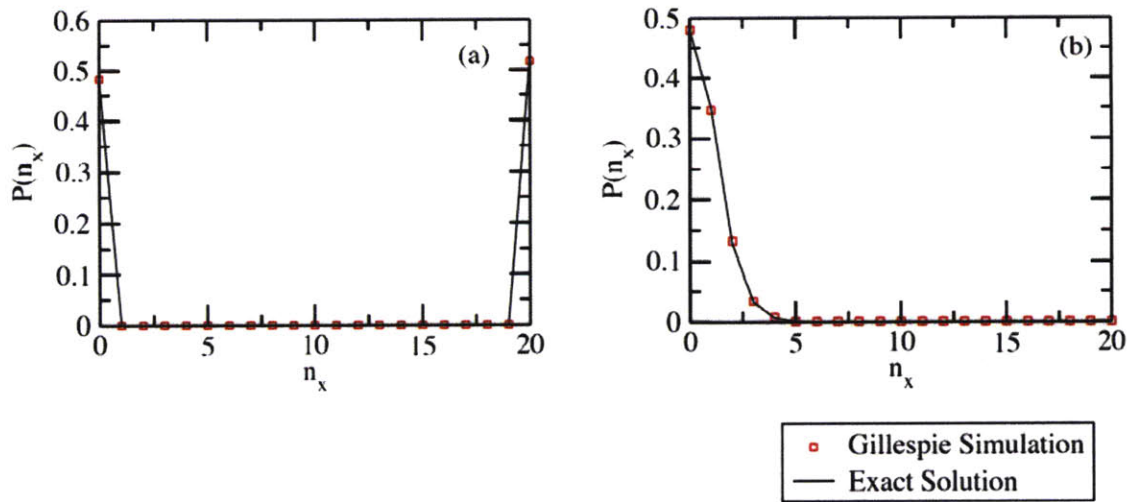
Fig S2.8 The particle number distribution function for the x species at the steady state for (a) a very strong positive feedback, $k_2 = 2s^{-1}$ and (b) for a very weak positive feedback $k_2 = 10^{-7}s^{-1}$. The values of the other parameters are, $k_1 = 0.0005\,s^{-1}$, $k_3 = 0.2s^{-1}$, $M = 20$ and $N = 15$, where, M and N are the numbers of particles at t=0 for the Z and Y species respectively. The solid line and the red points are obtained from the analytical solution and the Gillespie simulation respectively.
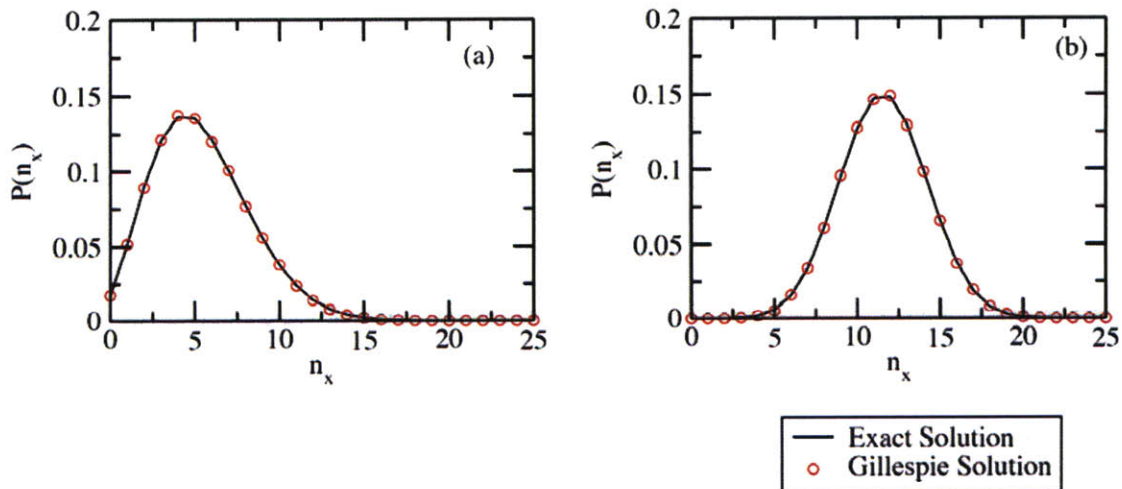

Fig. S2.9 Particle distribution functions for the X species do not show any bimodality when there is no positive feedback, *i.e.*, k_2=0. The plots show cases for (a) M=25, N=5, k_1=0.005s^{-1} and k_3=0.1 s^{-1}, (b) M=25, N=100, k_1=0.005s^{-1} and k_3=0.8 s^{-1}. The Gillespie solution is compared with the exact solution in Eq. B28.

Chapter 3

Stochastic bimodalities in deterministically mono-stable reversible chemical networks due to network topology reduction

3.1 Introduction

Diverse cellular functions are mediated by signal transduction and subsequent gene transcription events. The dynamical behavior of chemical reaction networks control and regulate these processes. The dynamics of spatially-homogeneous chemical reactions are often described by deterministic ordinary differential equations in terms of classical chemical kinetics (CCK) (Goldbeter and Koshland, 1981; Gomez-Uribe et al., 2007; Lai et al., 2004; Markevich et al., 2004). The mean-field character of such a treatment is exemplified by considering the following deterministic ordinary differential equation describing the dynamics of second order reactions such as $A + B \xrightarrow{k} AB$:

$$-\frac{d\langle A \rangle}{dt} = k\langle A \cdot B \rangle \approx k\langle A \rangle\langle B \rangle \tag{1}$$

where k is the rate coefficient. In writing Eq. 1, the number of molecules of each species is described by an average concentration (<A> or <B>) and the average of the product of the number of A and B molecules is replaced by the product of the average concentrations. In other words, stochastic and discrete features of the underlying molecular number levels, including fluctuations and associated correlations are ignored.

Cell signaling and gene transcription often involve small copy numbers of the pertinent molecules. Therefore, many important examples of stochastic fluctuations in determining cellular response have been reported (Artyomov et al., 2007a; Berg et al., 2000; Kepler and Elston, 2001a; Levine et al., 2007). Accurate analysis of these and other chemical processes – for which the underlying discrete molecular states or random nature

of individual interactions become important – requires methods able to capture such features. This is frequently done in via the chemical master equation (CME) (McQuarrie, 1967).

For example, when a continuous-deterministic CCK description of the dynamics of chemical reactions yields multiple steady states in some parameter range, the corresponding discrete-stochastic CME descriptions will generally produce a multimodal distribution of responses. This is because CCK closely follows modes of the underlying CME distribution, so presence of multiple steady states under the same set of parameters broadly indicates existence of multiple distribution modes (Samoilov and Arkin, 2006). Notably, CME distributions with this type of multimodality, e.g. bimodality with cells being either "on" or "off", can be realized for parameter ranges where a deterministic multi/bistability is predicted as well as outside of these regimes. The latter phenomenon results from stochastic sampling of parameters or dynamic states in the deterministically bistable regime, which is enabled by the fluctuations inherent in CME system trajectories that are not available under CCK.

A more intriguing class of phenomena is comprised of studies showing the existence of bimodal stochastic responses for systems whose deterministic description yields monostable solutions for all parameter ranges. Two types of reaction networks in this class have been reported. One is comprised of systems with absorbing states (i.e., systems comprised of irreversible chemical reactions). We have recently demonstrated that the necessary and sufficient conditions for such a system to exhibit purely stochastic bimodal responses are the existence of more than one absorbing state and feedback loops characterized by distinct time scales (Artyomov et al., 2007a). The focus of this paper is on chemical reaction networks without absorbing states, e.g. networks comprised of reversible chemical reactions, that exhibit bimodal stochastic responses when a deterministic treatment is devoid of instabilities in any parameter range.

An example of such a system is obtained by considering the following simple birth-death process for a species X (Lipshtat et al., 2006; Loinger et al., 2007):

$$\phi \underset{k_{-X}}{\overset{k_X}{\rightleftharpoons}} X \tag{2}$$

The rate constants in Eq. 2 can be chosen such that at steady state there are only few molecules of X present. This reaction can then be coupled to a fast "indicator" reaction as shown below:
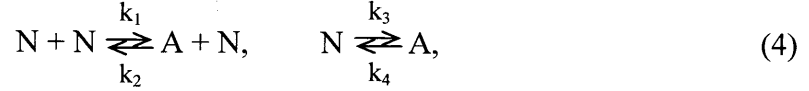
$$X \xrightarrow{k_Y} X + Y; \; Y \xrightarrow{k_{-Y}} \phi \qquad (3)$$

Species X gives birth to species Y with rate k_Y, and Y can be degraded with the rate k_{-Y}. If the rate constants k_Y and k_{-Y} are chosen to be much larger than k_X and k_{-X}, an adiabatic concentration of Y is established corresponding to the particular value of X being sampled stochastically. When a small number of X molecules is present, on average, one can see the signature of the discreteness of X in multiple peaks appearing in the steady-state probability distribution of Y (see Fig. 3.1, simulations carried out with standard Gillespie algorithm(Gillespie, 1977)), provided that the rates of the reactions (3) are fast enough that peaks in the steady-state distribution of Y are resolved. The "indicator reaction" effectively amplifies the discrete nature of the molecules of X which is why this scenario can be called the "discreteness amplification" scenario for obtaining multi-peaked distributions for deterministically monostable systems. A particular example of this scenario that was presented in Ref ((Lipshtat et al., 2006)) can be obtained from the reaction scheme (2-3) by restricting possible numbers of X molecules to zero or one. Referring to the state with X = 0 as the inactive state of a gene and X = 1 as the active state, a bimodal distribution of cellular response is obtained for conditions where the adiabatic limit is approached (see above).

Our focus is on a different class of chemical reaction networks without absorbing states that can exhibit purely stochastic bimodalities. Kinetic schemes in this class have been described previously (Samoilov et al., 2005b), but the underlying reason for a bimodal stochastic response in the absence of any deterministic instabilities remained unclear. Here, we show that a previously unreported phenomenon, network topology reduction, is one of the mechanisms that could result in this unusual behavior.

3.2 Model development

We start by considering the following simple chemical reaction network:

$$N + N \underset{k_2}{\overset{k_1}{\rightleftarrows}} A + N, \qquad N \underset{k_4}{\overset{k_3}{\rightleftarrows}} A, \tag{4}$$

$$A \underset{k_6}{\overset{k_5}{\rightleftarrows}} B \tag{5}$$

The deterministic kinetic equations for this system are

$$\frac{dN}{dt} = -k_1 N^2 + k_2 \, N \, A - k_3 N + k_4 A$$

$$\frac{dA}{dt} = k_1 N^2 - k_2 \, N \, A + k_3 N - k_4 A - k_5 A + k_6 B \tag{6}$$

$$\frac{dB}{dt} = k_5 A - k_6 B$$

Eq. 6 makes clear that the quadratic equation obtained for steady-state concentrations of A (or B) can only result in a single stable fixed point for all possible values of rate parameters. Stochastic simulation of this reaction network for some choices of parameter values, however, yields a bimodal response for the number of B molecules (Fig. 3.2). This phenomenon cannot be explained by the arguments described above in the "amplification of discreteness" scenario.

One of the peaks in the bimodal distribution in Fig. 3.2 corresponds to the monostable deterministic steady state solution of Eq. 6, but the other is different. The explanation for this unexpected stochastic bimodality can be found by considering the time courses of the concentrations of N and B simultaneously (Fig. 3.2, inset). One notices that values of B corresponding to the peak that is not centered around the deterministic solution (i.e., B~27) are sampled when the number of molecules of N in the system accidentally, by way of stochastic fluctuations, becomes zero. In this situation, the reaction network effectively reduces in size because all reactions where N is among the reactants cannot occur, which leads to certain kinetic degrees of freedom otherwise available to the system to be temporarily "frozen out", subsequently constraining it onto a smaller dynamical manifold. The latter may potentially display different temporal or stationary features, thus contributing another behavioral mode to discrete-stochastic reaction network properties that may then be reflected in the overall species state distribution.

While a general analytical investigation of specific CME mechanisms underlying such phenomena is substantially outside the scope of this work and would need to be further addressed/discussed elsewhere, the described mechanism provides a compelling example of how this type of deviant chemical and biochemical dynamics can arise even in seemingly simple reaction mechanisms, making them relevant for *in vivo* and *in vitro* applications. For example, of the six reactions in the scheme described by Eqs. 4-5 only three are still possible if N is eliminated. If parameters are chosen appropriately, this reduced network may be realized for a sufficient amount of time to allow sampling of its steady state, with the second peak in the bimodal response shown in Fig. 3.2 corresponding to it. The only way for the system to escape from being "trapped" or "frozen" in the reduced network is thorugh the occurrence of a reaction that produces N, i.e. the one that converts A to N here.

This type of behavior was, in fact, observed by Samoilov *et al(Samoilov et al., 2005b)*and it was attributed to the non-linearity of the noise appearing in the system coupled to the "noise generator" of reactions of type (4) (or reaction 9 below). These authors considered (Samoilov et al., 2005b) the much-studied and biochemically ubiquitous futile cycle (Berg et al., 2000; Goldbeter and Koshland, 1981; Gomez-Uribe et al., 2007; Levine et al., 2007), which interconverts X and X* with the help of enzymes E_+ and E_- according to the standard Michaelis-Menten mechanism, with E_+ subject to noise (9):

$$X + E_+ \underset{k_{+2}}{\overset{k_{+1}}{\rightleftharpoons}} X{\cdot}E_+ \overset{k_{+3}}{\rightarrow} X^* + E_+ \tag{7}$$

$$X^* + E_- \underset{k_{-2}}{\overset{k_{-1}}{\rightleftharpoons}} X^*{\cdot}E_- \overset{k_{-3}}{\rightarrow} X + E_- \tag{8}$$
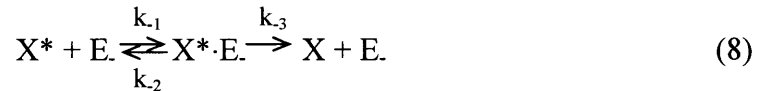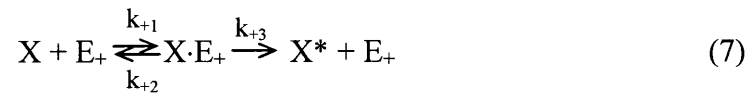
In (Samoilov et al., 2005b), the peculiar bimodal steady-state distribution of species X has been attributed to a sufficiently non-linear yet monomodal external noise distributions imposed on the forward enzyme E_+ in the Michaelis-Menten reaction network. The specific example considered in (Samoilov et al., 2005b) achieved the required non-linearity of noise appearing in the system reactions (7-8) by having it coupled to a "noise generator" reaction mechanism of type (4) (or reaction 9 below).

$$N + N \underset{k_{-21}}{\overset{k_{21}}{\rightleftarrows}} E_+ + N \qquad N \underset{k_{-22}}{\overset{k_{22}}{\rightleftarrows}} E_+, \qquad (9)$$
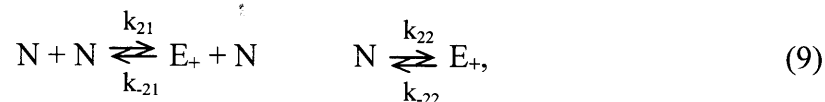
The deterministic steady-state equations for the reaction system (7-9) are polynomials of up to 6th order and it is not straightforward to show that there is no bistability for *all* possible sets of rate constants. This can be circumvented by using the topological rules described by Feinberg and coworkers (Craciun et al., 2006), which allow us to conclude that the system can not admit more than one positive steady-state, regardless of parameter values. However, a bimodal steady-state distribution of X (and X*) was found in(Samoilov et al., 2005b) for a fully discrete-stochastic description of this chemical network in a narrow range of parameters (Fig. 3.3). Just like in the simple example considered previously, the behavior of the generator reaction (9) may be viewed as having two network topologies. For non-zero values of N, the complete network (7-9) is explored and a peak around its steady state solution emerges. But when the number of molecules of N stochastically becomes zero – the effective topology is reduced similar to that of mechanism 4-5 (see Fig. 3.3).

The necessary conditions for observing bimodality due to network topology reduction is that the steady-states for complete and reduced topologies are sufficiently different so that two distinct peaks can be resolved. The other necessary condition is that system stays "arrested" in the reduced topology for a time scale sufficient for sampling its steady state. Samoilov et al.(Samoilov et al., 2005b) actually achieved this for the driver reaction (9) by "kinetically" arresting the system in the reduced topology by setting k_{-22} (9) to be smaller than reaction rates pertinent for the reduced network (7-8). Since reaction k_{-22} is the only possible way to return from the reduced topology to the complete network, this kinetic restriction fulfils the second necessary requirement.

The transient nature of the bimodality observed by Samoilov *et al* (Samoilov et al., 2005b) is related to the violation of the second necessary condition above, which requires that the system be able to spend sufficient time in the $N = 0$ state and that is possible only for a relatively small number of N molecules in the steady-state. Therefore, the observed bimodality disappears in Samoilov *et al* (Samoilov et al., 2005b) when increasing the number of molecules of N participating in the reaction. For $N + E_+ = 40$,

the peak around X= 790 corresponding to reduced topology becomes very small compared to the situation when N + E₊ = 35 (Fig. 3.3). In spite of the fact that peaks are well separated, the reduced topology is rarely sampled because fluctuations leading to N=0 are rare and the time spent in this state is short when N_{tot} is large.

When decreasing the number of molecules of N in the system, steady-state concentrations for complete and reduced topologies are very close to each other, and can not be resolved in the simulation. For N + E₊ = 35, one sees two distinct peaks corresponding to X = 1455 (complete system) and X = 1120 (reduced topology). However, for N + E₊ = 30, steady-states for the full and reduced topologies are X = 1766 and X=1639, respectively. These peaks can not be resolved completely due to intrinsic noise of the level ± 70 molecules at steady-state conditions (Fig.3.3). With the further decrease of concentration, the two peaks merge in to one.

3.3 Solution of Fokker-Planck Equation

We next turn to the possibility of treating systems of this type with continuous approximation methods, particularly the Fokker-Planck equation. The interest is stimulated by the fact that bimodality in this class of systems occurs due to special behavior at the single point where N=0. Is N=0 still a special point when N can take on non-integer values in a continuous approximation (Gardiner, 2004). For arbitrary small N, as long as it is not exactly 0, the effective topology is still the complete topology of the network, and the effect of the second attractor is not obvious.

Consider again the simplest kinetic scheme (4,5) featuring the network-topology reduction induced bimodality. The corresponding master-equation

$$\frac{dP_{n,a,b}(t)}{dt} = -\{k_1 n(n-1) + k_2 na + k_3 n + k_4 a + k_5 a + k_6 b\}P_{n,a,b}(t) +$$

$$+ k_1 n(n+1)P_{n+1,a-1,b}(t) + k_2(a+1)(n-1)P_{n-1,a+1,b}(t) + k_3(n+1)P_{n+1,a-1,b}(t) \quad (10)$$

$$+ k_4(a+1)P_{n-1,a+1,b}(t) + k_5(a+1)P_{n,a+1,b-1}(t) + k_6(b+1)P_{n,a-1,b+1}(t)$$

can be transformed into a Fokker-Planck equation according to standard rules (Gardiner, 2004), and after taking care of conservation of mass law $n + a + b = T$ becomes

$$\frac{\partial P(n,a,t)}{\partial t} = \frac{\partial}{\partial n}\left[f_1(n,a)P(n,a,t)\right] + \frac{\partial}{\partial a}\left[f_2(n,a)P(n,a,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial n^2}\left[f_3(n,a)P(n,a,t)\right] +$$

$$+ \frac{\partial^2}{\partial a \partial n}\left[f_4(n,a)P(n,a,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial a^2}\left[f_5(n,a)P(n,a,t)\right] \tag{11}$$
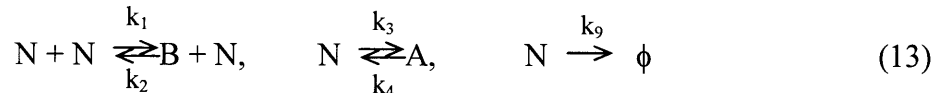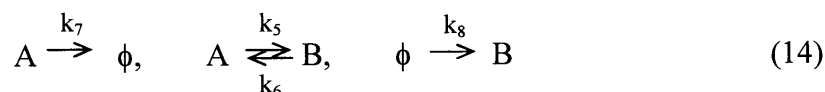
with

$$\begin{aligned}
f_1 &= k_1 n^2 - k_2 na + k_3 n - k_4 a \\
f_2 &= -k_1 n^2 + k_2 na - k_3 n + k_4 a + k_5 a - k_6(T - a - n) \\
f_3 &= k_1 n^2 + k_2 na + k_3 n + k_4 a \\
f_4 &= -k_1 n^2 - k_2 na - k_3 n - k_4 a \\
f_5 &= k_1 n^2 + k_2 na + k_3 n + k_4 a + k_5 a + k_6(T - a - n)
\end{aligned} \tag{12}$$

The time dependent partial differential equation (10) was solved numerically with reflecting boundary conditions and the steady state distribution was determined in the very long time limit. In Fig. 3.4 the numerical solution of the Fokker-Planck equation is shown along with the steady-state distribution obtained from the stochastic simulations. One can see immediately that continuous approximation correctly reproduces network topology reduction effects. An analogy can be drawn with diffusion on a surface where there is a point-like sink to understand why the Fokker Planck equation reproduces behavior that appears to arise from discreteness. Even if the surface is curved such that the mass is concentrated well away from the sink, in the very-long time limit the mass will escape through the sink due to negligible, but still non-zero, diffusive motion. The described Fokker-Planck equation has the character of diffusion in a potential well with an additional "finite" point-like sink at the boundary.

3.4 Discussion

Finally, it can be argued that the fact that both systems (4-5) and (7-9) are "closed" systems strictly obeying conservation of mass introduces additional non-linearity at the boundary that is necessary to observe the mechanism of network topology reduction. In order to address this point we have constructed "open" system exhibiting the stochastically bimodal distribution due to topology reduction:

$$N + N \underset{k_2}{\overset{k_1}{\rightleftarrows}} B + N, \qquad N \underset{k_4}{\overset{k_3}{\rightleftarrows}} A, \qquad N \overset{k_9}{\longrightarrow} \phi \tag{13}$$

$$A \xrightarrow{k_7} \phi, \quad A \underset{k_6}{\overset{k_5}{\rightleftarrows}} B, \quad \phi \xrightarrow{k_8} B \tag{14}$$

As one can see from the Fig. 3.5, two peaks are observed in the steady-state histogram of B molecules. First peak at B~37 molecules correspond to the complete network when both eqns (13) and (14) are dictating steady-state of the network. The second peak at B~91 molecule correspond to the steady-state of the reduced network which consists only of equations (14). The second peak is observed when number of molecules of N stochastically goes to zero (see inset of Fig. 3.5).

In this work we have identified the mechanism that allows multipeaked steady-state distributions for the systems without absorbing states characterized by a single deterministic attractor (e.g. chemical networks consisting of purely reversible chemical reactions that have single stable solution for their ODE chemical equations). This mechanism can be realized in both closed, mass-conserving, system and in open, steady-state systems. The network topology reduction relies on the stochastic fluctuations in the particular network architectures that allows effective reduction in the number of possible reactions due to exhaust of one of the components. The validity of continuous (Fokker-Planck) description of this mechanism was also studied.

3.5 Bibliography

Artyomov, M.N., Das, J., Kardar, M., and Chakraborty, A.K. (2007). Purely stochastic binary decisions in cell signaling models without underlying deterministic bistabilities. Proceedings of the National Academy of Sciences of the United States of America *104*, 18958-18963.

Berg, O.G., Paulsson, J., and Ehrenberg, M. (2000). Fluctuations and quality of control in biological cells: Zero-order ultrasensitivity reinvestigated. Biophysical Journal *79*, 1228-1236.

Craciun, G., Tang, Y.Z., and Feinberg, M. (2006). Understanding bistability in complex enzyme-driven reaction networks. Proceedings of the National Academy of Sciences of the United States of America *103*, 8697-8702.

Gardiner, C.W. (2004). Handbook of stochastic methods : for physics, chemistry, and the natural sciences, 3rd edn (Berlin ; New York: Springer).

Gillespie, D.T. (1977). Exact Stochastic Simulation of Coupled Chemical-Reactions. Journal of Physical Chemistry *81*, 2340-2361.

Goldbeter, A., and Koshland, D.E. (1981). An Amplified Sensitivity Arising from Covalent Modification in Biological-Systems. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences *78*, 6840-6844.

Gomez-Uribe, C., Verghese, G.C., and Mirny, L.A. (2007). Operating regimes of signaling cycles: Statics, dynamics, and noise filtering. Plos Computational Biology *3*, 2487-2497.

Kepler, T.B., and Elston, T.C. (2001). Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. Biophysical Journal *81*, 3116-3136.

Lai, K., Robertson, M.J., and Schaffer, D.V. (2004). The Sonic hedgehog signaling system as a bistable genetic switch. Biophysical Journal *86*, 2748-2757.

Levine, J., Kueh, H.Y., and Mirny, L. (2007). Intrinsic fluctuations, robustness, and tunability in signaling cycles. Biophysical Journal *92*, 4473-4481.

Lipshtat, A., Loinger, A., Balaban, N.Q., and Biham, O. (2006). Genetic toggle switch without cooperative binding. Physical review letters *96*, -.

Loinger, A., Lipshtat, A., Balaban, N.Q., and Biham, O. (2007). Stochastic simulations of genetic switch systems. Physical Review E *75*, -.

Markevich, N.I., Hoek, J.B., and Kholodenko, B.N. (2004). Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. Journal of Cell Biology *164*, 353-359.

McQuarrie, D.A. (1967). Stochastic Approach to chemical kinetics. Journal of Applied Probability *4*, 413-478.

Samoilov, M., Plyasunov, S., and Arkin, A.P. (2005). Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations. Proc Natl Acad Sci U S A *102*, 2310-2315.

Samoilov, M.S., and Arkin, A.P. (2006). Deviant effects in molecular reaction pathways. Nature Biotechnology *24*, 1235-1240.
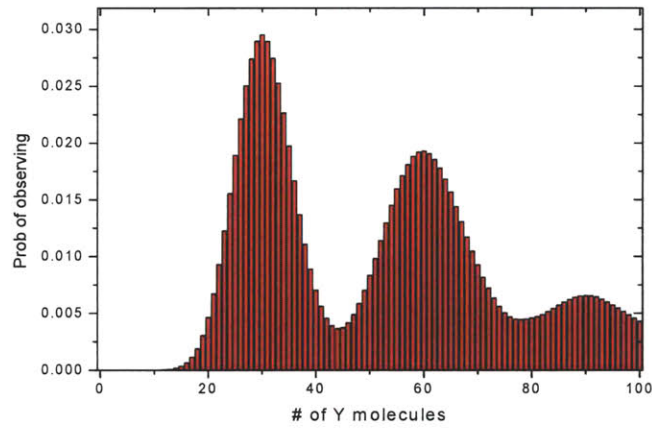
3.6 Figures for Chapter 3

Figure 3.1: Stochastic simulations for network (1-2) with parameters $kX = 0.01$, k-X=0.01, kY= 300, k-Y = 10. Steady-state distribution of molecules Y is shown.
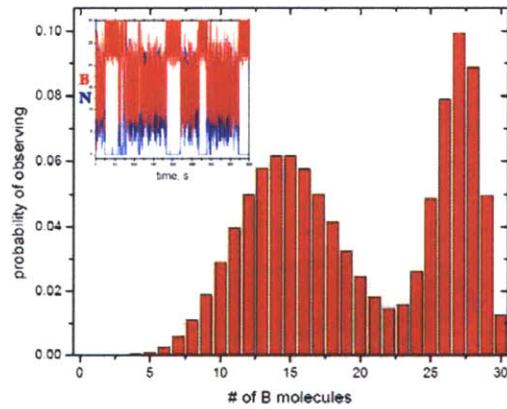

Figure 3.2: Stochastic Simulations of the network (3-4) with parameters k1=0.1; k2=10; k3=13; k4=0.03; k5=100; k6=10; N+A+B=30. Steady-state distribution of molecules B is shown. Peak at B~27 corresponds to steady-state of reduced network, while peak at B~15 corresponds to steady state of the complete network. Inset: time course of simulations shown for N (red) and B (black) species.
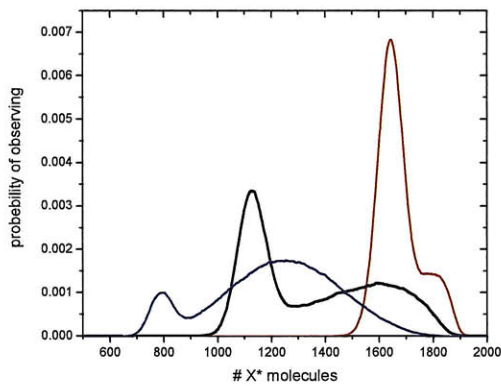
Figure 3.3: Steady-state probability distribution, for the reaction network (7-9) with parameters k1=40; k2=104; k3=104; k-1=200; k-2=100; k-3=5000; k21=10;   k-21=5; k22=10; k-22=0.2; X+X*=2000; E- = 50 (same as (Samoilov et al., 2005b)) and E+ + N = 30 (red), E+ + N = 35 (black), E+ + N = 40 (blue);.
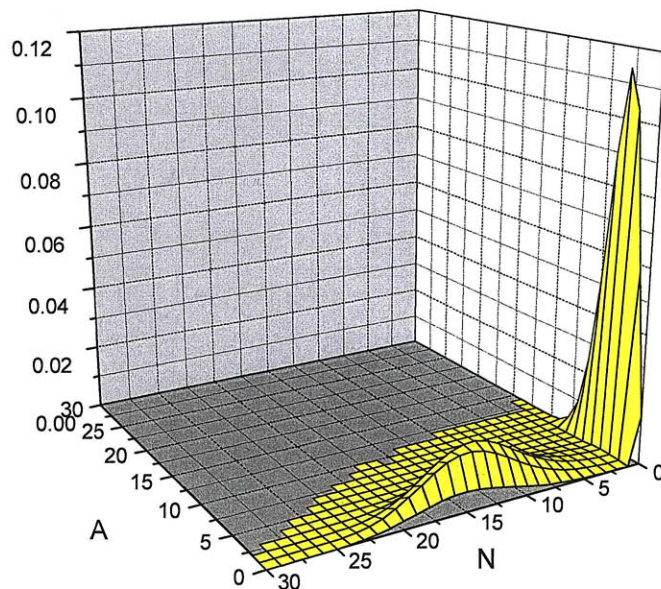
Figure 3.4: Numerical solution of the Fokker-Planck eqn (10) at infinite time, coinciding with results of stochastic simulations for network (6-8) with the same parameters as fig.3 and E+ + N = 30.

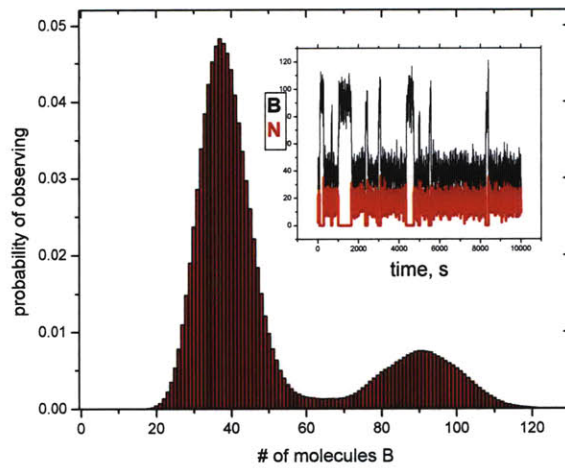**Figure 3.5:** Stochastic Simulations of the network (13-14) with parameters k1=1; k2=10; k3=13; k4=0.01; k5=100; k6=10; k7=1; k8=9; k9=0.4; Steady-state distribution of molecules B is shown. Inset: time course of simulations shown for N (red) and B (black) species.

PART II

Chapter 4

Introduction into SSC: algorithm, units conversion and examples

4.1 General description of the algorithm

In this chapter we briefly describe the simulation algorithm used for modeling reaction-diffusion processes. More general and thorough descriptions of the simulation algorithms can be found elsewhere (Lis et al., 2009). All the simulations presented in Part II were carried out with the help of SSC software ((Lis et al., 2009), see also Appendix to this Chapter) which performs spatially-resolved stochastic chemical master-equation simulations and allows efficient treatment of the combinatorial expansion problem in the size of simulation network.

Discretization is used for treatment of spatial components: space is discretized into subvolumes and all the particles withtin subvolume are assumed to be well mixed. In our simulations typical subvolume size is determined by the interaction length of proteins. We will illustrate the mechanics of SSC by example of T-cell surface proteins interacting with proteins on the surface of antigen presenting cell.

According to the experimental data interface between T-cell and APC is of the order of 1 μm^2 (Grakoui et al., 1999). In our simulation it is represented by a square of dimensions 1 μm × 1 μm (although SSC provides capabilities to explore other geometries). Simulation surface is divided into square chambers (subvolumes) of the

length L=100 angstrom (0.01 μm), which implies that we simulate a "chessboard" of 100×100 chambers. The dimensions of the individual chamber are chosen to correspond to the range of attractive interactions between T-cell membrane proteins (Yachi et al., 2006).

At any moment of time there might very well be more than one protein in each chamber. We are ignoring internal degrees of freedom and specific positions of proteins inside the chamber saying that proteins can react with each other provided that both are in the same chamber (*i.e.* they are within the range of interactions). Reactions between the species in different cells are not permitted. Diffusion of proteins is considered as a first order chemical reaction and is represented by protein hopping from one chamber to the neighboring chamber with corresponding rate constant.

The simulation consists of very large number of repetitions of two basic operations (which is very similar to original algorithm(Gillespie, 1977)):

1. Choosing subvolume where reactions take place
2. Choosing next process happening (might be reaction or diffusion – they are on the same footing)
3. Choosing the time when next process happens

And, after updating concentrations and positions, these steps are repeated all over again.

In first step, probability to choose specific subvolume is determined by a relative reaction flux in this subvolume compared to total reaction flux in all subvolumes.

We illustrate the procedure behind second step by considering the molecule A which can either diffuse to the neighboring chamber with rate constant k_{motion} or transform into molecule B (as in A→B) with rate constant $k_{A \to B}$. Probability that diffusion will occur is found with:

$$P(diffusion) = \frac{k_{motion}}{k_{motion} + k_{A \to B}} \tag{1}$$

$$P(reaction) = \frac{k_{A \to B}}{k_{motion} + k_{A \to B}} = 1 - P(diffusion) \tag{2}$$

We throw a random number within 0 to 1 interval, and if it turns out to be smaller than *P(diffusion)* then we make a diffusion move, otherwise we substitute one molecule of A with B.

The time when the process determined by above procedure occurs is defined by picking a random number from an exponential distribution $e^{-\lambda}$ with parameter $\lambda = k_{motion} + k_{A \to B}$. This implies that after throwing another random number from 0 to 1 (*rand*) we employ the formula:

$$t = \frac{1}{k_{motion} + k_{A \to B}} \ln\left(\frac{1}{rand}\right) \tag{3}$$

It is important to note, once again, that diffusive motion and chemical reactions – both first and second order, are treated on the same footing which differs from the usual way they are described in experiments: diffusion by the diffusion constant D (cm^2/s), first order reactions by rate constant k_1 (s^{-1}), second order reactions by rate constant k_2 ($M^{-1}s^{-1}$) *etc..* In a following section, we describe in detail how we shift to description where all the parameters have the same units and, thus can be compared to each other.

4.2 Rate constant unit conversion

The processes are characterized by the rate constants, but it is propensity of the reaction rather than its rate constant that enters into expression for relative probabilities. Propensity is defined as reaction flux in a given volume, i.e. number of reactions taking place every second. Thus, for

- first order reaction A→B

$$a_{A \to B} = k_{A \to B} \times (\# of \ molecules \ A \ per \ chamber), \tag{4a}$$

which for the case of the single molecule coincides with rate constant $a_{A \to B} = k_{A \to B}$. This is why expressions (1) and (2) have rate constants in them.

- second order reactions A+B→C

$$a_{A+B \to C} = k_{A+B \to C} \times (\# of \ molecules \ A \ per \ chamber) \times (\# of \ molecules \ B \ per \ chamber) \tag{4b}$$

- diffusion of molecules A out of the particular chamber

$$a_{diff \ A} = k_{diff \ A} \times (\# of \ molecules \ A \ in \ chamber) \tag{4c}$$

Note, that all propensities must have the same units, namely, $\dfrac{1}{s \times chamber \ area}$,

i.e. the physical meaning of *propensity* is the reaction flux in a individual chamber on the

surface. Unit consistency in equation (4) requires the rate constants to have the different units:

$$k_{A \to B} : \frac{1}{s} \tag{5a}$$

$$k_{A+B \to C} : \frac{chamber\,area}{molecules \times s} \tag{5b}$$

$$k_{diff\,A} : \frac{1}{s} \tag{5c}$$

One can see immediately, that units of (5b) are very different from the ones used in bulk measurements for k_2, namely $M^{-1}s^{-1}$ $\left(\frac{L}{mole \times s} \right)$. Below we show how to transform experimental value k_2 in units $M^{-1}s^{-1}$ to the computational parameter $k_{A+B \to C}$ in units $\frac{chamber\,area}{molecules \times s}$:

$$k_2 \frac{L}{mole \times s} = k_2 \frac{10^{-3}\,m^3}{6 \cdot 10^{23}\,molecules \times s} = \frac{k_2 \cdot 10^{-3}}{6 \cdot 10^{23}} \frac{10^{18}\,\mu m \times \mu m^2}{molecules \times s} =$$

$$= \frac{k_2 \cdot 10^{-3} \cdot 10^{18}}{6 \cdot 10^{23}}\,\mu m \times \frac{10^4\,chambers}{molecules \times s} = \frac{k_2 \cdot 10^{-3} \cdot 10^{18} \cdot 10^4}{6 \cdot 10^{23}}\,\mu m \times \frac{chamber}{molecules \times s} \tag{6}$$

where we have used the fact that our contact area of 1 μm^2 consists of 10^4 chambers. The last step is to convert three dimensional value of k_2 into the 2-dimensional $k_{A+B \to C}$. To do that, we have to know a characteristic length, corresponding to the confinement of the proteins to 2-dimensional membrane. Usually, we assume that membrane proteins can move in direction perpendicular to membrane within the distance of approximately d=10 Angstrom=10^{-3} μm. Thus:

$$\frac{k_2 \cdot 10^{-3} \cdot 10^{18} \cdot 10^4}{6 \cdot 10^{23}}\,\mu m \times \frac{chambers}{molecules \times s} \times \frac{1}{10^{-3}\,\mu m} = k_{A+B \to C} \frac{chambers}{molecules \times s} \tag{7}$$

$$k_{A+B \to C} = \frac{k_2 \cdot 10^{-3} \cdot 10^{18} \cdot 10^4}{6 \cdot 10^{23} \cdot 10^{-3}} \tag{8}$$

where k_2 is in units $M^{-1}s^{-1}$ and $k_{A+B \to C}$ is in units $\frac{chamber\,area}{molecules \times s}$. Note, that for each particular choice of the individual chamber size (which is determined by the range of

interactions), one should recalculate the value of $k_{A+B \to C}$ accordingly, while experimental value, naturally, does not change.

Rate of the diffusion process is estimated from the diffusion rate constant D, which is usually measured in units $\mu m^2/s$. Here, we can simply multiply by the number of chambers in 1 μm^2, which is 10^4 in this particular case:

$$k_{diff} = D \times 10^4 \; \frac{1}{s} \tag{9}$$

4.3 Appendix 1 to Chapter 4: BioInformatics Paper Describing SSC

Systems biology

Efficient stochastic simulation of reaction–diffusion processes via direct compilation

Mieszko Lis[1,*], Maxim N. Artyomov[2], Srinivas Devadas[1] and Arup K. Chakraborty[2,3,4]

[1]Computer Science and Artificial Intelligence Laboratory and the Departments of [2]Chemistry, [3]Chemical Engineering and [4]Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

ABSTRACT

We present the Stochastic Simulator Compiler (SSC), a tool for exact stochastic simulations of well-mixed and spatially heterogeneous systems. SSC is the first tool to allow a readable high-level description with spatially heterogeneous simulation algorithms and complex geometries; this permits large systems to be expressed concisely. Meanwhile, direct native-code compilation allows SSC to generate very fast simulations.

Availability: SSC currently runs on Linux and Mac OS X, and is freely available at http://web.mit.edu/irc/ssc/.

Contact: mieszko@csail.mit.edu

Supplementary Information: Supplementary data are available at *Bioinformatics* online.

1 BACKGROUND

Cells interact with their environment via receptors that bind to extracellular molecules; these events are then translated into functions by biochemical signaling networks. Non-linearities arising from the complex topology of such networks often make it difficult to intuit qualitative behavior of signaling modules. Moreover, recent imaging experiments have revealed that signaling components are organized into spatial patterns that modulate signaling (Grakoui *et al.*, 1999; Lee *et al.*, 2003). Finally, extrinsic and intrinsic stochastic effects, which make each cell's response unique, can be important when small numbers of signaling molecules are involved (Artyomov *et al.*, 2007). As computational studies are increasingly becoming necessary complements to genetic, biochemical and imaging experiments in unraveling this non-intuitive behavior of cell signaling networks, efficient and easy to use tools that can carry out stochastic simulations of biochemical networks, both in well-mixed and spatially inhomogeneous approximations, have become key technologies.

Since the original stochastic simulation algorithm (Gillespie, 1977), basic computer science techniques have reduced the rate at which the per-step computation time grows with the number of possible reactions to logarithmic growth (Gibson and Bruck, 2000; Li and Petzold, 2006; Wylie *et al.*, 2006), or optimized performance by noting that a few reactions account for most events (Cao *et al.*, 2004; McCollum *et al.*, 2006); more recently, Slepoy *et al.* (2008)

have reduced per-step computation to expected constant time via an elegant composition–rejection algorithm. Similar techniques have been applied to reduce spatially heterogeneous simulation time to logarithmic (Elf and Ehrenberg, 2004). The combinatorial growth of the instantiated reaction network size, another limiting factor for complex systems, has been addressed either by generating species and reactions on the fly (Faeder *et al.*, 2005; Lok and Brent, 2005) during a Gillespie-based simulation, by representing each molecule separately (Morton-Firth and Bray, 1998), or ingeniously do away with explicit counts altogether by adjusting the sampling distribution (Danos *et al.*, 2007; Yang *et al.*, 2008).

Efficient formulation of such simulations in a general programming language like C or FORTRAN, however, is not a trivial task: while simulating a few reactions is fast even with a simple implementation, a system with thousands of reactions and subvolumes demands more complex algorithms which are much more tricky to code. The programming burden has been reduced by libraries (e.g. Li *et al.*, 2008) as well as by simulators for well-mixed (e.g. Gillespie *et al.*, 2006; Mauch, 2009) and spatially inhomogeneous (e.g. Hattne *et al.*, 2005; Meier-Schellersheim *et al.*, 2006) models. File formats like SBML (Hucka *et al.*, 2008), developed to express biochemical models, can be read by several simulators.

The modeling task is further complicated by the explosion in combinatorial complexity which arises when modeling post-translational modification or reactions local to one molecule in a complex (Hlavacek *et al.*, 2006): in SBML (and, indeed, in most simulators) all possible species and each combination of every possible reacting complex must be written out as a separate reaction, which renders expressing even modestly complex reaction networks impractical. To mitigate these limitations, BioNetGen (Faeder *et al.*, 2009) and κ (Danos and Laneve, 2004) have proposed higher level specifications where the reactants in each reaction are written as patterns covering many possible species; such descriptions not only naturally correspond to the intuitive concept of a biochemical reaction, but are significantly smaller and therefore more readable as well as much less error-prone.

The main contribution of the Stochastic Simulation Compiler (SSC) that we present here lies in combining a higher level specification required for modeling larger systems with the ability to model spatially heterogeneous systems. It differs from BioNetGen and κ because their syntax and expansion algorithms offer no

---

*To whom correspondence should be addressed.

support for spatially inhomogeneous containers, while SSC supports multiple regions with arbitrarily complex shapes specified using Constructive Solid Geometry (CSG); meanwhile, while MesoRD allows such regions and geometries, it suffers from the combinatorial complexity limitations described above. In addition, SSC produces fast simulations (cf. Supplementary Material) by directly generating machine code tailored to a specific architecture.

2 IMPLEMENTATION

2.1 Tool flow

The tool flow resembles a programming language compiler. The user writes a high-level description of the reaction system (see Supplementary Material for examples), using patterns to select and change specific parts of compounds (similar to how a cell biologist would describe a known or hypothesized cell signaling network). Regions are specified using CSG, a technique that employs simple operations (e.g. union, intersection, difference and scale) on basic shapes (such as spheres, cubes, cylinders, etc.) to describe arbitrarily complex geometries and widely used in solid modeling and computer graphics (see, e.g. Requicha and Voelcker, 1977). Any reaction can be restricted to a subset of regions, and diffusions within and among the regions are written using the same high-level pattern syntax as reactions. The compiler then expands the model starting with initial species and reaction patterns, creating the necessary instances with specific properties and connections as well as specific reactions operating on each of those compounds in each region; meanwhile, the regions are discretized into cubic subvolumes. This intermediate representation is used to produce a simulator executable, which, in turn, simulates the model signaling pathway.

2.2 Reaction expansion

Most biologically relevant signaling reactions are conceptually *local*, that is, they 'see' only a part of a larger molecule or complex (say, a single phosphorylation site). Therefore, we write reactions and diffusions locally, using pattern matching to recognize and modify parts of complexes, and rely on the compiler to derive all the possible cases in all regions. Similarly, only initially present compounds are specified; the compiler generates the rest from the initial set and the reactions.

Formally, the reactions and diffusion form a graph term-rewriting system, which is fully evaluated to generate the simulator. Briefly, each expansion step considers a rule in the system, finding all combinations of substrates in the relevant region that match the rule. The rule is then applied to each match, possibly resulting in new compounds, and a compound-specific reaction is created for the specific substrate combination. Any new compounds not excluded by predefined limits (used to prevent infinite expansion) are added to the region where the reaction took place and any regions reachable by following the given diffusion patterns; the cycle then repeats until no more new compounds have been created. (See Supplementary Material for details of the expansion process).

2.3 Direct code generation

We obtain the efficiency of hand-optimized code by directly generating assembly code from the fully expanded set of reachable species and reactions. This allows us to avoid the interpretive

overhead of consulting dependency graphs to determine which copy counts and propensities must be recomputed.

The generated code is also tailored for model complexity and processor architecture. For most sizes, the compiler creates a separate, straight-line segment of code for each possible reaction in a region; each segment is parameterized only on the subvolume (or, in the case of diffusion, two subvolumes), and directly updates and propagates the affected propensities (see Section 2.4). This avoids pipeline stalls and cache flushes caused by mispredicted branches, and reduces the number of data memory reads and writes (which are the performance bottleneck) to the absolute minimum. (See Supplementary Material for a detailed description of the code generation method).

2.4 Reaction–diffusion simulation algorithm

The simulation algorithm is similar to the logarithmic-time versions of the direct stochastic simulation algorithm (Li and Petzold, 2006; Wylie et al., 2006). The simulation-time representation details may be found in the Supplementary Material; briefly, the reactions in each subvolume (or on each boundary between subvolumes) are arranged in an n-ary heap with the leaves corresponding to individual reaction propensities and each node carrying the combined propensity of the reactions underneath — the topmost node for each subvolume is, then, the propensity of any reaction taking place within. The subvolume and boundary reaction propensities are, in turn, themselves arranged in a heap where each leaf is either a subvolume or a boundary propensity; the topmost node is the propensity of any reaction in the system taking place (and, hence, the range from which the random number should be selected).

Simulation proceeds as follows: a random number r is selected from range $[0, R)$ where R is the propensity of any reaction taking place; then the subvolume and reaction corresponding to r is selected by n-ary search in the heap. Next, the reaction is 'executed', that is, the copy numbers of the affected species are adjusted as the reaction dictates. Finally, the propensity of each reaction whose substrate copy counts were altered is recomputed, and the partial propensities are propagated up the propensity heap until the new R is recomputed and the cycle can be repeated.

Since the propensity heap in each subvolume (or boundary) has height logarithmic in the number of reactions within, and the heap above is logarithmic in the number of subvolumes and boundaries, the total tree depth scales roughly logarithmically in the number of reactions in the system. Both the reaction selection/search and copy number/propensity update step, therefore, run in time logarithmic in the number of reactions.

3 PERFORMANCE

We compared spatially homogeneous SSC against BioNetGen 2.0.46 (Faeder et al., 2009) (since, like SSC, it builds reaction networks from pattern-matching rules), and against simulators built with the StochKit library (Li et al., 2008); because of the complexity of the larger models, we had SSC automatically generate the required StochKit C++ configurations. To test real-world performance, we selected two toy systems and two more realistic systems with various reaction counts: a dimer decay model (Gillespie, 2001) with four reactions, a simplified EGFR signaling model (Blinov et al., 2006) with 64 reactions, a model for the earliest events in

T-cell signaling (Wylie *et al.*, 2006) with 1120 reactions, and an enhanced version of the same with 2422 reactions. To test spatially heterogeneous models, we compared with the latest development revision of MesoRD (Hattne *et al.*, 2005), SVN r559; we used the T-cell signaling model above where single molecules (but not compounds) were permitted to diffuse around a membrane interface, which was divided into 100, 10 000, and 50 000 subvolumes. All simulations produced the same results (modulo random seed variation and precision loss during floating point arithmetic). To focus on measuring only the simulation time, we disabled all output except the final species counts, and repeated each experiment 5-fold to account for initial random seed variation and possible effects of other processes executing on the system.

We found that SSC consistently outperformed the faster of the two spatially homogeneous simulators we tested by 2 to 6, with the advantage growing with the size of the model (see Supplementary Fig. 3). For spatially heterogeneous simulation, we found that SSC was 50 faster than MesoRD, although both scaled very well with the number of subvolumes (see Supplementary Fig. 4).

4 CONCLUSIONS

We have described the SSC, a new tool for exact stochastic simulations of biochemical reaction networks. SSC is, to our knowledge, the first tool to combine a succinct high-level description (which avoids combinatorial complexity explosion) with spatially resolved simulation where species and reactions may be restricted to specific regions of arbitrarily complex shapes, and unique in employing direct native machine code generation to produce fast simulators.

ACKNOWLEDGEMENTS

REFERENCES

Artyomov,M.N. *et al.* (2007) Purely stochastic binary decisions in cell signaling models without underlying deterministic bistabilities. *Proc. Natl Acad. Sci. USA*, **104**, 18958.

Blinov,M.L. *et al.* (2006) A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. *Biosystems*, **83**, 136.

Cao,Y. *et al.* (2004) Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *J. Chem. Phys.*, **121**, 4059.

Danos,V. and Laneve,C. (2004) Formal molecular biology. *Theor. Comput. Sci.*, **325**, 69.

Danos,V. *et al.* (2007) Scalable simulation of cellular signaling networks. In *Proceedings of APLAS, Singapore, Nov 29 Dec 1 2007*, Springer.

Elf,J. and Ehrenberg,M. (2004) Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. *Syst. Biol.*, **1**, 230.

Faeder,J.R. *et al.* (2005) Rule-based modeling of biochemical networks. *Complexity*, **10**, 22 41.

Faeder,J.R. *et al.* (2009) Rule-based modeling of biochemical systems with BioNetGen. In Maly,I.V. (ed), *Methods in Molecular Biology. Systems Biology*, vol. 500. Humana Press, Clifton, NJ.

Gibson,M.A. and Bruck,J. (2000) Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A*, **104**, 1876.

Gillespie,C.S. *et al.* (2006) Tools for the SBML community. *Bioinformatics*, **22**, 628.

Gillespie,D.T. (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, **81**, 2340.

Gillespie,D.T. (2001) Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.*, **115**, 1716.

Grakoui,A. *et al.* (1999) The immunological synapse: a molecular machine controlling T cell activation. *Science*, **285**, 221.

Hattne,J. *et al.* (2005) Stochastic reaction-diffusion simulation with MesoRD. *Bioinformatics*, **21**, 2923.

Hlavacek,W.S. *et al.* (2006) Rules for modeling signal transduction systems. *Science STKE*, **344**, re6.

Hucka,M. *et al.* (2008) Systems biology markup language (SBML) level 2: structures and facilities for model definitions. *Nat. Prec.* [Epub ahead of print, doi: 10.1038/npre.2008.2715.1, November 5, 2007].

Lee,K.H. *et al.* (2003) The immunological synapse balances T cell receptor signaling and degradation. *Science*, **302**, 1218.

Li,H. and Petzold,L.R. (2006) Logarithmic direct method for discrete stochastic simulation of chemically reacting systems. *Technical report*, UCSB Computer Science and Engineering Group, Santa Barbara, CA.

Li,H. *et al.* (2008) Algorithms and software for stochastic simulation of biochemical reacting sytems. *Biotechnol. Prog.*, **24**, 56 61.

Lok,L. and Brent,R. (2005) Automatic generation of cellular reaction networks with Moleculizer 1.0. *Nat. Biotechnol.*, **23**, 131 136.

Mauch,S. (2009) Cain: stochastic simulations for chemical kinetics. Available at http://cain.sourceforge.net/ (last accessed date May 10, 2009).

McCollum,J.M. *et al.* (2006) The sorting direct method for stochastic simulation of biochemical systems with varying reaction execution behavior. *Comput. Biol. Chem.*, **30**, 39.

Meier-Schellersheim,M. *et al.* (2006) Key role of local regulation in chemosensing revealed by a new molecular interaction based modeling method. *PLOS Comput. Biol.*, **2**, e82.

Morton-Firth,C.J. and Bray,D. (1998) Predicting temporal fluctuations in an intracellular signalling pathway. *J. Theor. Biol.*, **192**, 117 128.

Requicha,A. and Voelcker,H. (1977) Constructive solid geometry. *Technical Memorandum no. 25*, Production Automation Project, University of Rochester, Rochester, NY.

Slepoy,A. *et al.* (2008) A constant-time kinetic Monte Carlo algorithm for simulation of large biochemical reaction networks. *J. Chem. Phys.*, **128**, 205101.

Wylie,D.C. *et al.* (2006) A hybrid deterministic-stochastic algorithm for modeling cell signaling dynamics in spatially inhomogeneous environments and under the influence of external fields. *J. Phys. Chem. B*, **110**, 12749.

Yang,J. *et al.* (2008) Kinetic Monte Carlo method for rule-based modeling of biochemical networks. *Phys. Rev. E*, **78**, 031910.

4.4 Appendix 2 to Chapter 4: Examples of SSC codes

This appendix is a result of work of summer research student Mykyta Artomov, who was testing SSC under my supervision.

The reaction networks treated in this appendix are exactly same networks initially simulated in famous Gillespie 1977 paper (Gillespie, 1977).

1. Isomerisation reaction
Consider the following reaction:
$$S1 \longrightarrow Z \qquad \text{with reaction rate c}$$

1a. Parameter set 1: c = 0.5, S1 = 1000, time step = 0.1

Create a file isomerisation.rxn, and put the SSC code into it:
```
region World
     box width 1 height 1 depth 1
subvolume edge 1
rxn x:S1 at 0.5 -> destroy x

new S1 at 1000
record S1
```
To compile the code, run:
```
[zxcv@zxcv]$ ssc  isomerisation.rxn
```
To start the simulation:
```
[zxcv@zxcv]$ ./isomerisation –e 10 –t 0.1
```

-e 10 - end the simulation after time 10
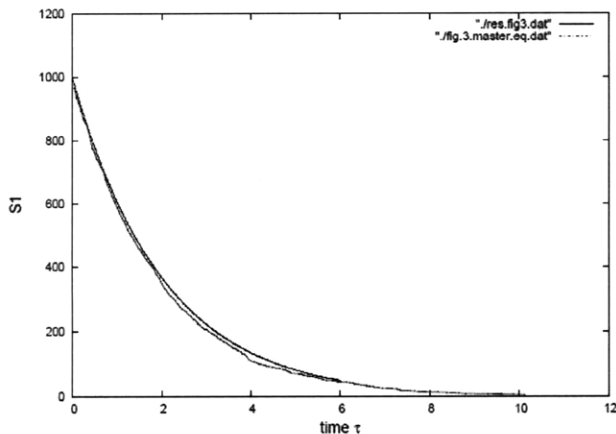-t 0.1 – the time step of output equals 0.1

Fig.1 Smooth line – plot of master equation solution, oscillating line – stochastic solution

1b Parameter set 2: c = 0.5 S1 = 5000, time step = 0.1

Do the same procedure as in 1.1., but notice, that initial number of S1 species is different.
SSC code:
```
region World
    box width 1 height 1 depth 1
subvolume edge 1
rxn x:S1 at 0.5 -> destroy x

new S1 at 5000
record S1
```


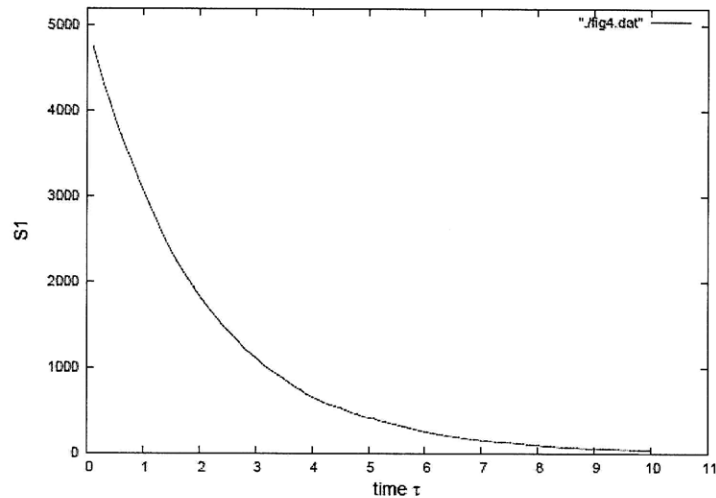**Fig.2** Simulation results for c=0.5, S10 = 5000

1c Parameter set.3. c = 0.5 S1 = 10000, time step = 0.1
SSC code:
```
region World
    box width 1 height 1 depth 1
subvolume edge 1
rxn x:S1 at 0.5 -> destroy x

new S1 at 10000
record S1
```
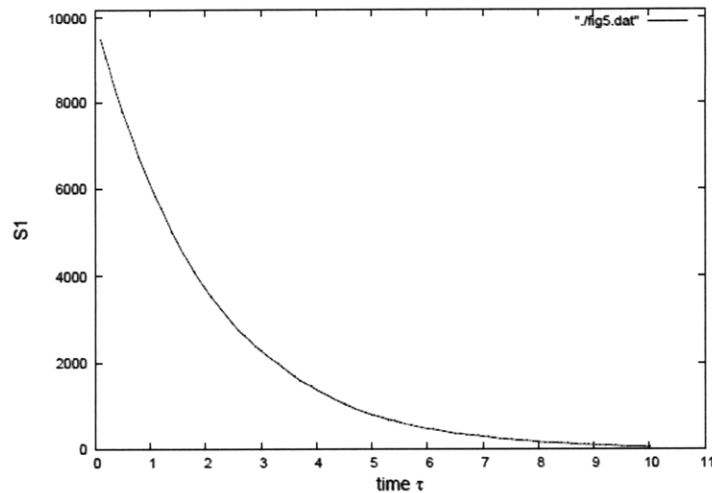
Fig.3 Simulation results for c = 0.5, S10 = 10000

The more initial number of S1 species is, the smoother line you have and it is similar to master equation solution.

2. Autocatalytic reaction

$$S1 + S2 \longrightarrow 2S2 \text{ with reaction rate } c1$$
$$2S2 \longrightarrow Z \text{ with reaction rate } c2$$
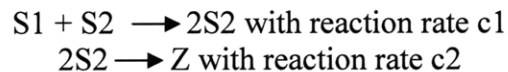
Create a file autocatalytic.rxn and put into it one of the described SSC codes.

2.1. c1*S1=5 , c2=0.005, S2=10 and S2=3000, time step=0.1. Deterministic steady state S2=1000

SSC code:
```
region World
      box width 1 height 1 depth 1
subvolume edge 1
rxn x:S1 y:S2 at 5 -> new y
rxn x:S2 y:S2 at 0.005 -> destroy x; destroy y

new S1 at 1
new S2 at 10
record S2
```

SSC code S2=3000:
```
region World
      box width 1 height 1 depth 1
subvolume edge 1
rxn x:S1 y:S2 at 5 -> new y
rxn x:S2 y:S2 at 0.005 -> destroy x; destroy y
```

89

```
new S1 at 1
new S2 at 3000
record S2
```

To compile the code, run:
```
[zxcv@zxcv]$ ssc autocatalytic.rxn
```

To start the simulation:
```
[zxcv@zxcv]$ ./autocatalytic –e 20 –t 0.1
```
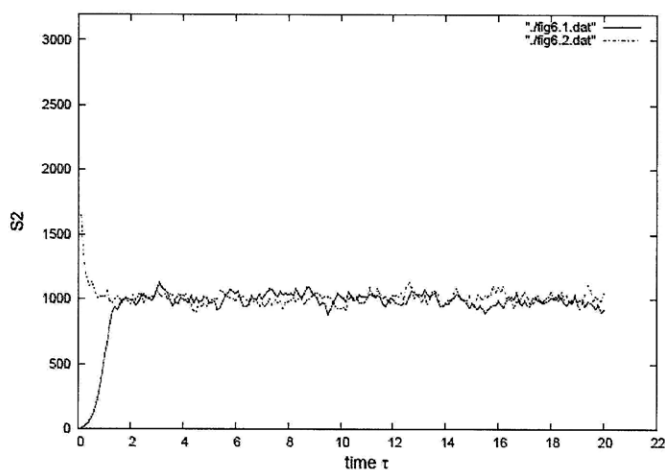

Fig. 4 Simulation results for autocatalytic reaction: $c1S1 = 5$, $c2 = 0.005$, $S20 =$ 10 - line, $S20 = 3000 -$ dashed line

To build a distribution in steady state:
```
[zxcv@zxcv]$ ./autocatalytic –e 50 –d 5
```
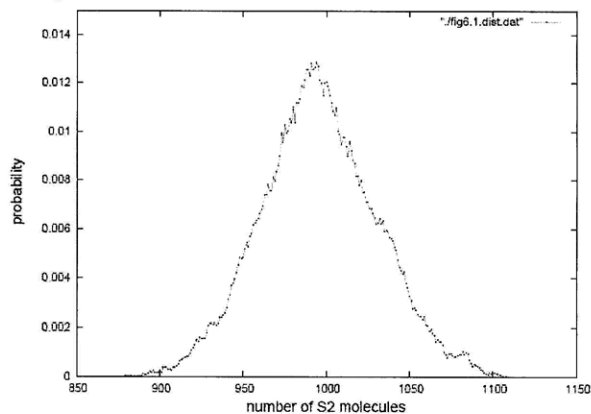

Fig.5. S2 species distribution

3. Lotka reactions

$$S1 + S2 \longrightarrow 2S2 \quad c1$$

$$S2 + S3 \longrightarrow 2S3 \qquad c2$$
$$S3 \longrightarrow Z \qquad c3$$
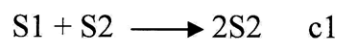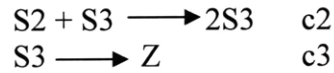
Create the file lotka.rxn and put into it one of the following SSC codes.

Parameter Set: c1*S1=10, c2=0.01, c3=10, S2=1000, S3=1000, time step=0.1.
Deterministic steady state S2=S3=1000.
SSC code:

```
region World
     box width 1 height 1 depth 1
subvolume edge 1
rxn x:S1 y:S2 at 10 -> new y
rxn x:S2 y:S3 at 0.01 -> destroy x; new y
rxn x:S3 at 10 -> destroy x


new S1 at 1
new S2 at 1000
new S3 at 1000
record S2
record S3
```

To compile the code, run:
```
[zxcv@zxcv]$ ssc  lotka.rxn
```

To start the simulation:
```
[zxcv@zxcv]$ ./lotka –e 25 –t 0.1
```
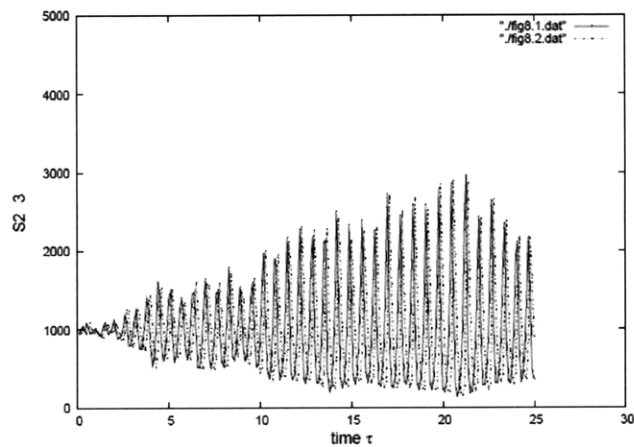

Fig.6. S2 – smooth line, S3 – dashed line. c1S1=10, c2=0.01, c3=10, S2=1000, S3=1000
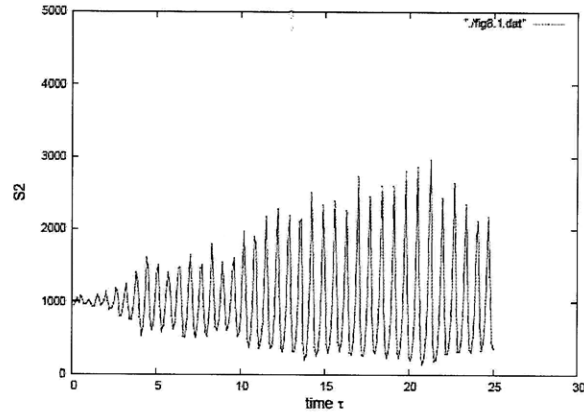
Fig.7. S2 time dependence. c1S1 = 10, c2 = 0.01, c3 = 10, S20 = 1000, S30 = 1000
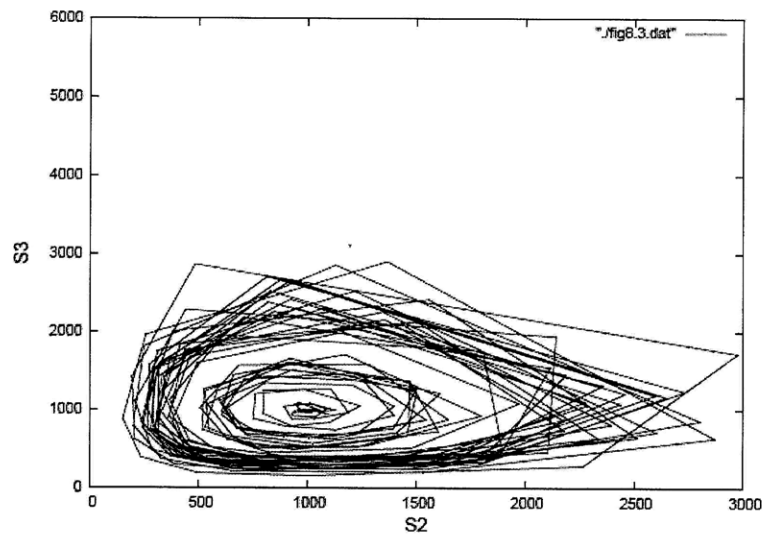
Fig.8. Phase diagram

To build a distribution in steady state:
```
[zxcv@zxcv]$ ./lotka –e 50 –d 15
```
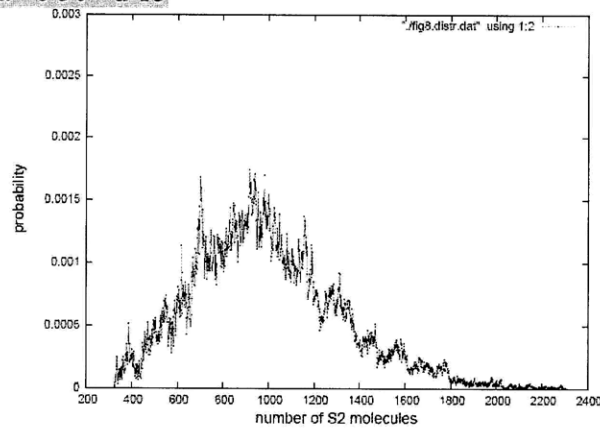

Fig.9. S2 species distribution
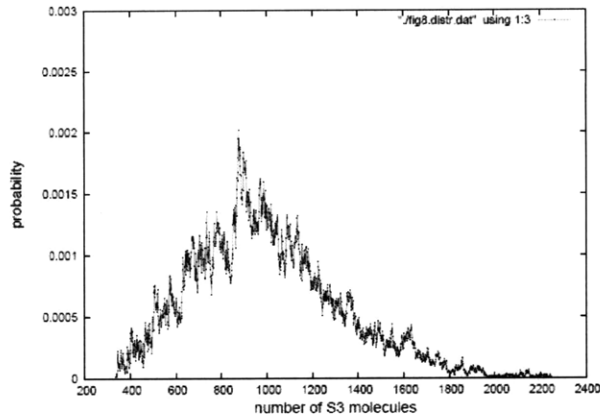
Fig.10. S3 species distribution

4. The Brusselator scheme

$$X1 \longrightarrow Y1, c1$$
$$X2 + Y1 \longrightarrow Y2 + Z1, c2$$
$$2Y1 + Y2 \longrightarrow 3Y1, c3$$
$$Y1 \longrightarrow Z2, c4$$
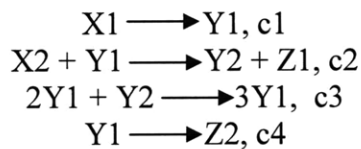
Create the file brusselator.rxn and put into it one of the following SSC codes

Initial conditions: c1*X1=5000, c2*X2=50, c3=0.00005, c4=5, Y1=1000, Y2=2000 (deterministic steady state), time step=0.01.

SSC code:
```
region World
      box width 1 height 1 depth 1
subvolume edge 1
rxn x:X1 at 0.00001 -> new Y1
rxn x:X2 y:Y1 at 0.00001 -> destroy y; new Y2
rxn x:Y1 y:Y2 z:Y1 at 0.00005 -> destroy y; new x
rxn x:Y1 at 5 -> destroy x

new X1 at 500000000
new Y1 at 1000
new X2 at 5000000
new Y2 at 2000
record Y1
record Y2
```

To compile the code, run:
```
[zxcv@zxcv]$ ssc  brusselator.rxn
```

To start the simulation:
```
[zxcv@zxcv]$ ./brusselator –e 20 –t 0.01
```

93
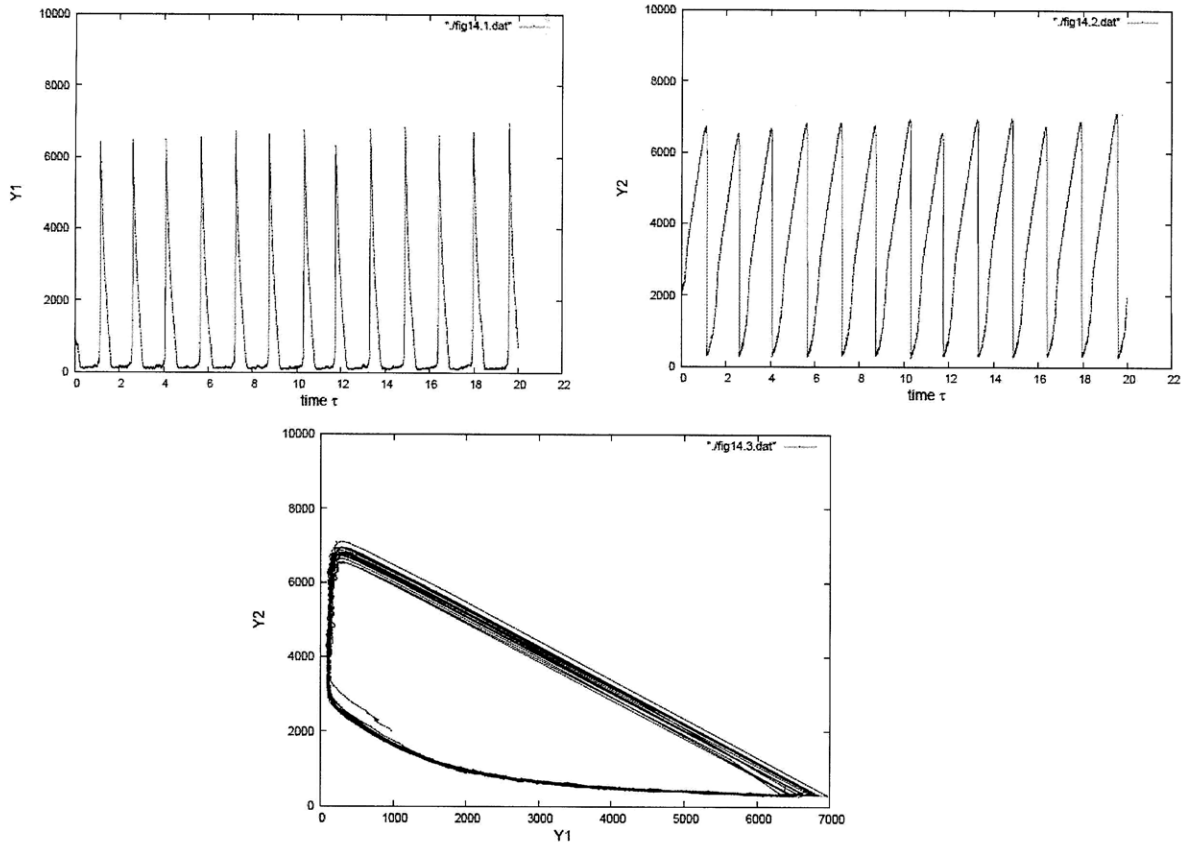
Fig.11. Results stochastic simulations run of the Brusselator reactions, with c1*X1 = 5000, c2*X2 = 50, c3 = 0.00005, c4 = 5, Y1 = 1000, Y2 = 2000
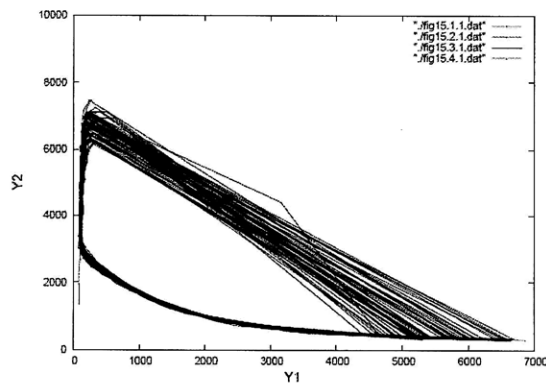
Fig.12. Plot of four stochastic simulation runs, each having the same reaction parameters as in fig.19, but different initial values of Y1 and Y2.

To build a distribution:
[zxcv@zxcv]$ ./brusselator –e 250 –d 50
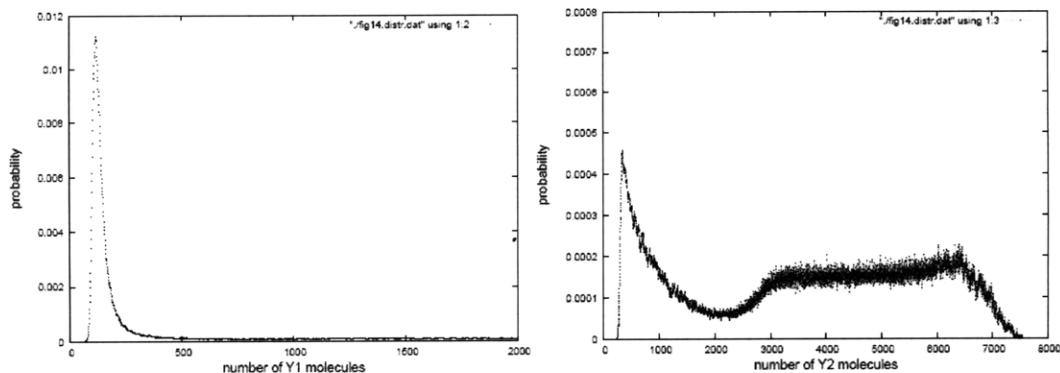
Fig.13. Y1,Y2 species distributions

5. The Oregonator scheme

$$X1 + Y2 \longrightarrow Y1 \quad c1$$
$$Y1 + Y2 \longrightarrow Z1 \quad c2$$
$$X2 + Y1 \longrightarrow 2Y1 + Y3 \quad c3$$
$$2Y1 \longrightarrow Z2 \quad c4$$
$$X3 + Y3 \longrightarrow Y2 \quad c5$$
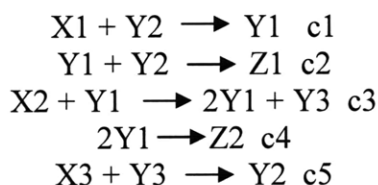
Create the file oregonator.rxn and put into it one of the following SSC codes

Initial conditions: $c1 = 0.001$, $c2 = 0.1$, $c3 = 0.1$, $c4 = 0.016$, $c5 = 0.1$, $X1 = 2000$, $X2 = 1040$, $X3 = 260$, $Y1 = 500$, $Y2 = 1000$, $Y3 = 2000$.

SSC code:
```
region World
    box width 1 height 1 depth 1
subvolume edge 1
rxn x:X1 y:Y2 at 0.001 -> destroy y; new Y1
rxn x:Y1 y:Y2 at 0.1 -> destroy x; destroy y
rxn x:X2 y:Y1 at 0.1 -> new y; new Y3
rxn x:Y1 y:Y1 at 0.016 -> destroy x; destroy y
rxn x:X3 y:Y3 at 0.1 -> destroy y; new Y2

new X1 at 2000
new Y1 at 500
new X2 at 1040
new Y2 at 1000
new X3 at 260
new Y3 at 2000
record Y1
record Y2
record Y3
```

To compile the code, run:
[zxcv@zxcv]$ ssc oregonator.rxn
To start the simulation:
[zxcv@zxcv]$ ./oregonator –e 25 –t 0.01
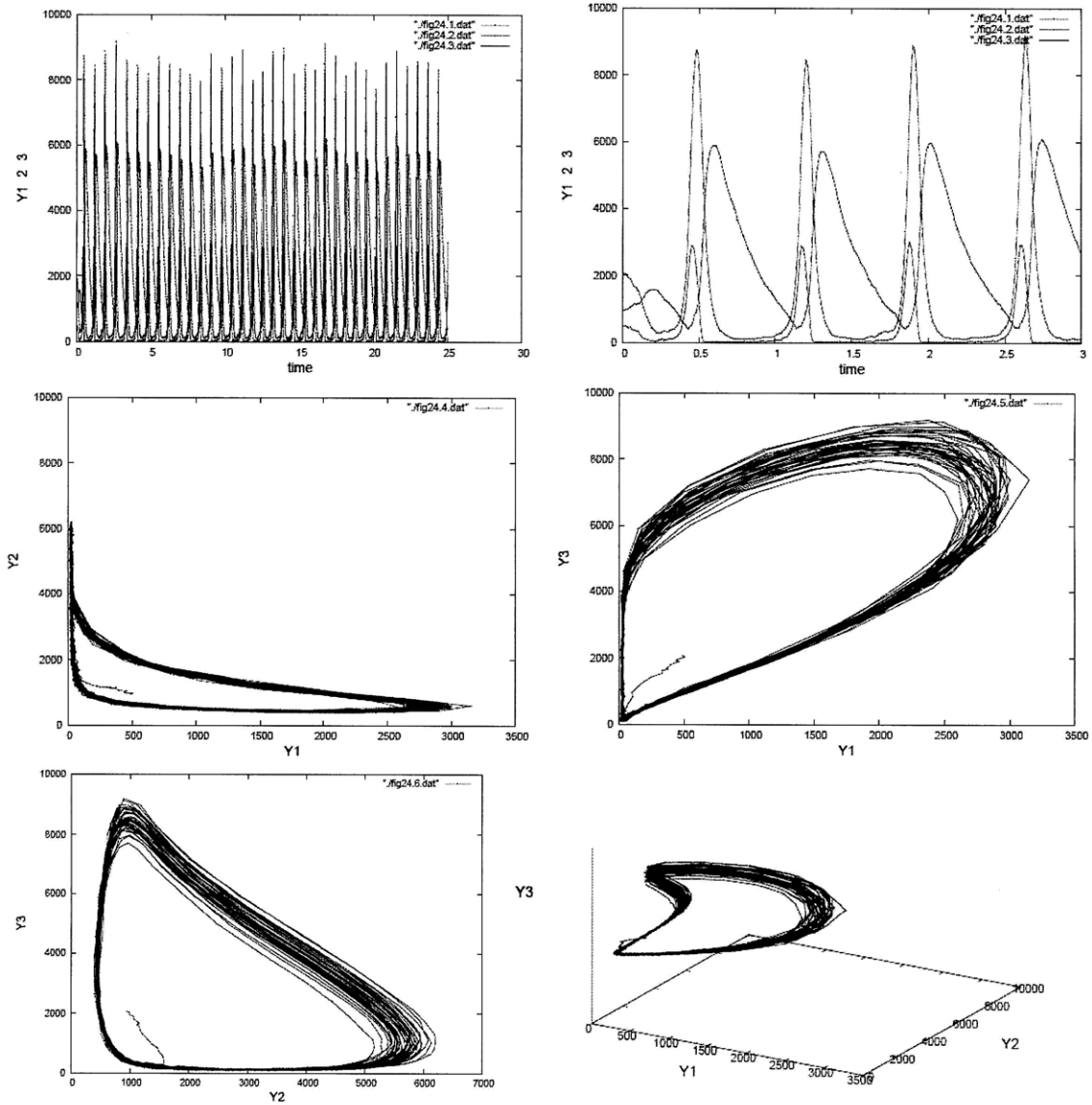
Fig.14. The Oregonator scheme simulation results. c1 = 0.001, c2 = 0.1, c3 = 0.1, c4 = 0.016, c5 = 0.1, X1 = 2000, X2 = 1040, X3 = 260, Y1 = 500, Y2 = 1000, Y3 = 2000
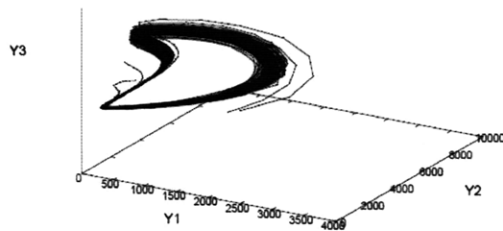
Fig.15. Plot of four simulations runs same as fig.24, but different initial values of Y1,Y2,Y3

To build a distribution in steady state:
```
[zxcv@zxcv]$ ./oregonator –e 100 –d 15
```
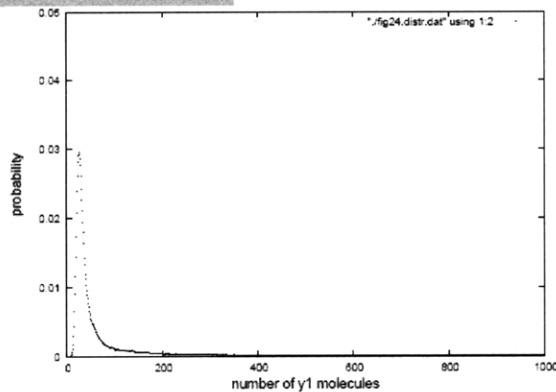

Fig.16. Y1 species distribution
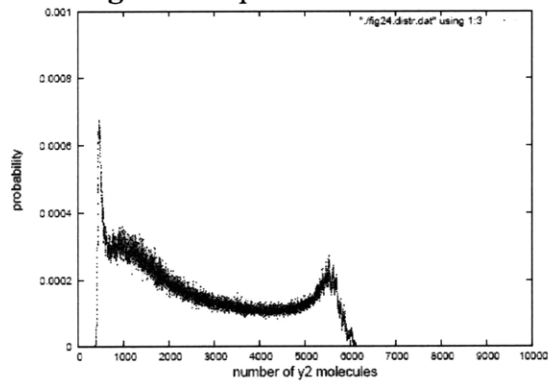
Fig.17. Y2 species distribution
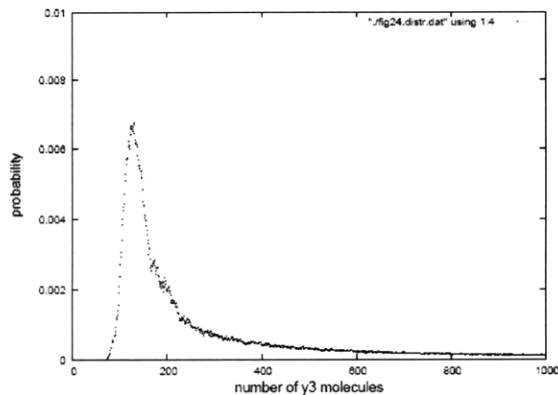
Fig.18. Y3 species distribution = 10000, Y1 = 2000, Y2 = 4000, Y3 = 8000

97

4.5 Bibliography

Gillespie, D.T. (1977). Exact Stochastic Simulation of Coupled Chemical-Reactions. Journal of Physical Chemistry *81*, 2340-2361.

Grakoui, A., Bromley, S.K., Sumen, C., Davis, M.M., Shaw, A.S., Allen, P.M., and Dustin, M.L. (1999). The immunological synapse: a molecular machine controlling T cell activation. Science (New York, N.Y *285*, 221-227.

Lis, M., Artyomov, M.N., Devadas, S., and Chakraborty, A.K. (2009). Efficient stochastic simulation of reaction-diffusion processes via direct compilation. Bioinformatics (Oxford, England) *25*, 2289-2291.

Yachi, P.P., Ampudia, J., Zal, T., and Gascoigne, N.R.J. (2006). Altered Peptide Ligands Induce Delayed CD8-T Cell Receptor Interaction--a Role for CD8 in Distinguishing Antigen Quality. Immunity *25*, 203-211.

Chapter 5

Dissecting the role of CD4 and CD8 co-receptors in T cell signaling: A puzzle resolved?

5.1 Introduction

CD4 and CD8 are membrane proteins expressed on T helper cells and cytotoxic T lymphocytes, respectively, that are known to augment the sensitivity and response of T cells to cognate pMHC ligands (Holler and Kranz, 2003; Li et al., 2004a). Textbooks ascribe the ability of these co-receptors to enhance T cell responses to two main effects: 1] Binding of CD4 and CD8 to MHC class II and class I molecules helps stabilize weak TCR-pMHC interactions. 2] The Src kinase, Lck, which is bound to the cytoplasmic tail of co-receptors, is efficiently recruited to the TCR complex upon co-receptor binding to the MHC, thereby enhancing the initiation of TCR signaling (Janeway et al., 2008).

Surface Plasmon resonance (SPR) experiments show that the half-life characterizing co-receptor-MHC interactions is shorter than 35 milliseconds (off-rate greater than 20 s^{-1}, the resolution of SPR instruments) for both CD4 and CD8 (Gao et al., 2002; Wyer et al., 1999; Xiong et al., 2001)). It is difficult to understand how the two effects noted above can be potentiated by such fleeting interactions. For example, consider the effect of co-receptor-MHC interactions in stabilizing the TCR-pMHC complex. A typical good agonist pMHC ligand is bound to TCR for approximately 10 000 milliseconds (corresponding to an off-rate of 0.1 s^{-1}) (Stone et al., 2009). This implies that during the lifetime of the TCR-pMHC bond a co-receptor would disengage from MHC approximately 1000 times, thereby making stabilization of TCR-pMHC interactions via co-receptor-MHC binding unlikely.

Yet, expression of CD8 was found to stabilize pMHC binding to CD8+ T-cell surfaces (Luescher et al., 1995; Wooldridge et al., 2005), and augment sensitivity (Holler

and Kranz, 2003). In contrast, past studies (Hamad et al., 1998) and recent in situ measurements at intercellular junctions show that CD4 does not stabilize the interactions of TCR with class II pMHC molecules (Huppa et al., Nature 2009). But, CD4 does enhance the sensitivity of T helper cells (Huppa et al., Nature 2009; Li et al., 2004). As the binding affinity of CD4 for the MHC ectodomain is just 2-4 times weaker than that characterizing CD8-MHC interactions (Gao et al., 2002), and the half-lives of both co-receptor MHC interactions are a 1000-fold shorter than the TCR-agonist pMHC bond, these results are difficult to reconcile. To shed light on this puzzle and to understand the potentially different ways in which CD4 and CD8 may augment TCR signaling, we carried out computer simulations of the well-established earliest events in TCR signaling.

5.2 Simulation results

We carried out computer simulations of a T cell-APC interface of 1 μm^2 area contained 300 T-cell receptors, 100 co-receptors and 100 pMHC complexes (these concentrations are typical for in vitro experiments, (Grakoui et al., 1999). The biochemical reactions that could occur upon the appropriate proteins encountering each other were (Scheme 1 in Fig. 5.1): TCR-pMHC binding and unbinding, co-receptor binding to MHC and coreceptor binding to TCR via Lck (Fig. 5.1). TCR, co-receptors, and pMHC were allowed to diffuse on the T cell and APC surface. In order to determine the apparent dissociation rate of pMHC molecules off the T-cell surface we initialized the simulation with 100 pMHC proteins bound to TCRs. We then simulated the biochemical reactions noted above using the Stochastic Simulation Compiler (Lis et al., 2009) that allows efficient implementation of the Gillespie algorithm (Gillespie, 1977) to study cell signaling processes, including protein motion and stochastic effects (details in SI). Several replicate simulations for each scenario were carried out, and average values of the dissociation rate of pMHC proteins in the presence and absence of coreceptors were obtained.

If a pMHC molecule dissociated from the T cell surface during the simulation, we removed it, thereby preventing rebinding. This mimics experiments where antibodies are used to achieve the same end. The parameters used to simulate situations with and without co-receptors were identical (Table 1), and we studied the effects of varying the

dissociation rate of the coreceptor-MHC bond (k_{off}), keeping the on-rate the same. These calculations aimed to explore whether the measured higher affinity of CD8 for MHC class I proteins, compared to CD4-MHC class II interactions, could explain why CD8, but not CD4, is observed to stabilize TCR-pMHC interactions (Hamad et al., 1998; Huppa, 2009; Luescher et al., 1995; Wooldridge et al., 2005).

Based on arguments noted above, we expected differences in the half-life of co-receptor-MHC interactions to have a minimal effect on the dissociation rate of pMHC molecules from the T cell surface. However, our results (Fig. 5.2a) indicate that this is not necessarily true. For co-receptor – MHC interactions with a k_{off} value of the order of 20 s^{-1}, the effective half-life of pMHC molecules bound to the T cell surface is enhanced by about a factor of 1.5. However, a 4-fold lower co-receptor-MHC affinity ($k_{off} \sim 80$ s^{-1}), results in an effective dissociation rate from the T cell surface that is indistinguishable from simulation results without the co-receptor. These results recapitulate the experimental observation that CD4, which binds MHC class II proteins with a 2 to 4-fold lower affinity compared to CD8 binding to MHC class I (Gao et al., 2002), does not stabilize the TCR-pMHC bond, but CD8 does (Hamad et al., 1998; Huppa, 2009; Luescher et al., 1995; Wooldridge et al., 2005). This is pleasing because it shows that experimental results for CD4 and CD8 (Hamad et al., 1998; Huppa, 2009; Luescher et al., 1995; Wooldridge et al., 2005) are not in conflict with each other.

However, what was wrong with the argument made earlier which suggested that the fleeting interactions between MHC proteins and CD4 or CD8 could not stabilize TCR-pMHC bonds? That argument did not account for the fact that co-receptor associated Lck also interacts with TCR. Thus, the co-receptor's interactions with a TCR-pMHC complex are bivalent, with one arm binding to MHC and the other to TCR (cartoon on right in Fig.5.1). Thus, if the co-receptor dissociates from MHC, it is still bound via Lck to the TCR-pMHC complex, and this enables rapid rebinding to MHC. A similar effect is in play if Lck dissociates from the TCR. Our results show that such cooperative interactions can cause co-receptor-mediated stabilization of the TCR-pMHC bond only if co-receptor-MHC interactions have a half-life not much larger than 20 s^{-1} (fig. S5.3).

Although CD4 does not stabilize TCR-pMHC interactions (Hamad et al., 1998;

Huppa, 2009), like CD8, it does enhance T cell responses (Huppa, 2009; Li et al., 2004a). This raises the question of whether CD4 can enhance Lck recruitment to the TCR complex, and more generally, whether CD4 and CD8 enhance T cell sensitivity to antigen in different ways.

To examine these issues, we carried out computer simulations of the type used to obtain the results in Fig. 5.2a except that the set of possible biochemical reactions was augmented to include the earliest event in TCR signaling, phosphorylation of the TCR ITAMs by Lck (scheme 2 in Fig.5.1). Multiple phosphorylation states of the ITAMs on ζ-chains of the TCR were represented by two phosphorylation states - partially and fully phosphorylated TCRs (see, for example, (Wylie et al., 2007b)). To assess the role of the co-receptor in TCR triggering, we studied two situations: 1] Lck is present as a free membrane-associated molecule and there are no co-receptors; 2] Lck is associated with the co-receptor. We carried out computer simulations for these two situations, with all parameters being identical (Table 1), and compared the levels of TCR phosphorylation (readout of signal strength). Varying the off-rate of coreceptor-MHC interactions did not affect the qualitative results of these simulations (Fig. S5.2).

Fig. 5.2b shows simulation results of signal strength as a function of the off-rate characterizing TCR-pMHC interactions. The simulations correctly recapitulate experimental observations in that TCR phosphorylation discriminates between stimulatory and non-stimulatory ligands. The border between stimulating and non-stimulating peptides ($\sim$0.1-1 s^{-1} in the simulations carried out with co-receptors present, black line in Fig. 5.2b) is dependent on the on-rate of the TCR-pMHC interaction. The simulations correspond to this rate being 10^4 M^{-1}s^{-1}, as is experimentally measured for typical good agonists. But, for higher value of the on-rate, pMHC ligands that bind TCR with off-rates larger than 1 s^{-1} are stimulatory.

Importantly, co-receptors clearly enhance TCR phophorylation. As shown in Fig. 5.2b, if a threshold amount of TCR phosphorylation is required for downstream digital signaling modules (Altan-Bonnet and Germain, 2005; Das et al., 2009; Reyes et al., 2005) to be activated (resulting in T cell responses), peptides that bind to TCR with off-rates in the range of 0.04-0.2 s^{-1} are stimulatory only when the co-receptor is present, but barely stimulate TCR phosphorylation without the co-receptor. pMHC ligands that bind

TCR with longer half-lives are stimulatory even withot the co-receptor. This is consistent with reports of co-receptor-dependent and independent ligands in both CD4 and CD8 systems (Holler and Kranz, 2003; van Bergen et al., 2001). Dose-response curves obtained from the simulations further support this point (Fig. S5.1).

We used the computational models to parse the relative contributions of TCR-pMHC stabilization and Lck recruitment to coreceptor-mediated signal enhancement in a way that is difficult to accomplish experimentally. We carried out computer simulations where the only effect of co-receptors was to enhance Lck recruitment. This was achieved by simulating systems where ITAM phosphorylation was allowed only if TCR and pMHC were directly bonded, and not if they were a part of the TCR-pMHC-co-receptor complex. These simulations showed results (Fig. 5.2c) similar to those in Fig. 5.2b. We also carried out computer simulations where the only effect of the co-receptor was to enhance the stability of the TCR-pMHC bond by a factor of 1.5 (as per results in Fig. 5.2a for k_{off} = 20 s^{-1}, corresponds to CD8). This was accomplished by simulating situations without the co-receptor (no enhancement of Lck recruitment), but with TCR-pMHC half-lives enhanced to mimic co-receptor mediated stabilization. The results (blue line, Fig. 5.2c) show that stabilization of the TCR-pMHC bond makes a minor contribution to co-receptor-mediated enhancement of TCR signaling. Thus, even for CD8, the main effect of the co-receptor is to enhance Lck recruitment to the TCR complex, and for CD4 it is the only effect.

How can fleeting CD4-MHC interactions, which do not stabilize the TCR-pMHC bond, enhance Lck recruitment? To answer this question, we analyzed the simulation results using the following simple arguments and calculations.

When two membrane proteins capable of binding approach within the range of interactions, there are two possible outcomes: the proteins associate and form a complex, or they diffuse away from each other and leave the domain of interactions. This is because there are two driving forces, attractive interactions pulling them together and random diffusive forces pushing proteins apart from each other. As both processes are stochastic, each possible outcome has a certain probability of occurrence. We calculated the probability for Lck association with the TCR complex for the case when Lck is co-receptor associated and when it is not. Differences in the probability with which Lck

binds to the TCR and its mechanistic origin shed light on how the co-receptor enhances Lck recruitment.

When Lck is not co-receptor-associated, vicinal TCR and Lck can either bind or diffuse away. The on-rate of Lck-TCR association can be estimated in the following way. It has to be large enough for the time required for TCR-Lck association to be longer than the lifetime of strong agonist pMHC-TCR complexes, otherwise these ligands would not trigger TCR in a co-receptor-independent way (Fig. 5.2b and (Holler and Kranz, 2003)). Also, the time required for TCR-Lck association must be shorter than the lifetime of endogenous pMHC-TCR bonds in order to prevent frequent spurious triggering of TCRs bound to endogenous ligands. These considerations imply that $k_{off,Ag} < k_{on,Lck} < k_{off,En}$. Based on measurements of TCR-pMHC off-rates, we took $k_{on,Lck} \sim 1 \ s^{-1} \ molec^{-1} (area\ of\ interaction)$. Effects of variations in $k_{on,Lck}$ are detailed in *appendix to chapter 5* (see Fig. S5.4). The diffusion constant of membrane-associated proteins is typically 0.01 μm^2/s (Dushek et al., 2008). Assuming the range of interactions to be of the order 100 angstroms (0.01 μm)(Yachi et al., 2006), we can compute the rate with which membrane proteins will leave the range of interactions due to diffusion to be $k_{motion} \approx 100 \ s^{-1}$. To find the probabilities escape versus binding, we compare the rates of motion and binding to obtain:

$$P(escape) = \frac{k_{motion,Lck} + k_{motion,TCR}}{k_{motion,Lck} + k_{motion,TCR} + k_{on,Lck}} = \frac{200}{201} = 0.995 \tag{1}$$

$$P(binding) = \frac{k_{on,Lck}}{2k_{motion} + k_{on,Lck}} = \frac{1}{201} = 0.005 \tag{2}$$

These estimates suggest that it is rather unlikely that Lck will form a bond with the TCR when it is not associated with the co-receptor.

Now consider the situation when Lck is associated with the co-receptor. Once the co-receptor is in the vicinity of a TCR-pMHC complex, it can either diffuse away or bind to the MHC. Experimental measurements (Gao et al., 2002) estimate the on-rate for co-receptor-MHC interactions to be very large - $k_{on,MHC-CD4(8)} > 10^5 \ M^{-1}s^{-1}$. This leads to the following estimate for the 2-dimensional value of this rate parameter (*appendix to*

chapter 5): $k_{on,MHC-CD4(8)}\ (2D) = 1670\ \dfrac{(area\ of\ interaction)}{molec \times s}$. Notice that this is 16 times

larger than the rate parameter corresponding to diffusive motion of proteins away from each other ($k_{motion} \sim 100$ s^{-1}). This suggests that the large on-rate for co-receptor-MHC interactions will combat diffusive forces effectively, enabling co-receptor binding to the MHC with high probability. Also, once this bond between two proteins anchored to apposed membranes is established, diffusion of the co-receptor and the MHC will be severely slowed down compared to a protein on a single membrane. These effects allow the co-receptor to effectively localize Lck to the TCR-pMHC complex.

To estimate the likelihood of successful Lck-TCR association given that the TCR-MHC-coreceptor complex is assembled, we must consider the following possible events: breaking TCR-MHC bond ($k_{off} \sim 1$ s^{-1}), breaking MHC-coreceptor bond ($k_{off} \sim 20$ s^{-1}), forming the TCR-Lck bond ($k_{on} \sim 1$ s^{-1}). Following the argument made earlier to compute the probability of binding versus diffusion for free Lck, the probability of Lck binding to the TCR complex is found to be:

$$P(binding) = \frac{k_{on,Lck}}{k_{off,MHC-CD4(8)} + k_{off,TCR-MHC} + k_{on,Lck}} \approx \frac{1}{20} = 0.05 \qquad (3)$$

Eqs. (2) and (3) show that Lck recruitment to the TCR complex is ten times more likely if it is associated with the co-receptor compared to when it is present as free Lck.

We conclude that co-receptors CD4 and CD8 augment T cell signaling primarily by enabling efficient recruitment of Lck to the TCR complex. The large on-rate of co-receptor-MHC interactions enables efficient recruitment of Lck by combating the effects of diffusive forces that tend to separate proteins within the range of interactions. Although, for reasons we have discussed, CD4 and CD8 have differential ability to stabilize TCR-pMHC interactions, the impact of this difference in triggering TCR signaling appear to be minor.

5.3 Appendix to Chapter 5

5.3.1. Dose response curves for peptides of different strength

Another possible test of the model would be to study the dose-response curve that follows form simulation of the model. We choose three different peptides: strong agonist

105

(koff = 0.002), typical agonist (koff=0.02) and weak agonist (koff=0.1) to see how the dose response curves change with for different peptide quality:

We see (Fig. S5.1) that although strong peptides are capable of signaling even without coreceptor, ability of weak agonist to signal is critically dependent on the presence of coreceptor. As mentioned in the main text, one might parallel this behavior to experimentally determined classes of coreceptor-dependent and coreceptor-independent peptides.

5.3.2. Cooperative binding is required for MHC stabilization on T-cell surface

Fig. S5.3 shows the enhancement of half-life of MHC bound to T-cell surface in situation when Lck can not bind TCR' intracellular domain. In this situation MHC interacts separately with TCR and coreceptor molecules present on the T-cell surface and presence of coreceptor does not improve the half-life of MHC.

5.3.3. Parameter sensitivity studies

Parameter sensitivity studies were carried out for parameters that have no experimental data available (typed in black font in Table 1):

• Fig. S5.4 illustrates that variations of kon, Lck-TCR (rate of Lck engagement with TCR) within the range of $0.5 - 5$ s^{-1} do not change the qualitative picture resulting from simulations reported in the main text.

• Fig. S5.5 illustrates that variations of koff, Lck-TCR (rate of Lck disengagement with TCR) within the range of $0.5 - 5$ s^{-1} do not change the qualitative picture resulting from simulations reported in the main text.

• Fig. S5.6 illustrates that variations of kp (rate of phosphorylation of TCR by Lck) within the range of $0.02 - 0.2$ s^{-1} do not change the qualitative picture resulting from simulations reported in the main text.

• Fig. S5.7 illustrates that variations of kdp (rate of dephosphorylation of TCR) within the range of $0.05 - 0.5$ s^{-1} do not change the qualitative picture resulting from simulations reported in the main text.

5.3.4. SSC code for simulating dissociation of MHC off the T-cell surface

```
region World
    box width 100 height 100 depth 1
subvolume edge 1

--MONOMER DISAPPEARANCE

rxn m:MHC(t#, c#) at 100000000 -> destroy m

--COMPLEXES FORMATION

-- MHC/TCR bond
rxn m:MHC(t#, c#) t:TCR(m#, c#) at kon -> m.t # t.m
rxn m:MHC(t#, c#1) CD4(m#1, t#) t:TCR(m#, c#) at kon -> m.t # t.m
rxn m:MHC(t#, c#) t:TCR(m#, c#1) CD4(t#1, m#) at kon -> m.t # t.m
rxn m:MHC(c#2, t#) t:TCR(m#, c#1) CD4(t#1, m#2) at kon -> m.t # t.m

rxn TCR(m#1) MHC(t#1, p="ag") at koffAg -> break 1
rxn TCR(m#1) MHC(t#1, p="en") at koffEn -> break 1


-- TCR/CD4 bond
rxn t:TCR(c#, m#) c:CD4(t#, m#) at konLck1 -> t.c # c.t
rxn t:TCR(c#, m#1) MHC(t#1, c#) c:CD4(t#, m#) at konLck -> t.c # c.t
rxn t:TCR(c#, m#) c:CD4(t#, m#1) MHC(c#1, t#) at konLck -> t.c # c.t
rxn t:TCR(m#2, c#) c:CD4(t#, m#1) MHC(c#1, t#2) at konLck -> t.c # c.t

rxn TCR(c#1, m#2) MHC(t#2) CD4(t#1) at koffLck -> break 1
rxn TCR(c#1, m#) CD4(t#1) at koffLck -> break 1


-- MHC/CD4 bond
rxn m:MHC(c#, t#) c:CD4(m#, t#) at konCD -> m.c # c.m
rxn m:MHC(c#, t#1) TCR(m#1, c#) c:CD4(m#, t#) at konCD -> m.c # c.m
rxn m:MHC(c#, t#) c:CD4(m#, t#1) TCR(c#1, m#) at konCD -> m.c # c.m
rxn m:MHC(t#2, c#) c:CD4(m#, t#1) TCR(c#1, m#2) at konCD -> m.c # c.m

rxn MHC(c#1) CD4(m#1) at koffCD -> break 1

--initial species

new MHC(p="ag", t#1) TCR(p="p0",m#1) at count_ag  --start with MHC bound to T-cell surface through TCR
new MHC(p="en") at count_en
new TCR(p="p0" ) at 200
new CD4(lck="basal") at 100

diffusion at 0
diffusion MHC(t#, c#) at kdiff
diffusion TCR(m#, c#) at kdiff
diffusion CD4(m#, t#) at kdiffCD4

record MHC(p="ag",t#1) TCR(m#1)
```

5.3.5. SSC code for simulation phosphorylation due to MHC-TCR interactions

```
--SPACE SETTINGS
region World
    box width 100 height 100 depth 1
    subvolume edge 1

--COMPLEXES FORMATION
-- MHC/TCR bond
rxn m:MHC(t#, c#) t:TCR(m#, c#) at kon -> m.t # t.m
rxn m:MHC(t#, c#1) CD4(m#1, t#) t:TCR(m#, c#) at kon -> m.t # t.m
rxn m:MHC(t#, c#) t:TCR(m#, c#1) CD4(t#1, m#) at kon -> m.t # t.m
rxn m:MHC(c#2, t#) t:TCR(m#, c#1) CD4(t#1, m#2) at kon -> m.t # t.m

rxn TCR(m#1) MHC(t#1, p="ag") at koffAg -> break 1

-- TCR/CD4 bond
rxn t:TCR(c#, m#) c:CD4(t#, m#) at konLck1 -> t.c # c.t
rxn t:TCR(c#, m#1) MHC(t#1, c#) c:CD4(t#, m#) at konLck-> t.c # c.t
rxn t:TCR(c#, m#) c:CD4(t#, m#1) MHC(c#1, t#) at konLck -> t.c # c.t
rxn t:TCR(m#2, c#) c:CD4(t#, m#1) MHC(c#1, t#2) at konLck -> t.c # c.t

rxn TCR(c#1, m#2) MHC(t#2) CD4(t#1) at koffLck -> break 1
rxn TCR(c#1, m#) CD4(t#1) at koffLck -> break 1

-- MHC/CD4 bond
rxn m:MHC(c#, t#) c:CD4(m#, t#) at konCD -> m.c # c.m
rxn m:MHC(c#, t#1) TCR(m#1, c#) c:CD4(m#, t#) at konCD -> m.c # c.m
rxn m:MHC(c#, t#) c:CD4(m#, t#1) TCR(c#1, m#) at konCD -> m.c # c.m
rxn m:MHC(t#2, c#) c:CD4(m#, t#1) TCR(c#1, m#2) at konCD -> m.c # c.m

rxn MHC(c#1) CD4(m#1) at koffCD -> break 1

-- MODIFICATIONS TO SPECIES
-- Phosphorylation/Dephosphorylation
rxn t:TCR(p="p0", c#1, m#2) MHC(t#2) CD4(lck="basal", t#1) at 2*kp -> t.p = "p1"
rxn t:TCR(p="p1", c#1, m#2) MHC(t#2) CD4(lck="basal", t#1) at kp -> t.p = "p2"

rxn t:TCR(p="p1", m#) at 2*kdp -> t.p = "p0"
rxn t:TCR(p="p2", m#) at kdp -> t.p = "p1"

--initial species
new MHC(p="ag") at count_ag
new TCR(p="p0" ) at 300
new CD4(lck="basal") at 100

diffusion at 0
diffusion MHC(t#, c#) at kdiff
diffusion TCR(m#, c#) at kdiff
diffusion CD4(m#, t#) at kdiff

--OUTPUT
record TCR(p="p2")
```

Tables:

Table 5.1. Rate parameters (in units s^{-1}) used in simulations. Experimentally derived parameters are shown in red (see review (Gao et al., 2002) for references), there are no experimental data available for black entries. See *appendix to chapter 5* showing that our qualitative conclusions do not change upon varying these parameters, except if Lck does not associate with TCR. Unit conversion rules can be found in chapter 4.

Rate parameter	Reaction described
150	$k_{on, TCR-MHC}$ TCR-MHC on rate, (exp~10^4 $M^{-1}s^{-1}$)
0.02	$k_{off,TCR-AgMHC}$ TCR-MHC off rate for agonist peptide (exp~0.02 s^{-1})
1000	$k_{on,CD4(8)-MHC}$ MHC-coreceptor (CD4/CD8) on-rate(exp~10^5 $M^{-1}s^{-1}$)
20	$k_{off,CD4(8)-MHC}$ MHC-coreceptor (CD4/CD8) off rate (exp~20 s^{-1})
50	k_{diff}, rate of diffusion of membrane surface proteins (exp~0.01 $\mu m^2/s$)
20	$k_{off,TCR-EnMHC}$ TCR-MHC off rate for endogenous peptide (exp~20 s^{-1})
0.05	k_p, rate of phosphorylation of TCR by Lck
0.2	k_{dp},rate of dephosphorylation of TCR
1	$k_{on, Lck-TCR}$, rate of Lck engagement with TCR
1	$k_{off, Lck-TCR}$, rate of Lck disengagement with TCR

5.4 Bibliography

Altan-Bonnet, G., and Germain, R.N. (2005). Modeling T cell antigen discrimination based on feedback control of digital ERK responses. PLoS biology *3*, e356.

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

Dushek, O., Mueller, S., Soubies, S., Depoil, D., Caramalho, I., Coombs, D., and Valitutti, S. (2008). Effects of Intracellular Calcium and Actin Cytoskeleton on TCR Mobility Measured by Fluorescence Recovery. PLoS ONE *3*, e3913.

Gao, G.F., Rao, Z., and Bell, J.I. (2002). Molecular coordination of alphabeta T-cell receptors and coreceptors CD8 and CD4 in their recognition of peptide-MHC ligands. Trends in immunology *23*, 408-413.

Gillespie, D.T. (1977). Exact Stochastic Simulation of Coupled Chemical-Reactions. Journal of Physical Chemistry *81*, 2340-2361.

Grakoui, A., Bromley, S.K., Sumen, C., Davis, M.M., Shaw, A.S., Allen, P.M., and Dustin, M.L. (1999). The immunological synapse: a molecular machine controlling T cell activation. Science (New York, N.Y *285*, 221-227.

Hamad, A.R.A., O'Herrin, S.M., Lebowitz, M.S., Srikrishnan, A., Bieler, J., Schneck, J., and Pardoll, D. (1998). Potent T Cell Activation with Dimeric Peptide-Major Histocompatibility Complex Class II Ligand: The Role of CD4 Coreceptor. pp. 1633-1640.

Holler, P.D., and Kranz, D.M. (2003). Quantitative analysis of the contribution of TCR/pepMHC affinity and CD8 to T cell activation. Immunity *18*, 255-264.

Huppa, J.B., Davis, M. M. (2009). Nature, in press.

Janeway, C., Murphy, K.P., Travers, P., Walport, M., and Janeway, C. (2008). Janeway's immuno biology (New York: Garland Science).

Li, Q.J., Dinner, A.R., Qi, S., Irvine, D.J., Huppa, J.B., Davis, M.M., and Chakraborty, A.K. (2004). CD4 enhances T cell sensitivity to antigen by coordinating Lck accumulation at the immunological synapse. Nat Immunol *5*, 791-799.

Lis, M., Artyomov, M.N., Devadas, S., and Chakraborty, A.K. (2009). Efficient stochastic simulation of reaction-diffusion processes via direct compilation. Bioinformatics (Oxford, England) *25*, 2289-2291.

Luescher, I.F., Vivier, E., Layer, A., Mahiou, J., Godeau, F., Malissen, B., and Romero, P. (1995). CD8 modulation of T-cell antigen receptor-ligand interactions on living cytotoxic T lymphocytes. Nature *373*, 353-356.

Reyes, B.M.R., Danese, S., Sans, M., Fiocchi, C., and Levine, A.D. (2005). Redox Equilibrium in Mucosal T Cells Tunes the Intestinal TCR Signaling Threshold. pp. 2158-2166.

Stone, J.D., Chervin, A.S., and Kranz, D.M. (2009). T-cell receptor binding affinities and kinetics: impact on T-cell activity and specificity. Immunology *126*, 165-176.

van Bergen, J., Kooy, Y., and Koning, F. (2001). CD4-independent T cells impair TCR triggering of CD4-dependent T cells: a putative mechanism for T cell affinity maturation. European Journal of Immunology *31*, 646-652.

Wooldridge, L., van den Berg, H.A., Glick, M., Gostick, E., Laugel, B., Hutchinson, S.L., Milicic, A., Brenchley, J.M., Douek, D.C., Price, D.A., and Sewell, A.K. (2005). Interaction between the CD8 coreceptor and major histocompatibility complex class I stabilizes T cell receptor-antigen complexes at the cell surface. Journal of Biological Chemistry *280*, 27491-27501.

Wyer, J.R., Willcox, B.E., Gao, G.F., Gerth, U.C., Davis, S.J., Bell, J.I., van der Merwe, P.A., and Jakobsen, B.K. (1999). T cell receptor and coreceptor CD8 alphaalpha bind peptide-MHC independently and with distinct kinetics. Immunity *10*, 219-225.

Wylie, D.C., Das, J., and Chakraborty, A.K. (2007). Sensitivity of T cells to antigen and antagonism emerges from differential regulation of the same molecular signaling module. Proceedings of the National Academy of Sciences of the United States of America *104*, 5533-5538.

Xiong, Y., Kern, P., Chang, H., and Reinherz, E. (2001). T Cell Receptor Binding to a pMHCII Ligand Is Kinetically Distinct from and Independent of CD4. The Journal of biological chemistry *276*, 5659-5667.

Yachi, P.P., Ampudia, J., Zal, T., and Gascoigne, N.R.J. (2006). Altered Peptide Ligands Induce Delayed CD8-T Cell Receptor Interaction--a Role for CD8 in Distinguishing Antigen Quality. Immunity *25*, 203-211.

5.5 Figures for Chapter 5

5.5.1. Main Figures
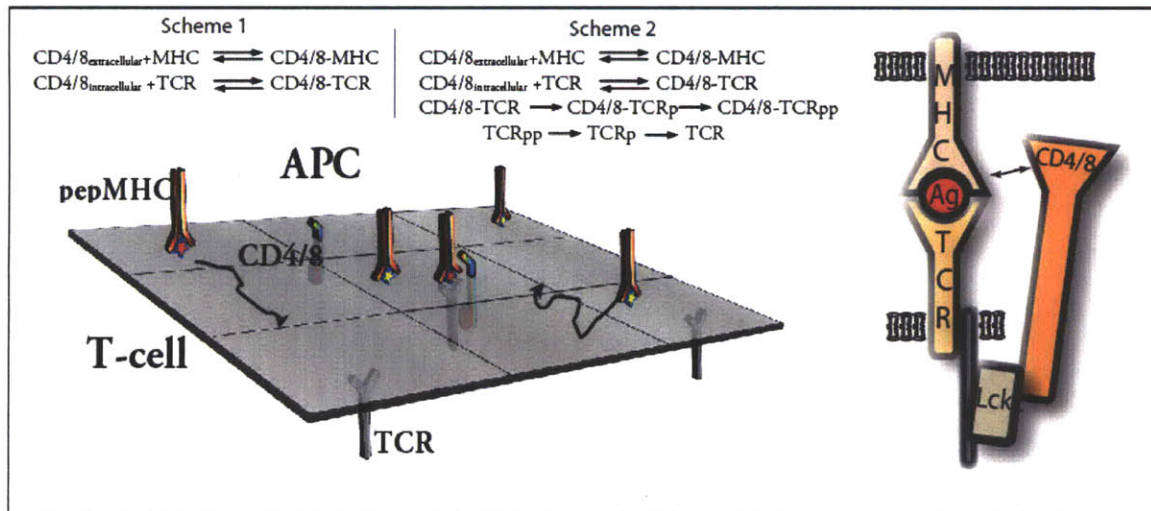
Fig. 5.1. Pictorial representation of the computer simulations that were carried out. Three kinds of proteins (MHC, coreceptor, TCR) were allowed to diffuse on the surface that represents 1 μm^2 of the T-cell/APC interface. Proteins were allowed to interact in accord with the indicated biochemical reactions. In the first set of simulations, reactions from scheme 1 were implemented; in the second set of simulations reactions from scheme 2 were implemented. Reaction rate parameters for these biochemical reactions are provided in the Table 1. Cartoon of the Lck-TCR complex that leads to cooperative interactions between the TCR, pMHC and coreceptor/Lck is shown on the right.
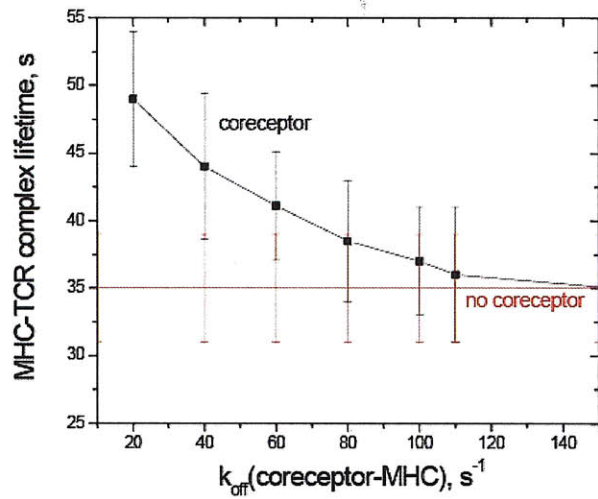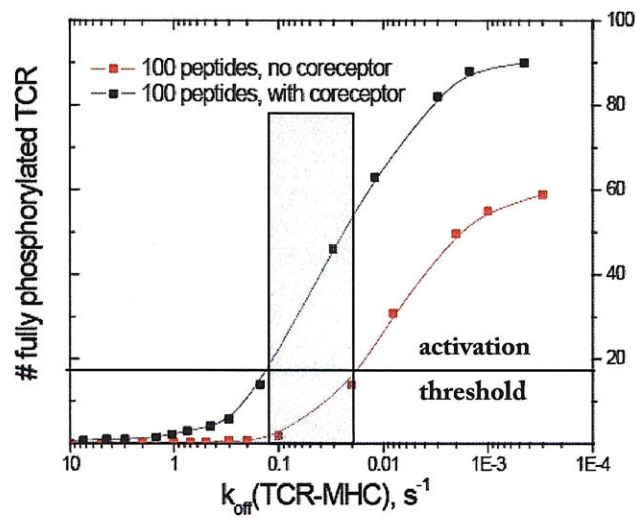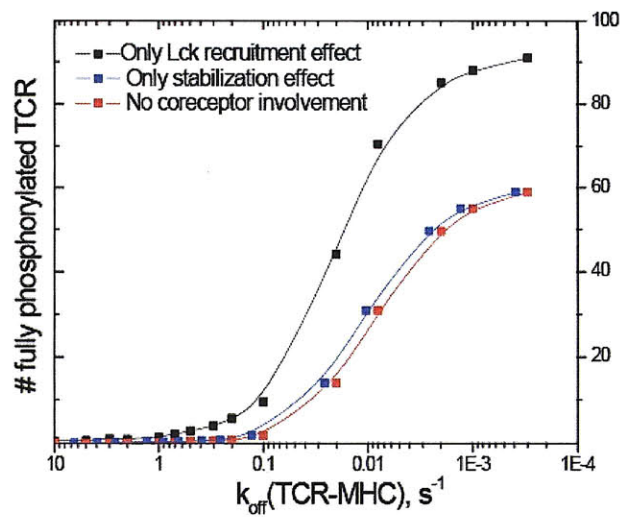
Fig. 5.2 a) Effective half-life of MHC on the T-cell surface as a function of $k_{off}^{MHC-coreceptor}$ (which is proportional to affinity of the MHC-coreceptor interaction) as obtained

in the first set of simulations (scheme 1 in Fig.1). At $k_{off} \sim 20$ s^{-1} the half-life is enhanced by ~1.5 times in the presence of coreceptor, while at $k_{off} \sim 80$ s^{-1} half-lives with and without coreceptor are statistically indistinguishable.

b) Levels of TCR phosphorylation as a function of k_{off} of the TCR-pMHC interaction. The results are obtained from simulations of 1 μm^2 of the T-cell/APC contact area with the following protein concentrations: 300 TCR per μm^2, 100 co-receptors per μm^2 (black curve), or no coreceptor present (red curve). Horizontal line indicates a threshold value of TCR phosphorylation threshold required to potentiate downstream signaling and T cell activation. Blue region represents the range of peptides that are coreceptor-dependent.

c) Signal enhancement measured by TCR phosphorylation. The red curve indicates phosphorylation level in the absence of coreceptor (for peptides of different potency, as measured by k_{off} of TCR-pMHC complex). The blue curve represents phosphorylation levels if coreceptor can only stabilize pMHC-TCR interactions, but not recruit Lck. The black curve represents phosphorylation level when coreceptors can enhance Lck recruitment but not stabilize the TCR-pMHC bond. How computer simulations could separate these effects is described in the text.

5.5.2 Supplementary Figures

(a)　　　　　　　　　　(b)　　　　　　　　　　(c)
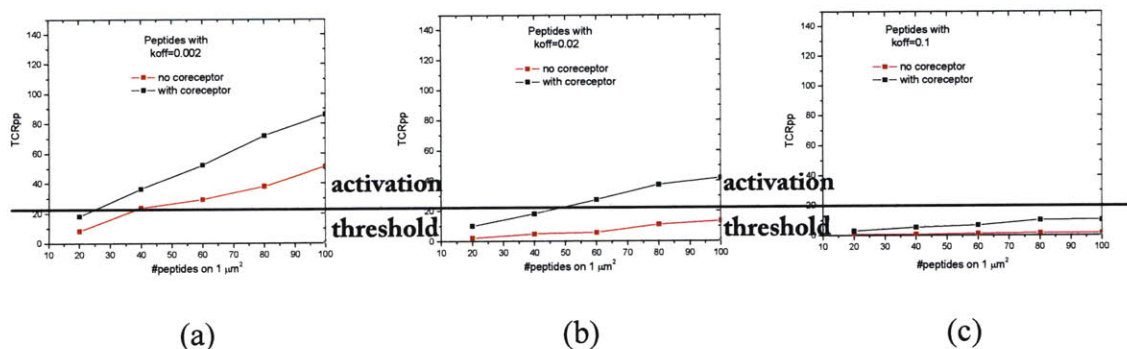
Fig. S5.1 Amount of fully phosphorylated TCR as a function of amount of peptides presented in the contact area (Dose-Response curves) for strong agonist(a), typical agonist(b), weak agonist(c)
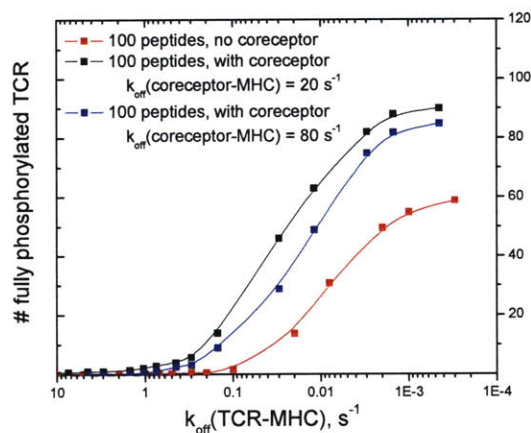
Fig. S5.2 Amount of fully phosphorylated TCR as a function of peptide quality (as determined by $k_{off}^{TCR-pMHC}$). Coreceptor mediated enhancement does not change qualitatively when varying stability of coreceptor-MHC interactions between 20 s^{-1} (black curve) and 80 s^{-1} (blue curve).
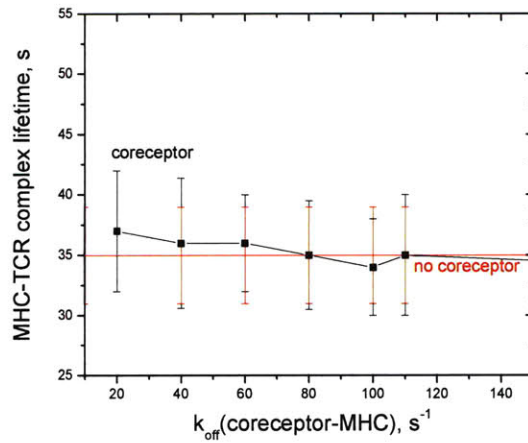
Fig. S5.3 Effective half-life of MHC on the T-cell surface as a function of $k_{off}^{MHC-coreceptor}$ (which is proportional to affinity of the MHC-coreceptor interaction) as obtained in the first set of simulations (scheme 1 in Fig.1) with association between Lck and TCR set to zero (k_{on}(lck-TCR)=0). One sees that in the absence of cooperative binding (see fig. 2a of main text) MHC is not stabilized on the surface.
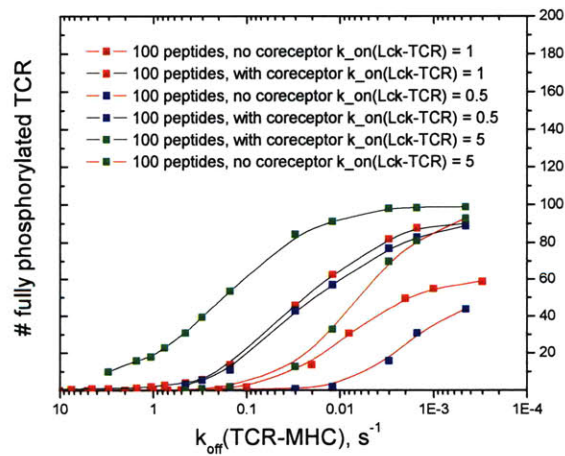
Fig. S5.4 Parameter sensitivity studies for k_{on}(Lck-TCR). Amount of fully phosphorylated TCR is plotted a function of peptide quality (as determined by $k_{off}^{TCR-pMHC}$). TCR phosphorylation with and without coreceptor is shown for 3 cases: k_{on}(Lck-TCR) = 0.5; 1; 5 s^{-1} (the rest of the parameters are the same as in Table 1). Coreceptor mediated enhancement does not change qualitatively when varying $k_{off}^{TCR-pMHC}$.
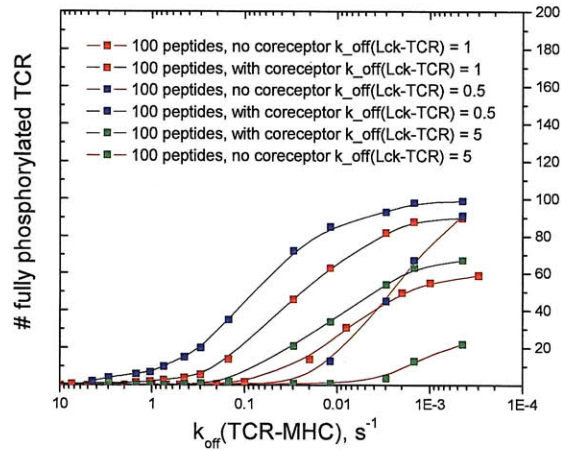
Fig. S5.5 Parameter sensitivity studies for k_{off}(Lck-TCR). Amount of fully phosphorylated TCR is plotted a function of peptide quality (as determined by $k_{off}^{TCR-pMHC}$). TCR phosphorylation with and without coreceptor is shown for 3 cases: k_{off}(Lck-TCR) = 0.5; 1; 5 s^{-1} (the rest of the parameters are the same as in Table 1). Coreceptor mediated enhancement does not change qualitatively when varying $k_{off}^{TCR-pMHC}$.
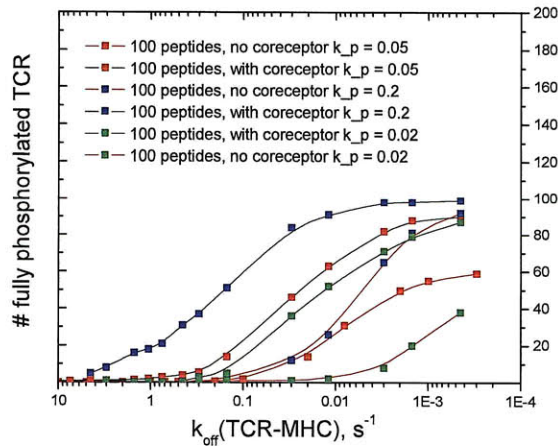
Fig. S5.6 Parameter sensitivity studies for k_p. Amount of fully phosphorylated TCR is plotted a function of peptide quality (as determined by $k_{off}^{TCR-pMHC}$). TCR phosphorylation with and without coreceptor is shown for 3 cases: k_p = 0.02; 0.05; 0.2 s^{-1} (the rest of the parameters are the same as in Table 1). Coreceptor mediated enhancement does not change qualitatively when varying k_p.
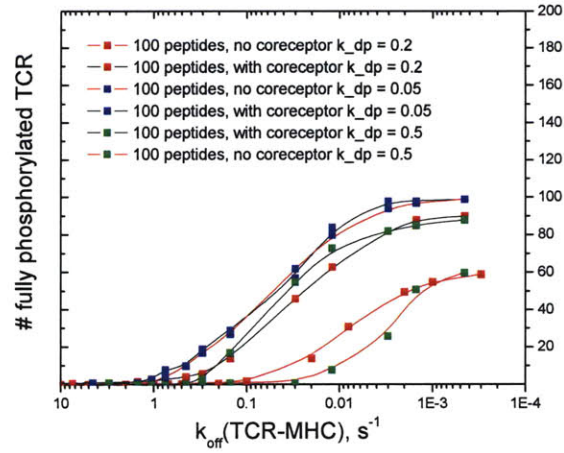
Fig. S5.7 Parameter sensitivity studies for k_{dp}. Amount of fully phosphorylated TCR is plotted a function of peptide quality (as determined by $k_{off}^{TCR\text{-}pMHC}$). TCR phosphorylation with and without coreceptor is shown for 3 cases: k_{dp} = 0.05; 0.2; 0.5 s^{-1} (the rest of the parameters are the same as in Table 1). Coreceptor mediated enhancement does not change qualitatively when varying k_{dp}.

Chapter 6

Mechanisms of signal enhancement by non-cognate peptides in CD4 and CD8 T-cells

6.1 Introduction

Interactions between peptide-MHC complex and T-cell receptor (TCR) play major role in shaping response of the T-cell to its environment allowing for pathogen recognition and proper T-cell homeostasis. Normally, during the immune response, T-cell activates only upon TCR stimulation with peptide-MHC complexes containing peptides derived from pathogenic proteins. Although peptides derived from self-proteins are not able to stimulate naïve T-cell by themselves, they actively participate in the action of the immune system. Firstly, thymocytes are tested in thymus against the pool of self-peptides to ensure appropriate sensitivity of the T-cell receptor (through positive and negative selection) (Palmer and Naeher, 2009). Secondly, it was reported that peripheral presence of weak ligands, such as self peptides, is required for proper homeostasis of the naïve T-cell population (Surh and Sprent, 2008). Thirdly, some self-derived peptides can enhance signals that were initiated by the pathogenic-peptides (Krogsgaard et al., 2007a; Krogsgaard et al., 2005). Interestingly, the manner of enhancement was found to be quite different in the CD4 and CD8 T-cells (Ebert et al., 2009; Lo et al., 2009; Yachi et al., 2005a; Yachi et al., 2007). Our paper focuses on the last phenomenon and provides computational insight into the possible differences between the CD4 and CD8 systems that could be responsible for the observed difference in the co-enhancement quality of the self-peptides in the mixtures of self- and non-self peptides.

From the biophysical point of view, difference between self- and nonself-derived peptides is manifested through the different affinities of interactions between TCR and

peptide-MHC complex (Holler and Kranz, 2003; Stone et al., 2009). Mature T-cell receptors that have undergone thymic selection are generally expected to interact very weakly with self-peptides (although exceptions do exist (Rosette et al., 2001)). Typically, self-peptides have k_{off} (TCR-pepMHC) > 10 s^{-1} while pathogenic peptides capable of activating T-cell have k_{off} (TCR-pepMHC) < 1 s^{-1} as measured by BiaCore experiments (Stone et al., 2009). Thus, one can view the value of k_{off} ~ 1 s^{-1} as a kind of threshold in k_{off} values that ensures recognition of the peptide.

It should be noted, however, that T-cell sensitivity (for example, as determined by the threshold value of k_{off}) to TCR stimulation is not set in stone and varies whilst T-cell progresses through different developmental periods. It is achieved through different "rheostat"-like mechanisms with most notable being miR181a regulations (Li et al., 2007). For instance, thymocytes are more responsive than mature T-cells and can respond to self peptides due to the change of the activating threshold as regulated by miR181a (Ebert et al., 2009).

Mature T-cells, on the other side, have a reasonably high threshold of activation such that self-derived peptides presented to T-cells on the lipid bilayer or APCs do not provoke T-cell activation. Yet, mature T-cells are able to sense the presence of self-peptides in the mixture with cognate peptides as indicated by signal augmentation compared to the situation when only cognate peptides are presented to the T-cell (Ebert et al., 2009; Lo et al., 2009; Yachi et al., 2005a; Yachi et al., 2007). This indicates that mechanism of enhancement is different from sensitivity modulation. The exact molecular mechanism remains the subject of the debates particularly because of the contrasting results obtained with CD4 and CD8 T-cells (Ebert et al., 2009; Lo et al., 2009; Yachi et al., 2005a; Yachi et al., 2007).

In CD8 T-cells it has been found that all tested non-activating-peptides are able to synergize the effect of cognate peptides, while in CD4 cells only some of the null-peptides are capable of enhancing the signal provided by agonist (Li et al., 2007; Yachi et al., 2007). In this paper we show that these results are, in fact, consistent with each other and can arise from a single mechanism.

Involvement of non-cognate peptides along with agonist peptides presents an interesting puzzle because of the dramatic difference in the time scales associated with

stability of different pepMHC-TCR complex. For the activating peptides lifetime of the bound TCR-pepMHC complex is about 10-100 seconds ($k_{off} \sim 0.1$-0.01 s^{-1}), while self-derived peptides that are always present on the APC surface in complex with MHC form extremely unstable bond with TCR with lifetime about 50 milliseconds ($k_{off} > 20$ s^{-1}). Three orders of magnitude in difference between k_{off}'s suggest that different mechanisms may operate in each situation.

First step to reconciling these rate differences was due to Wylie *et al* who hypothesized that Lck might have two states capable of carrying out the phosphorylation, basally active and fully active states (Wylie et al., 2007a). While the former has limited kinase activity, the latter is active enough to phosphorylate even short lived complexes such as self-peptide-MHC-TCR. This assumption was recently confirmed in experiments studying the properties of membrane bound phosphotase CD45 that actively maintains the balance between three forms of Lck (inactive, basally active, fully active) (McNeill et al., 2007; Zamoyska, 2007). Particularly, CD45 promotes transformation of the inactive form of Lck to the basally active, but also inactivates fully active form down to the basally active one. In other words, it is CD45 who assures that Lck is in the basally active state.

This function of CD45 protein, together with its peculiar physical dimensions, plays the crucial role in the spatial coordination of the kinase activity of Lck. Namely, the length of the extracellular domain of CD45 is approximately 40 A, which is considerably larger compared to approximately 15 A of the total length of the TCR-pepMHC connection. The dimensional mismatch makes it impossible for phosphotase CD45 to be localized in close proximity to the bound TCR-pepMHC complex (Burroughs and van der Merwe, 2007). So, Lck can be securely promoted to the fully active state while in contact with TCR-pepMHC complex, but it will be deactivated shortly upon departure from the safe zone around TCR-pepMHC. This, as one will see in more details below, creates a potential for the signal enhancement: if fully activated Lck encounters TCR engaged with the self-pepMHC before it gets deactivated, it might as well phosphorylate its ζ-chains, thus, increasing the total TCR phosphorylation level.

As one can see from the above considerations, the mechanism behind the self-peptide co-enhancement includes profound contribution of the spatial component.

Although there have been conceptual studies dealing with the principles behind the self-peptide enhancement (Li et al., 2004a; Wylie et al., 2007a), coherent consideration that takes into account both protein-protein interactions and spatial localization and incorporates *all* available biophysical measurements is still lacking. Here, we report the first study that meets these criteria and, thus, allows us to analyze different possible enhancement mechanisms in details.

We consider the situation when antigen presenting cell (APC) presents the mixture of cognate and non-cognate peptides on its surface. We analyze three different explicit models that describe signal enhancement due to the presence of non-signaling peptides on the APC surface. Each model represents distinct mechanism that is characterized by a certain parameter regime of validity and specific assumptions about the molecular interactions. We find in all cases that interplay between spatial motion of membrane proteins and protein-protein interactions are critical to the signal enhancement mechanism.

This explicit description of synergism due to non-cognate peptide-MHCs allows us to address the important debate regarding the differences in the signal enhancement in CD4 and CD8 T-cells (Ebert et al., 2009; Lo et al., 2009; Yachi et al., 2005a; Yachi et al., 2007). We search for the possibility that differences in CD4 vs CD8 experiments arise from a common mechanism due to differences in the strength of molecular interactions. For example, it is well known that CD8 has larger affinity for MHC molecules than CD4 (Gao et al., 2002). We, therefore, examine the models with respect to parameters that describe structural and molecular differences between CD4 and CD8 coreceptors, such as interactions with MHC receptor, interactions between TCR and Lck bound to intracellular domain of the coreceptor and efficacy of ζ-chain phosphorylation by Lck. In all three models we, indeed, are able to identify the critical properties that are responsible for the observed experimental distinction between CD4 and CD8 self-peptide co-stimulation.

Finally, the prospective taken in this work, namely, a heavy emphasis on spatial movement of proteins in the membrane in early T-cell signaling, allows us to comment on the importance and function of the heterogeneteities existing in the T-cell membrane before any TCR signaling. Important example is microclustering of T-cell receptors on

the surface (Campi et al., 2005; DeMond et al., 2008; Dustin, 2009). Our work offers the way of thinking for more detailed studies of spatial effects in T-cell signaling that will address the questions like what is optimal spatial distribution of TCR leading to the optimal signaling (i.e. what is optimal size and pattern of microclusters) *etc*.

6.2 Model 1 of self-peptide enhancement

Model 1 is described by the following general picture. At rest, T-cell surface contains a mixture of T-cell receptors and coreceptors (Lck is assumed to be constantly associasted with coreceptor, at least on the time scale of the early signaling events). Large amounts of CD45 ensure that Lck, both free and coreceptor associated one, is in the basally active state, thus, at the initial point we assume that all of the Lck is at the basally active state. The activity of Lck is defined through its ability to phosphorylate engaged T-cell receptor, reflected by the ζ-chain phosphorylation rate (k_p and $k_{p,act}$ – see Table 1) (Kersh et al., 1998). We define the level of basal activity of Lck (k_p) as the one that is consistent with kinetic proofreading requirements sufficient to deliver TCR phosphorylation by an agonist ($k_{off, TCR-AgMHC} < 0.1s^{-1}$), but not by the self-peptide. Upon engagement with the antigen-presenting cell, coreceptor-associated Lck is recruited to the engaged TCR-pepMHC complex. This process is regulated by two interactions: coreceptor-MHC binding (on-rate $k_{on,CD4(8)-MHC}$ and off-rate $k_{off,CD4(8)-MHC}$) and Lck association with intracellular domain of TCR (on-rate $k_{on,Lck-TCR}$ and off-rate $k_{off,Lck-TCR}$). In this model we assume that Lck can carry out ζ-chains phosphorylation only if it is associated with TCR that is engaged with MHC at the same moment. Note that peptide identity does not play direct role, the likelihood of TCR phosphorylation is very different for self and non-self-peptides due to different lifetimes of the pepMHC-TCR complex.

Only upon complete ζ-chain phosphorylation, adapter proteins associated with T-cell receptor can upgrade Lck to the fully active state provided that Lck remains associated with TCR sufficiently long. This process is coarse-grained into one activation reaction described by the rate k_{act}. Fully active state of Lck is characterized by increased ζ-chain phosphorylation rate ($k_{p,act}$) and can carry out phosphorylation even before engagement with TCR, rather, spatial proximity is sufficient condition for kinase action. Deactivation under the action of CD45 is also described through single reaction

characterized by the rate k_{deact}; it is assumed that deactivation can only take place when coreceptor is not associated with either TCR or MHC and, thus, approach of oversized CD45 protein is possible.

Ultimately, coreceptor bearing fully active Lck can disassemble from MHC-TCR complex after which it diffuses freely around the T-cell surface. During this period, fully active form of Lck survives for a limited amount of time until CD45 downgrades Lck to basaly active form (this timescale is set by the choice of k_{deact} to be approximately $1/k_{deact}$). If, during this time, it encounters TCR that is transiently associated with the self-derived pepMHC, Lck can phosphorylate new TCR provided that fully active Lck is sufficiently active. The sufficient condition is that $k_{p,act} \geq k_{off,TCR-EnMHC}$, i.e. endogenous peptide-MHC comlpex is engaged with TCR long enough for the fully active Lck to phosphorylate its ζ-chains. This mode of action is local with respect to the initial agonist-MHC engagement position because fully active Lck can only go so far before it becomes deactivated to the basal level. This mechanism is summarized pictorially on the Fig. 6.1.

We see therefore, that self peptides involvement in the signaling process is critically dependent on the spatial component and mutual locations of cognate pepMHC, non-cognate MHC and coreceptor. We can deduce the most sensitive parameters controlling enhancement by self peptides to be diffusion rate of the coreceptor, rate of deactivation of the fully active form of Lck, and, of course, concentration of the self-peptides. It comes as no surprise in this model that artificially co-localized self- and agonist peptides are capable of the strong enhancement of the signal, *e.g.* when put together on the quantum dot or bound by a molecular tether (Anikeeva et al., 2006; Krogsgaard et al., 2007b).

In order to study the behavior of the model in details we have performed spatially resolved stochastic simulations mimicking 1 μm^2 interface between T-cell and antigen-presenting cell. We assume that there are 300 T-cell receptors, 100 coreceptors engaged with Lck and 100 MHC-peptide complexes freely diffusing in this area. The read-out of the simulation was the amount of fully phosphorylated TCRs at steady-state. For simplicity, multitude of phosphorylation site on ζ-chain of TCR was modeled by introducing two states – partially phosphorylated TCR and fully phosphorylated TCR (Kersh et al., 1998; Wylie et al., 2007a). The presence of the phosphotases that

constitutively dephosphorylate ζ-chains was modeled by the dephosphorylation reaction characterized by the rate k_{dp}.

Table 1 presents parameters used in the simulations, and lists the basic set of reactions. The red entries indicate parameters available from biophysical measurements. These table entries were recomputed from available experimental information (all the numbers are in units of s^{-1} in accord with the discussion in methods section, experimentally measured values corresponding to the simulation parameter are also provided). Black entries had to be estimated theoretically. In fact, in many cases experimental data provide indirect, yet very strict bounds on the experimentally unavailable parameters.

For instance, the phosphorylation and dephosphorylation by basal Lck rates are readily estimated from the fact that kinetic proofreading scheme leading to fully phosphorylated TCR must separate peptide-MHC complexes into two classes in accord with their $k_{off,TCR-pepMHC}$. Particularly, peptide-MHC complexes with $k_{off} > 1$ s^{-1} should never be able to stimulate the T-cell (i.e., fully phosphorylate TCR in this model) while peptide-MHCs with $k_{off} < 1$ s^{-1} deliver different levels of TCR phosphorylation (see Fig. 6.1). Rate of Lck association with internal part of TCR and Lck activation rate are estimated following the same reasoning. There is, however, no reliable argument to anticipate the value of two parameters.

Firstly, phosphorylation rate by fully activated Lck is free parameter of the model that has been varied extensively. As one will see below, it is varying this rate parameter we find that the model exhibits regimes reminiscent of CD4 and CD8 behavior with respect to self-peptides for different values of phosphorylation rate by fully activated Lck.

Secondly, Lck deactivation rate from fully active to the basal state is completely free parameter. As mentioned above, this parameter regulates the lifetime of the freely diffusing fully active Lck. Thus, two extreme scenarios are possible: one, when Lck has lifetime sufficient to encounter other TCR-MHC complexes, and second, when Lck is deactivated immediately upon coreceptor disengagement from TCR-MHC. This, in fact, leads to two very different mechanism of signal enhancement that are considered in models 1 and 2. Here, in first model, we set deactivation rate to a sufficiently small

value. Physically this means that coreceptor with activated Lck will have a reasonable lifetime after disengagement from the original TCR-pepMHC. During this lifetime, activated Lck will phosphorylate TCRs transiently complexed with self-peptides-MHC.

Next, we see how the levels of phosphorylation delivered by agonist alone compare with the phosphorylation triggered by the mixture of agonist and null-peptide (k_{off} =20 s^{-1}). According to experimental data, mixtures perform better due to synergistic effects of the non-cognate peptides. Indeed, as one can see on the Fig. 6.3 addition of non-cognate peptides dramatically increases the signal. Moreover, for small concentrations of agonist, we see that 10 agonist peptides presented in the sea of the self-peptides perform as good as ~30 agonist peptides alone.

What is the parameter, responsible for the amount of enhancement delivered by self-peptides? Rate of deactivation of coreceptor associated Lck, which, basically, represents the activity of the CD45 and diffusion rate of the free coreceptor both determine how far away can receptor with fully active Lck diffuse from the agonistMHC-TCR complex which initially activated Lck. The bigger is the area covered by diffusion of fully active Lck, the bigger is the harvest of TCRs bound to endogenous pepMHC that Lck can phosphorylate. As one can see from Fig. 6.3 (red open circles), increasing deactivation rate leads to the smaller signal enhancement levels.

Molecular and structural differences between the CD4 and CD8 coreceptors can be ascribed to several parameters in this model. Firstly, these are parameters corresponding to the interaction of coreceptor with MHC. It is generally accepted that CD8 affinity to the MHC class I molecules is 2-3 times higher than CD4 affinity to MHC class II molecules (Gao et al., 2002). Secondly, the rate of Lck binding to TCR can be different due to different structure of intracellular and transmembrane domains of CD4 and CD8 proteins. Finally, the rate of phosphorylation by the fully active Lck could also be different for CD4 and CD8 proteins as a consequence of different spatial organization of coreceptor-Lck-TCR complex.

Experiments show that enhancement effect depends on the self-peptides identity, which in kinetic reaction-diffusion scheme translates into stability of self-pepMHC-TCR (k_{off}), for tested pool of peptides in case of CD4 T-cells, while all tested null-peptides

enhance the signal equally well in CD8 T-cells. Can we describe these experimental differences varying parameters distinguishing CD4 and CD8 T-cells?

In our simulations behavior of CD8 cells would be manifested in that all the self-peptides in the wide range of $k_{off,TCR-EnMHC}$ (e.g. 10 s^{-1} to 300 s^{-1}) would be able to enhance the signal initiated by the agonist. On the other side, experimentally observed CD4-like behavior would correspond to the situation when only limited range of self-peptides could co-enhance agonist-MHC (e.g. only 10 s^{-1} to 50 s^{-1}).

We find in our simulations that varying parameters related to coreceptor association with MHC or Lck binding to TCR does not change the range of self-peptides that are able to enhance the agonist-derived signal. Rather these parameters affect the magnitude of enhancement for all the self-peptides equally. However, when the effectiveness of TCR phosphorylation by fully activated Lck is perturbed, one sees that range of co-enhancing self-peptides changes. If the phosphorylation rate by fully active Lck is high, then practically all the self peptides can synergize the signaling with equal efficiency (see Fig. 6.4a), which correspond to the observations made on CD8 T-cells. If, however, phosphorylation by fully active Lck is moderately stronger then by basal Lck, the quality (k_{off}) of the self-peptide plays a crucial role (see Fig. 6.4b), like it was observed in case of CD4 T-cells. Biologically this would mean that CD8 coreceptor associates Lck with TCR more tightly and carries out phosphorylation more effectively then CD4.

Also, note that in CD4-like regime (Fig. 6.4b) only strong agonist can be synergized by sufficiently good peptides, whereas in CD8-like case (Fig. 6.4a) this effect is independent of quality of the agonist.

Summing up, in the described mechanism, the difference in behavior of the CD4 and CD8 T-cell is due to the distinct ways of coordinating Lck nearby intracellular part of the TCR, which leads to the higher efficiency of phosphorylation in CD8 T-cells compared to CD4 cells. Indeed, the recent experiments (Mallaun et al., 2008) indicate that the potency of coreceptor associated kinase is critically dependent on the close proximity between coreceptor and T-cell receptor. This difference in CD4 versus CD8 Lck proximity to intracellular domain of TCR can be a consequence of the difference in affinities of the coreceptors to MHC or simply follow from the different structures of

these proteins. We identify it as a critical, parameters, but the cause of variation lies actually in the physical differences for CD4 and CD8 coreceptors ((association with MHC, structure of the intracellular domain).

6.3 Model 2 of self-peptide enhancement

In this model we explore a different set of assumptions that describe the self-peptide involvement in the signaling process. Here, we consider the regime when deactivating ability of CD45 is very high (k_{deact} is large) and lifetime of fully active Lck is very small, i.e. it becomes deactivated almost immediately upon disengagement from TCR-pepMHC complex. Thus, the only the possibility to extend action of fully active Lck is through stable association with intracellular domains of TCR for as long as TCR is engaged to peptide-MHC complex (i.e., $k_{off,Lck-TCR}$ is very small). Because of the stable binding, activated Lck is always localized within the original, successful TCR-pepMHC interaction. Thus, signal enhancement can only occur when other MHCs that bear self-peptides occasionally engage with TCR in the region nearby the initial TCR-MHC complex (as shown pictorially on Fig. 6.5).

If, in order to reflect the above biological picture, we merely change $k_{off, Lck-TCR}$ in the model 1 to the smaller value, it leads to the constitutive signaling even with non-signaling peptides, thus, losing the distinction between agonist and non-agonist peptides. Such a behavior is due to stochastic activating of Lck: once every while, even self-peptides are capable of activating one or few fully active Lck molecules. Such rare events play no role in model 1 because Lck is deactivated shortly after it disengages from TCR. If Lck binds to TCR very stably then single fully active Lck has a lifetime determined not by deactivating ability of CD45 but rather by dissociation rate of Lck-TCR connection. Presence of active Lck allows for phosphorylation of bystander-TCR-pepMHC which, in turn, activates more Lck leading to global TCR phosphorylation. In this manner rare events trigger uncontrolled activation of T-cells. The only difference between weak and strong peptides in this situation is in average waiting time until such rare event occurs. Biologically this would imply that every T-cell could get activated by non-cognate peptide if we observe it for sufficiently long time, which is obviously non-realistic.

Hence, model 2 requires that, unlike in previous model, we use two different parameters describing the stability of Lck-TCR complex. As one can see from table 1, rate of Lck dissociation from TCR, $k_{off, Lck-TCR}$, takes two different values: very large, when TCR is not bound to pepMHC (i.e. Lck-TCR bond is very unstable without pepMHC), or small value, when TCR is bound to pepMHC (i.e., Lck-TCR bond is stable). As detailed above, additional discrimination between the rates of TCR binding with free Lck or MHC-coreceptor mediated Lck binding is necessary in this model, because otherwise free Lck could trigger unstimulated T-cell due to single stochastic engagement leading to long stable binding.

As one can see from Table 1, reaction network in this model is almost identical to the previous model with the only difference in rate constants and possibility to cross-phosphorylate bystander TCR by Lck engaged with the other TCR. In order to study this mechanism and do not mix it with model 1, deactivation rate of Lck was set to the very large value. This means that Lck will be deactivated almost instantaneously upon disengaging from TCR, hence, only proximal mode of cross-phosphorylation is possible. Two mechanisms described by models 1 and 2 are, therefore, not mutually exclusive, and, in reality, signal enhancement is likely to be the consequence of both of the mechanisms.

Since no changes are made in the basic signal initiation module, model 2 is equally capable of discriminating between agonist and non-agonist peptides (see Fig. 6.6) in the same manner as model 1. One finds, however, that enhancement mechanism of model 2 is less efficient compared to the model 1 (see below). It is because the mechanism of signal enhancement in Model 2 requires that 5 proteins ultimately appear in the very close proximity to each other (see Fig. 6.5) as opposed to just 3 proteins "colliding" productively in the Model 1.

In terms of models simulations the difference in enhancement strength is seen when varying TCR dephosphorylation rates in each model, which represent activity of generic phosphotases. To recognize the effect of dephosphorylation rate one should note that in both models TCR phosphorylation is arising from two sources: direct phosphorylation due to engaging with agonist-MHC, which provides the signal, and

phosphorylation due to non-agonist pepMHCs, which provides signal enhancement. Changing the dephosphorylation rate affects differentially each of the mechanisms.

Direct phosphorylation mechanisms and rate parameters are identical in both models. Since in both models dephosphorylation is possible only after disengagement of TCR from MHC (i.e. when phosphotase can approach it), then, even in the limit of the very strong dephosphorylation rate, TCRs bound to the agonist MHCs will get phosphorylated. In this limit, upon disengagement TCRs will get dephosphorylated almost instantaneously and thus phosphorylation levels (i.e. signal levels) are defined by the number of TCR bound to agonist-MHC at each instant of time. Note, that signal enhancement decreases to zero when phosphotases are very active because those TCRs are regularly exposed to phosphotases due to shortness of binding time with non-cognate peptide-MHC. Hence, we see that rate of dephosphorylation affects phosphorylation levels due to signal initiation less dramatically then signal enhancement. As one decreases $k_{dph,\ TCR}$ (i.e. reduces level of phosphotases activity), signal enhancement becomes more and more pronounced. By comparing the level of signal enhancement for equal values of $k_{dph,TCR}$ in both models we can judge the efficiency of each enhancement mechanism.

In simulations of model 2, $k_{dph,TCR}$ must be decreased compared to model 1 in order to observe similar levels of signal enhancement (see table 1 and Fig. 6.7). Indeed, when keeping dephosphorylation rate in Model 2 exactly the same as in Model 1, one observes no signal enhancement when comparing agonist-null-peptide mixture against the agonist alone. Even though this comparison allows us to judge the relative "strength" of signal enhancement in each mechanism, it does not provide the basis for discriminating between two models, because one can not say which value of the dephosphorylation rate corresponds to the level of phosphotases activity inside the real cell since $k_{dph,TCR}$ actually coarse-grains a number of processes into a single step

In spite of the different enhancement mechanism, model 2 exhibits features similar to model 1 with respect to the discrimination between CD4-like and CD8-like behavior. Similar to model 1, in model 2 there are several relevant parameters: rates of coreceptor-MHC association and dissociation, rate of Lck associating with, and

dissociating from, the internal domain of TCR, and, finally, the rate of phosphorylation by fully active Lck.

Varying relevant parameters we should be able to reproduce both CD4- and CD8- like behavior of T-cell activation module with respect to self-peptides. Namely, for different values of the parameter we should be able to see that the self-peptides of broad range contribute approximately equally to the signal enhancement, independent of their quality (CD8-like), or that self-peptides of restricted range are capable of co-enhancing (CD4-like behavior).

Decreasing coreceptor-MHC association rate or increasing their dissociation rates decreases levels of TCR phosphorylation due to signal enhancement, but does so equally for all self-peptides. Same results are obtained when varying parameters of TCR and Lck interactions. Therefore, it is impossible to reproduces both CD4 and CD8 behavior only changing these parameters. The discrimination of the self-peptides based on their quality can, however, be explained when varying the rate of TCR-phosphorylation by fully active Lck with CD8-like regime observed for more efficient phosphorylation by fully active Lck, and CD4-like regime for less efficient phosphorylation (Fig. 6.8). Biologically, this can be the consequence of the bigger stability of CD8-MHC interactions and differenes in structure between CD4 and CD8 coreceptors.

6.4 Model 3 of self-peptide enhancement

In this part we study spatial behavior of the model that can be considered as a modification of the "pseudodimer" model (Krogsgaard et al., 2007a). Note that in models 1 and 2 coreceptor could bind to the engaged pepMHC-TCR complex at the same time through both its extracellular domain (CD4/CD8+MHC interactions) and its intracellular domain (Lck associated with coreceptor + TCR interaction). This arrangement plays important role in the described mechanism of Lck recruitment to the engaged TCR. In this, a la "pseudodimer", model, it is assumed that due to steric reasons (Krogsgaard et al., 2007a) such an arrangement is not possible. Instead, receptor bound to engaged TCR through intracellular domain can attach to another MHC molecule forming "pseudodimer" structure with two different MHCs in the close proximity of a single coreceptor and its Lck kinase (Fig. 6.9).

As one can see from Table 1, the basic parameters of the model are the same as for model 2 with the exception of the dephosphorylation rate constant and rate of Lck association with TCR. Due to this similarity, we see that discrimination of the agonist and null peptides is preserved in this model (Fig. 6.10). However, since model 3 does not provide efficient Lck recruitment through coreceptor, rate of direct association between Lck and intracellular domain of engaged TCR has to be increased compared to the models 1 and 2 in order to achieve reasonable levels of signaling. At the same time, we find that enhancement strength, as understood in previous section, is about the same for model 3 as for model 1. This is reflected in that similar enhancement levels (Fig. 6.11) are achieved with same rate of dephosphorylation (Table 1). Improved enhancement in model 3 compared to model 2 is due to strong interactions (see $k_{on,coreceptor-MHC}$, Table 1) of the extracellular domain of coreceptor with bypassing MHCs, effectively recruiting them to the engagement site while in model 2 it was passive process of random occurrence of MHC nearby the site of initial engagement.

We do not find, however, that "pseudodimer" behaves like catalytic center phosphorylating multiple TCRs transiently engaging with non-cognate pepMHC part of pseudodimer (Krogsgaard et al., 2007a). Rather, because of the short-lived coreceptor MHC association, we find that multiple engagements of MHC-TCR lead to the signal enhancement

Finally, we study how in this model the differences between CD4 and CD8 receptors can be accommodated. Following the same procedure as for the previous two models, we find that in model 3 it is also effectiveness of phosphorylation by fully active Lck that is responsible for observed experimental disparity between CD4 and CD8 T-cells. The Fig. 6.12 indicates that for smaller value of $k_{p,act}$ one can observe the difference between the different non-cognate peptides enhancing the signal, while for larger value of $k_{p,act}$ all natural non-cognate peptides will deliver approximately similar level of signal enhancement.

6.5 Discussion

We have described three models detailing possible mechanisms of signal enhancement by non-cognate peptides. Each model is characterized by different set of

assumptions and employs parameters consistent with available direct biophysical measurements. Model 1 requires the smallest number of assumptions and parameters about the molecular interactions. Models 2 and 3, however, require additional assumption about ability of fully active Lck to cross phosphorylate bystander TCR while still bound to initial TCR. Also supplementary distinction of Lck binding with engaged/free TCR should be introduced to models 2 and 3. Note that models 1 and 2 are compatible with described mechanism (Chapter 5) of the Lck recruitment, while model 3 does not have a mechanism for coreceptor recruiting Lck to the engaged TCR. It must be also additionally stressed that models 1 and 2 are not mutually exclusive and actual enhancement process can employ both mechanisms at the same time.

Importance of the spatial coordination of T-cell receptors and MHC proteins is a feature common to all the models. The locality of the action of fully activated Lck restricts the region of active signaling to the proximity of the initial TCR-MHC productive encounter. This correlates well with recent experiments indicating that additional co-localization of TCRs (e.g. through the actin skeleton actions) improves signaling in T-cells (Anikeeva et al., 2006; DeMond et al., 2008; Krogsgaard et al., 2005). On the same basis, the experiments where peptide-MHC proteins are artificially co-localized with a biotin link (Krogsgaard et al., 2005) are also explained by either of three models.

Analysis of all models reveals that one of the important differences between CD4 and CD8 T-cells is the efficiency of TCR phosphorylation by fully activated TCR. This efficiency determines the ultimate level of sensitivity accessible by T-cells, because no pepMHC-TCR interaction can be "felt" by T-cell if lifetime of pepMHC-TCR complex is smaller than time required by fully active Lck to phosphorylate engaged TCR. All the considered models predict that fully active Lck is more efficient in CD8 T-cells (i.e. $k_{p,actCD8} > k_{p,actCD4}$). This fact has to be of special importance in the situation when T-cell operates on the edge of its sensitivity threshold, e.g. when it has to be positively selected during the thymic selection process. In accord with the considered models, it would be reasonable to predict that CD8 T-cells could be positively selected on the broader set of peptides than CD4 T-cells (as measured by $k_{off,\ pepMHC-TCR}$). This prediction is corroborated by recent experiments by Mark M. Davis group (Ebert et al., 2009) and

Allen's group (Lo et al., 2009) where endogeneous positively selecting peptides were identified for the first time and it was found that only these peptides are capable of co-enhancing activation signal derived from agonist.

We conclude by considering the possibilities of experimental discrimination between three models. Model 3 does not allow coreceptor to be simultaneously bound to the engaged TCR-MHC complex with both its intracellular domain bound to TCR and extracellular domain bound to MHC. FRET or similar structural studies could be of help when judging the reliability of this assumption. Models 1 and 2 are different in more subtle manner and are not mutually exclusive. Model 1 requires coreceptor to diffuse away from the site of original engagement, while model 2 provides very local mode of signal enhancement. Note, that Model 2 would be very insensitive to the variations in the phosphotase maintaining Lck in its basal state (CD45), while large increase of the CD45 levels will shut down any signal enhancement in Model 1. So, varying CD45 levels and measuring the change in signal enhancement, one can estimate experimentally the contribution of each of the model. Alternative experiment would be to physically co-localize coreceptor with TCR with some kind of covalent linker and determine the level of signal enhancement. With physical co-localization, only model 2 that can be realized and, hence, measured change in enhancement levels will describe the contribution of enhancement in accord to the model 1.

Finally, it follows from all of the considered mechanisms that colocalization of T-cell receptors on the surface will drastically improve signaling and sensitivity. This is particularly important because according to recent experiments (Campi et al., 2005; Dustin, 2009) T-cell receptors are not randomly distributed across the T-cell surface but rather are organized in the pre-formed microcluters. The formation of microclusters can be very important part of the spatial organization critical for the early T-cell signaling which we plan to assess in future work.

6.6 Appendix to Chapter 6

SSC code Model 1:

```
--COMPLEXES FORMATION
-- MHC/TCR bond
rxn m:MHC(t#, c#) t:TCR(m#, c#) at kon -> m.t # t.m
rxn m:MHC(t#, c#1) CD4(m#1, t#) t:TCR(m#, c#) at kon -> m.t # t.m
rxn m:MHC(t#, c#) t:TCR(m#, c#1) CD4(t#1, m#) at kon -> m.t # t.m
rxn m:MHC(c#2, t#) t:TCR(m#, c#1) CD4(t#1, m#2) at kon -> m.t # t.m

rxn TCR(m#1) MHC(t#1, p="ag") at koffAg -> break 1
rxn TCR(m#1) MHC(t#1, p="en") at koffEn -> break 1

-- TCR/CD4 bond
rxn t:TCR(c#, m#) c:CD4(t#, m#) at konLck1 -> t.c # c.t
rxn t:TCR(c#, m#1) MHC(t#1, c#) c:CD4(t#, m#) at konLck1 -> t.c # c.t
rxn t:TCR(c#, m#) c:CD4(t#, m#1) MHC(c#1, t#) at konLck1 -> t.c # c.t
rxn t:TCR(m#2, c#) c:CD4(t#, m#1) MHC(c#1, t#2) at konLck2 -> t.c # c.t

rxn TCR(c#1, m#2) MHC(t#2) CD4(t#1) at koffLck -> break 1
rxn TCR(c#1, m#) CD4(t#1) at koffLck2 -> break 1

-- MHC/CD4 bond
rxn m:MHC(c#, t#) c:CD4(m#, t#) at konCD -> m.c # c.m
rxn m:MHC(c#, t#1) TCR(m#1, c#) c:CD4(m#, t#) at konCD -> m.c # c.m
rxn m:MHC(c#, t#) c:CD4(m#, t#1) TCR(c#1, m#) at konCD -> m.c # c.m
rxn m:MHC(t#2, c#) c:CD4(m#, t#1) TCR(c#1, m#2) at konCD -> m.c # c.m

rxn MHC(c#1) CD4(m#1) at koffCD -> break 1

-- MODIFICATIONS
-- Phosphorylation/Dephosphorylation
rxn t:TCR(p="p0", c#1, m#2) MHC(t#2) CD4(lck="basal", t#1) at kp1 -> t.p = "p1"
rxn t:TCR(p="p1", c#1, m#2) MHC(t#2) CD4(lck="basal", t#1) at kp2 -> t.p = "p2"

rxn t:TCR(p="p0", c#1, m#2) MHC(t#2) CD4(lck="active", t#1) at kpact1 -> t.p = "p1"
rxn t:TCR(p="p1", c#1, m#2) MHC(t#2) CD4(lck="active", t#1) at kpact2 -> t.p = "p2"
rxn t:TCR(p="p0", m#2) MHC(t#2, c#1) CD4(lck="active", m#1) at kpact1 -> t.p = "p1"
rxn t:TCR(p="p1", m#2) MHC(t#2, c#1) CD4(lck="active", m#1) at kpact2 -> t.p = "p2"

rxn t:TCR(p="p1", m#) at kdp1 -> t.p = "p0"
rxn t:TCR(p="p2", m#) at kdp2 -> t.p = "p1"

-- Lck activation/deactivation
rxn c:CD4(lck="basal", t#1) TCR(c#1, p="p2", m#2) MHC(t#2) at kact -> c.lck = "active"
rxn c:CD4(lck="basal", m#1) TCR(p="p2", m#2) MHC(c#1, t#2) at kact -> c.lck = "active"
rxn c:CD4(lck="active", t#, m#) at kdeact -> c.lck = "basal"

--initial species
new MHC(p="ag") at count_ag
new MHC(p="en") at count_en
new TCR(p="p0" ) at 300
new CD4(lck="basal") at 100

diffusion at 0
diffusion MHC(t#, c#) at kdiff
diffusion TCR(m#, c#) at kdiff
diffusion CD4(m#, t#) at kdiff

record TCR(p="p2")
```

Rate parameters

'kon': 150,   'koffAg': 0.02,   'konCD': 1000,   'koffCD': 20.0,
'kp1': 0.1,   'kp2': 0.05,   'kdp1': 0.4,   'kdp2': 0.2,   'kdiff': 50,   'kact': 1,
'kdeact': 0.3,   'koffEn': 20,   'konLck1': 1,   'konLck2': 1,   'koffLck': 1,
'koffLck2':1,   'kpact1': 10000,   'kpact2': 10000,   'count_ag': 0,   'count_en': 0,

SSC code Model 2

```
--COMPLEXES FORMATION
-- MHC/TCR bond
rxn m:MHC(t#, c#) t:TCR(m#, c#) at kon -> m.t # t.m
rxn m:MHC(t#, c#1) CD4(m#1, t#) t:TCR(m#, c#) at kon -> m.t # t.m
rxn m:MHC(t#, c#) t:TCR(m#, c#1) CD4(t#1, m#) at kon -> m.t # t.m
rxn m:MHC(c#2, t#) t:TCR(m#, c#1) CD4(t#1, m#2) at kon -> m.t # t.m


rxn TCR(m#1) MHC(t#1, p="ag") at koffAg -> break 1
rxn TCR(m#1) MHC(t#1, p="en") at koffEn -> break 1
-- TCR/CD4 bond
rxn t:TCR(c#, m#) c:CD4(t#, m#) at konLck1 -> t.c # c.t
rxn t:TCR(c#, m#1) MHC(t#1, c#) c:CD4(t#, m#) at konLck1 -> t.c # c.t
rxn t:TCR(c#, m#) c:CD4(t#, m#1) MHC(c#1, t#) at konLck1 -> t.c # c.t
rxn t:TCR(m#2, c#) c:CD4(t#, m#1) MHC(c#1, t#2) at konLck2 -> t.c # c.t


rxn TCR(c#1, m#2) MHC(t#2) CD4(t#1) at koffLck -> break 1
rxn TCR(c#1, m#) CD4(t#1) at koffLck2 -> break 1
-- MHC/CD4 bond
rxn m:MHC(c#, t#) c:CD4(m#, t#) at konCD -> m.c # c.m
rxn m:MHC(c#, t#1) TCR(m#1, c#) c:CD4(m#, t#) at konCD -> m.c # c.m
rxn m:MHC(c#, t#) c:CD4(m#, t#1) TCR(c#1, m#) at konCD -> m.c # c.m
rxn m:MHC(t#2, c#) c:CD4(m#, t#1) TCR(c#1, m#2) at konCD -> m.c # c.m


rxn MHC(c#1) CD4(m#1) at koffCD -> break 1
-- MODIFICATIONS
-- Phosphorylation/Dephosphorylation
rxn t:TCR(p="p0", c#1, m#2) MHC(t#2) CD4(lck="basal", t#1) at kp1 -> t.p = "p1"
rxn t:TCR(p="p1", c#1, m#2) MHC(t#2) CD4(lck="basal", t#1) at kp2 -> t.p = "p2"


rxn t:TCR(p="p0", c#1, m#2) MHC(t#2) CD4(lck="active", t#1) at kpact1 -> t.p = "p1"
rxn t:TCR(p="p1", c#1, m#2) MHC(t#2) CD4(lck="active", t#1) at kpact2 -> t.p = "p2"
rxn t:TCR(p="p0", m#2) MHC(t#2, c#1) CD4(lck="active", m#1) at kpact1 -> t.p = "p1"
rxn t:TCR(p="p1", m#2) MHC(t#2, c#1) CD4(lck="active", m#1) at kpact2 -> t.p = "p2"


rxn t:TCR(p="p0", m#1, c#) MHC(t#1) CD4(lck="active") at kpact1cross -> t.p = "p1"
rxn t:TCR(p="p1", m#1, c#) MHC(t#1) CD4(lck="active") at kpact2cross -> t.p = "p2"


rxn t:TCR(p="p1", m#) at kdp1 -> t.p = "p0"
rxn t:TCR(p="p2", m#) at kdp2 -> t.p = "p1"
-- Lck activation/deactivation
rxn c:CD4(lck="basal", t#1) TCR(c#1, p="p2", m#2) MHC(t#2) at kact -> c.lck = "active"
rxn c:CD4(lck="basal", m#1) TCR(p="p2", m#2) MHC(c#1, t#2) at kact -> c.lck = "active"
rxn c:CD4(lck="active", t#, m#) at kdeact -> c.lck = "basal"


--initial species
new MHC(p="ag") at count_ag
new MHC(p="en") at count_en
new TCR(p="p0" ) at 300
new CD4(lck="basal") at 100
diffusion at 0
diffusion MHC(t#, c#) at kdiff
diffusion TCR(m#, c#) at kdiff
diffusion CD4(m#, t#) at kdiff

layout grid 100 * 100

record TCR(p="p2")
record MHC(p="ag", c#, t#)
```

SSC code Model 3:

```
limit TCR at 5
--COMPLEXES FORMATION
-- MHC/TCR bond
rxn m:MHC(t#) t:TCR(m#) at kon -> m.t # t.m

rxn TCR(m#1) MHC(t#1, p="ag") at koffAg -> break 1
rxn TCR(m#1) MHC(t#1, p="en") at koffEn -> break 1

-- TCR/CD4 bond
rxn t:TCR(c#) c:CD4(t#) at konLck -> t.c # c.t
rxn TCR(c#1) CD4(t#1) at koffLck -> break 1
rxn TCR(c#1,m#) CD4(t#1,m#) at koffLck2 -> break 1

-- MHC/CD4 bond
rxn m:MHC(c#) c:CD4(m#) at konCD -> m.c # c.m
rxn MHC(c#1) CD4(m#1) at koffCD -> break 1

-- MODIFICATIONS
-- Phosphorylation/Dephosphorylation
rxn t:TCR(p="p0", c#1, m#2) MHC(t#2) CD4(lck="basal", t#1) at kp1 -> t.p = "p1"
rxn t:TCR(p="p1", c#1, m#2) MHC(t#2) CD4(lck="basal", t#1) at kp2 -> t.p = "p2"

rxn t:TCR(p="p0", c#1, m#2) MHC(t#2) CD4(lck="active", t#1) at kpact1 -> t.p = "p1"
rxn t:TCR(p="p1", c#1, m#2) MHC(t#2) CD4(lck="active", t#1) at kpact2 -> t.p = "p2"
rxn t:TCR(p="p0", m#2) MHC(t#2, c#1) CD4(lck="active", m#1) at kpact1 -> t.p = "p1"
rxn t:TCR(p="p1", m#2) MHC(t#2, c#1) CD4(lck="active", m#1) at kpact2 -> t.p = "p2"

rxn t:TCR(p="p1", m#) at kdp1 -> t.p = "p0"
rxn t:TCR(p="p2", m#) at kdp2 -> t.p = "p1"

-- Lck activation/deactivation
rxn c:CD4(lck="basal", t#1) TCR(c#1, p="p2", m#2) MHC(t#2) at kact -> c.lck = "active"
rxn c:CD4(lck="basal", m#1) TCR(p="p2", m#2) MHC(c#1, t#2) at kact -> c.lck = "active"
rxn c:CD4(lck="active", t#, m#) at kdeact -> c.lck = "basal"

--initial species
new MHC(p="ag") at count_ag
new MHC(p="en") at count_en
new TCR(p="p0") at 300
new CD4(lck="basal") at 100
diffusion at 0
diffusion MHC(t#, c#) at kdiff
diffusion TCR(m#, c#) at kdiff
diffusion CD4(m#, t#) at kdiff

layout grid 100 * 100

record TCR(p="p2")
```

Table 6.1 Rate parameters used in simulations of Models 1-3, studying signal enhancement by non-cognate peptides. (Experimentally derived parameters are shown in red, fitted parameters are shown in black – there are no experimental data available for black entries)

Model #			Reaction described
1	2	3	
150	150	150	$k_{on, TCR-MHC.}$ TCR-MHC on rate (exp~10^4 $M^{-1}s^{-1}$)
0.02	0.02	0.02	$k_{off,TCR-AgMHC}$ TCR-MHC off rate for agonist peptide (exp~0.02 s^{-1})
1000	1000	1000	$k_{on,CD4(8)-MHC}$ MHC-coreceptor (CD4/CD8) on-rate (exp~10^5 $M^{-1}s^{-1}$)
20	20	20	$k_{off,CD4(8)-MHC}$ MHC-coreceptor (CD4/CD8) off rate (exp~20 s^{-1})
50	50	50	$k_{diff,}$ rate of diffusion of membrane surface proteins (~0.01 $\mu m^2/s$)
20	20	20	$k_{off,TCR-EnMHC}$ TCR-MHC off rate for endogenous peptide (exp~20 s^{-1})
0.05	0.05	0.05	k_p, rate of phosphorylation of TCR by basal Lck
0.2	0.02	0.2	$k_{dp,}$ rate of dephosphorylation of TCR
1	1	100	$k_{on, Lck-TCR,}$ rate of Lck engagement with TCR
1	0.0002	0.02	$k_{off, Lck-TCR,}$ rate of Lck disengagement with TCR
	20	20	
300	300	300	$k_{p,active,,}$ rate of phosphorylation by fully active Lck
1	1	1	$k_{act,}$ rate of Lck activation
0.3	10^5	10^5	$k_{deact,}$ rate of Lck deactivation when away from TCR-pepMHC complex
N/A	300	300	$k_{p,cross}$ rate of phosphorylation of the bystander TCR by the fully active Lck bound to another TCR

6.7 Bibliography

Anikeeva, N., Lebedeva, T., Clapp, A.R., Goldman, E.R., Dustin, M.L., Mattoussi, H., and Sykulev, Y. (2006). Quantum dot/peptide-MHC biosensors reveal strong CD8-dependent cooperation between self and viral antigens that augment the T cell response. Proceedings of the National Academy of Sciences of the United States of America *103*, 16846-16851.

Burroughs, N.J., and van der Merwe, P.A. (2007). Stochasticity and spatial heterogeneity in T-cell activation. Immunological reviews *216*, 69-80.

Campi, G., Varma, R., and Dustin, M.L. (2005). Actin and agonist MHC-peptide complex-dependent T cell receptor microclusters as scaffolds for signaling. pp. 1031-1036.

DeMond, A.L., Mossman, K.D., Starr, T., Dustin, M.L., and Groves, J.T. (2008). T Cell Receptor Microcluster Transport through Molecular Mazes Reveals Mechanism of Translocation. Biophysical Journal *94*, 3286-3292.

Dustin, M.L. (2009). The Cellular Context of T Cell Signaling. Immunity *30*, 482-492.

Ebert, P.J., Jiang, S., Xie, J., Li, Q.J., and Davis, M.M. (2009). An endogenous positively selecting peptide enhances mature T cell responses and becomes an autoantigen in the absence of microRNA miR-181a. Nature immunology *10*, 1162-1169.

Gao, G.F., Rao, Z., and Bell, J.I. (2002). Molecular coordination of alphabeta T-cell receptors and coreceptors CD8 and CD4 in their recognition of peptide-MHC ligands. Trends in immunology *23*, 408-413.

Holler, P.D., and Kranz, D.M. (2003). Quantitative analysis of the contribution of TCR/pepMHC affinity and CD8 to T cell activation. Immunity *18*, 255-264.

Kersh, E.N., Shaw, A.S., and Allen, P.M. (1998). Fidelity of T cell activation through multistep T cell receptor zeta phosphorylation. Science (New York, N.Y *281*, 572-575.

Krogsgaard, M., Juang, J., and Davis, M.M. (2007a). A role for "self" in T-cell activation. Seminars in immunology *19*, 236-244.

Krogsgaard, M., Juang, J., and Davis, M.M. (2007b). A role for "self" in T-cell activation. Seminars in Immunology *19*, 236-244.

Krogsgaard, M., Li, Q.J., Sumen, C., Huppa, J.B., Huse, M., and Davis, M.M. (2005). Agonist/endogenous peptide-MHC heterodimers drive T cell activation and sensitivity. Nature *434*, 238-243.

Li, Q.J., Chau, J., Ebert, P.J., Sylvester, G., Min, H., Liu, G., Braich, R., Manoharan, M., Soutschek, J., Skare, P., *et al.* (2007). miR-181a is an intrinsic modulator of T cell sensitivity and selection. Cell *129*, 147-161.

Li, Q.J., Dinner, A.R., Qi, S., Irvine, D.J., Huppa, J.B., Davis, M.M., and Chakraborty, A.K. (2004). CD4 enhances T cell sensitivity to antigen by coordinating Lck accumulation at the immunological synapse. Nature immunology *5*, 791-799.

Lo, W.L., Felix, N.J., Walters, J.J., Rohrs, H., Gross, M.L., and Allen, P.M. (2009). An endogenous peptide positively selects and augments the activation and survival of peripheral CD4+ T cells. Nature immunology *10*, 1155-1161.

Mallaun, M., Naeher, D., Daniels, M.A., Yachi, P.P., Hausmann, B., Luescher, I.F., Gascoigne, N.R.J., and Palmer, E. (2008). The T cell receptor's alpha-chain connecting peptide motif promotes close approximation of the CD8 coreceptor allowing efficient signal initiation. J Immunol *180*, 8211-8221.

McNeill, L., Salmond, R.J., Cooper, J.C., Carret, C.K., Cassady-Cain, R.L., Roche-Molina, M., Tandon, P., Holmes, N., and Alexander, D.R. (2007). The differential regulation of Lck kinase phosphorylation sites by CD45 is critical for T cell receptor signaling responses. Immunity *27*, 425-437.

Palmer, E., and Naeher, D. (2009). Affinity threshold for thymic selection through a T-cell receptor-co-receptor zipper. Nat Rev Immunol *9*, 207-213.

Rosette, C., Werlen, G., Daniels, M.A., Holman, P.O., Alam, S.M., Travers, P.J., Gascoigne, N.R., Palmer, E., and Jameson, S.C. (2001). The impact of duration versus extent of TCR occupancy on T cell activation: a revision of the kinetic proofreading model. Immunity *15*, 59-70.

Stone, J.D., Chervin, A.S., and Kranz, D.M. (2009). T-cell receptor binding affinities and kinetics: impact on T-cell activity and specificity. Immunology *126*, 165-176.

Surh, C.D., and Sprent, J. (2008). Homeostasis of Naive and Memory T Cells. Immunity *29*, 848-862.

Wylie, D.C., Das, J., and Chakraborty, A.K. (2007). Sensitivity of T cells to antigen and antagonism emerges from differential regulation of the same molecular signaling module. Proceedings of the National Academy of Sciences of the United States of America *104*, 5533-5538.

Yachi, P.P., Ampudia, J., Gascoigne, N.R., and Zal, T. (2005). Nonstimulatory peptides contribute to antigen-induced CD8-T cell receptor interaction at the immunological synapse. Nature immunology *6*, 785-792.

Yachi, P.P., Lotz, C., Ampudia, J., and Gascoigne, N.R. (2007). T cell activation enhancement by endogenous pMHC acts for both weak and strong agonists but varies with differentiation state. The Journal of experimental medicine *204*, 2747-2757.

Zamoyska, R. (2007). Why is there so much CD45 on T cells? Immunity *27*, 421-423.

6.8 Figures for Chapter 6
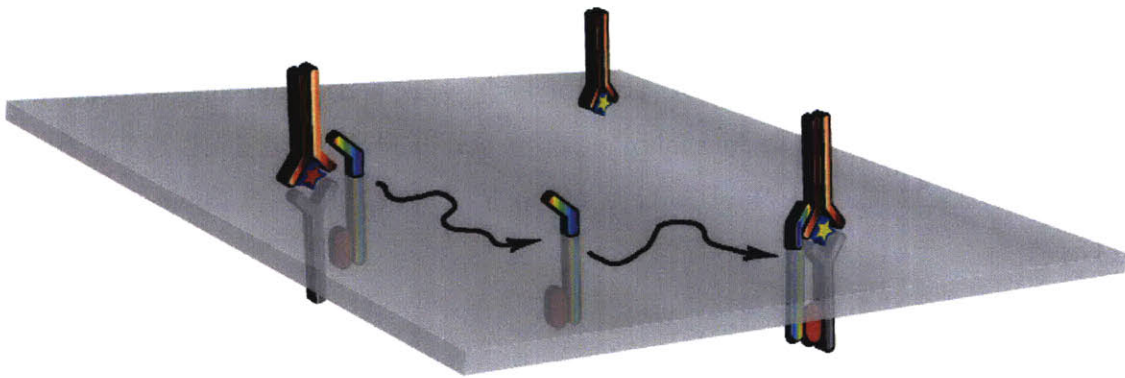
Fig. 6.1 Pictorial description of Model 1. The grey surface represents interface between T-cell (underneath the surface) and APC (above the surface). MHC proteins (brownish) at APC present two types of peptides: agonist peptides (red star) and non-cognate peptides (yellow star). Coreceptor (rainbow color) that spans through the interface is constitutively associated with Lck (red oval) associates with MHC and TCR and after full activation of Lck disengages, diffuses around and occasionally encounters TCR engaged with non-cognate pepMHC and phosphorylate it.
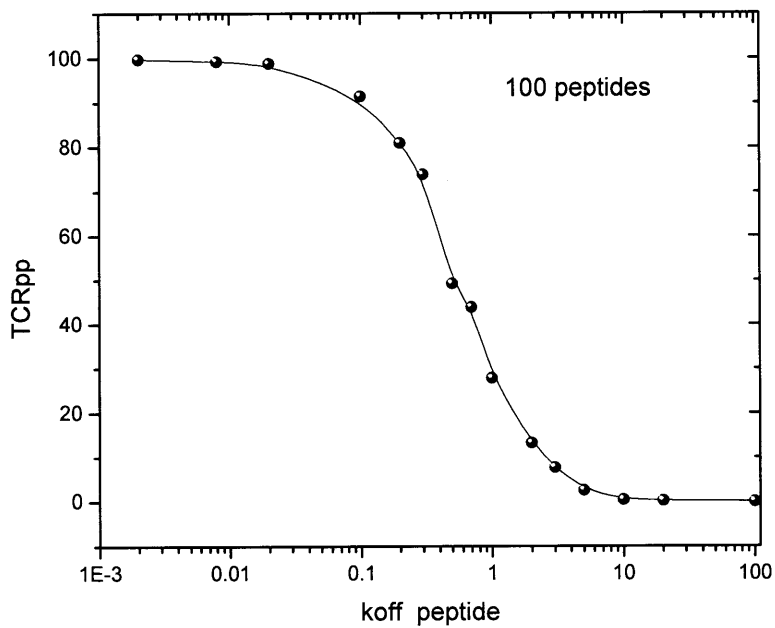
Fig. 6.2 Peptide potency as a function of k_{off} of the peptidesMHC-TCR for the model where Lck can be basally active or fully active. Simulation of 1 μm^2 of the T-cell/APC contact area. Concentration of TCR is 300 per μm^2, concentration of coreceptor is 100 per μm^2.

Fig. 6.3 Signaling levels of the agonist peptide alone (closed squares) compared to the mixture of agonist and non-cognate peptide (open circles). The curve in red describes the behaviour of the mixture of agonist and non-agonist peptides with the rate of Lck deactivation twice larger than that for the black curve (Table 1). Increasing deactivation reduces the lifetime of fully activated Lck and hence reduces the signal enhancement
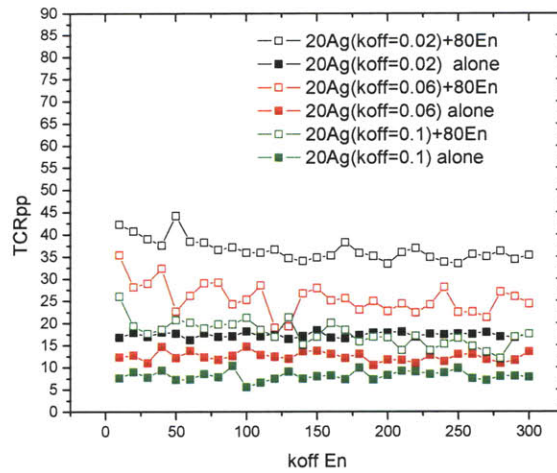
(a)
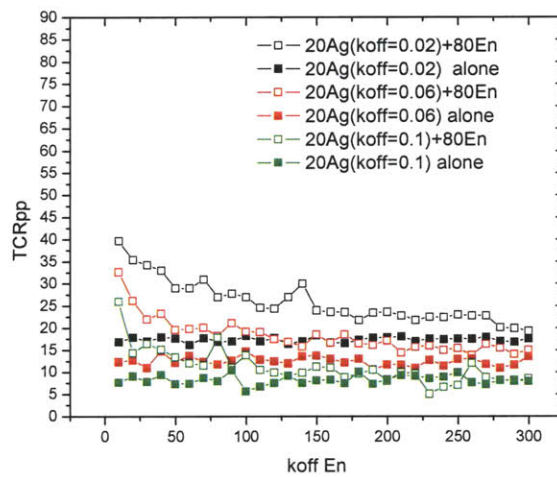
(b)

Fig. 6.4 The TCR phosphorylation stimulated by agonist of indicated k_{off} (three types of Ag are presented: strong with k_{off} = 0.02 in black, intermediate with k_{off}=0.06 in red and weak with k_{off}=0.1 in green) or mixture of agonist together with non-stimulating peptide (denoted as En, endogenous). x-Axis shows the k_{off} for EnMHC-TCR interaction. Panel (a) corresponds to the high value of the rate of phosphorylation by fully active Lck ($k_{ph,act}$ = 300), while panel (b) has low rate of phosphorylation by fully active Lck ($k_{ph,act}$=100). As on can see from panel (a) non-cognate peptides of all kinds synergize equally well when $k_{ph,act}$ is large, which is reminiscent of CD8 T-cells behavior. On panel (b) enhancement level differs for different non-cognate peptides, which is similar to observed CD4 behavior.
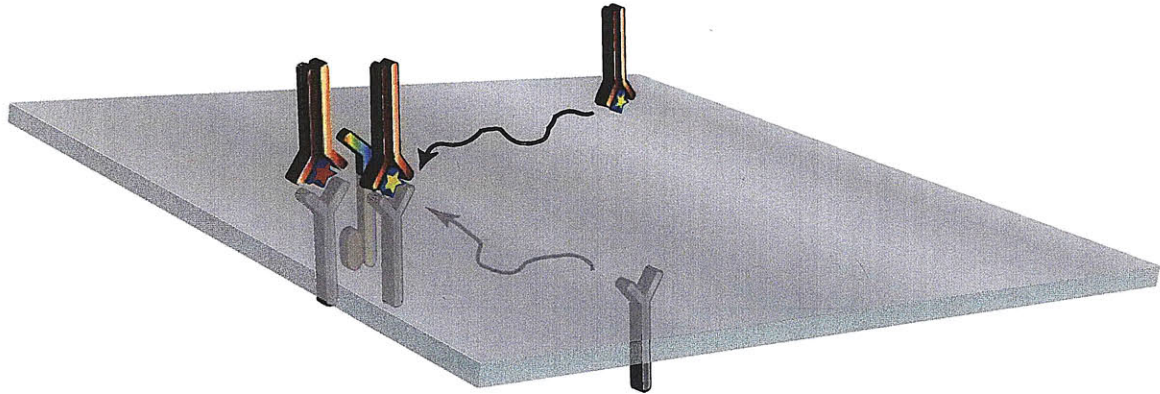
Fig. 6.5 Pictorial description of Model 2. Coreceptor binds stably to the assembled TCR-agonistMHC complex and associated Lck gets fully activated. Fully activated Lck is capable of phosphorylating bypassing TCRs that engage with any pepMHC in the close vicinity of the original complex. Upon disassembly from original complex fully activated Lck is deactivated almost instantaneously, which ensures the proximal mode of signal enhancement.
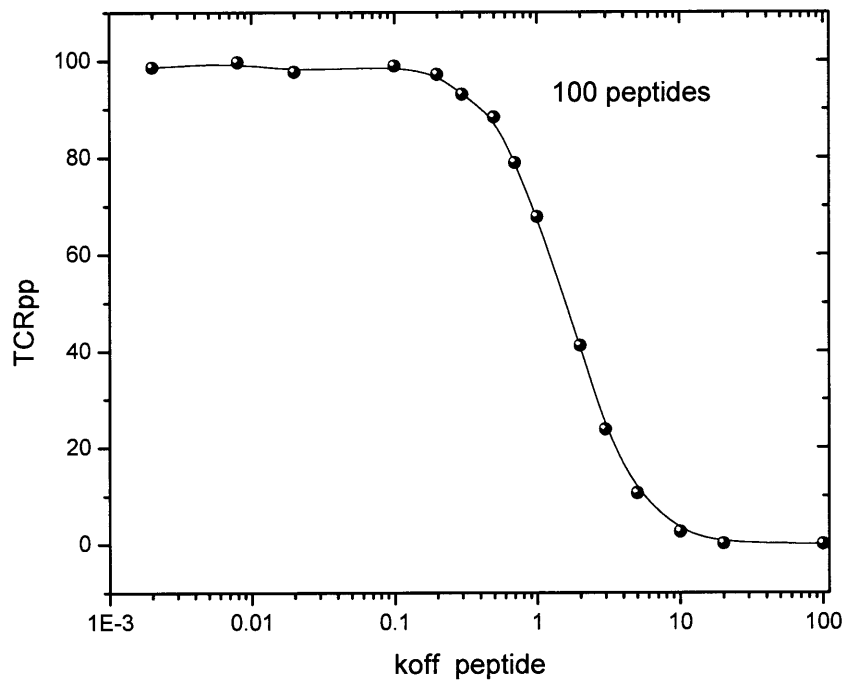
Fig. 6.6 Peptide potency as a function of k_{off} of the peptidesMHC-TCR for the model 2. Simulation of 1 μm^2 of the T-cell/APC contact area. Concentration of TCR is 300 per μm^2, concentration of coreceptor is 100 per μm^2.
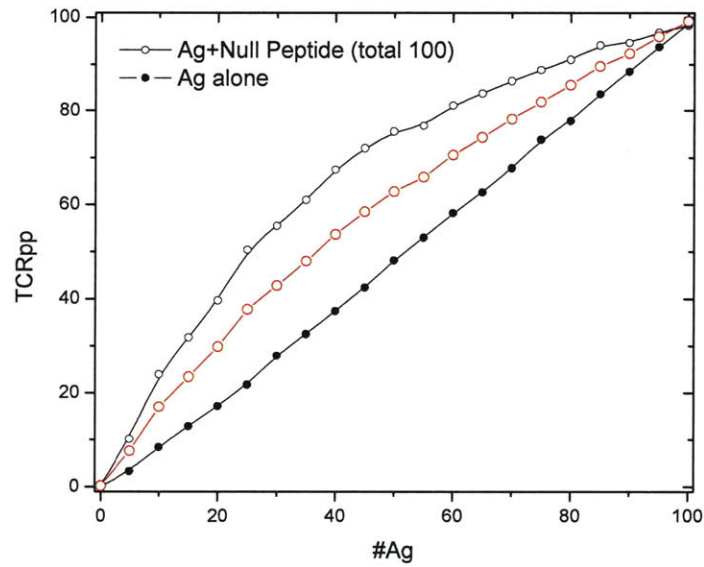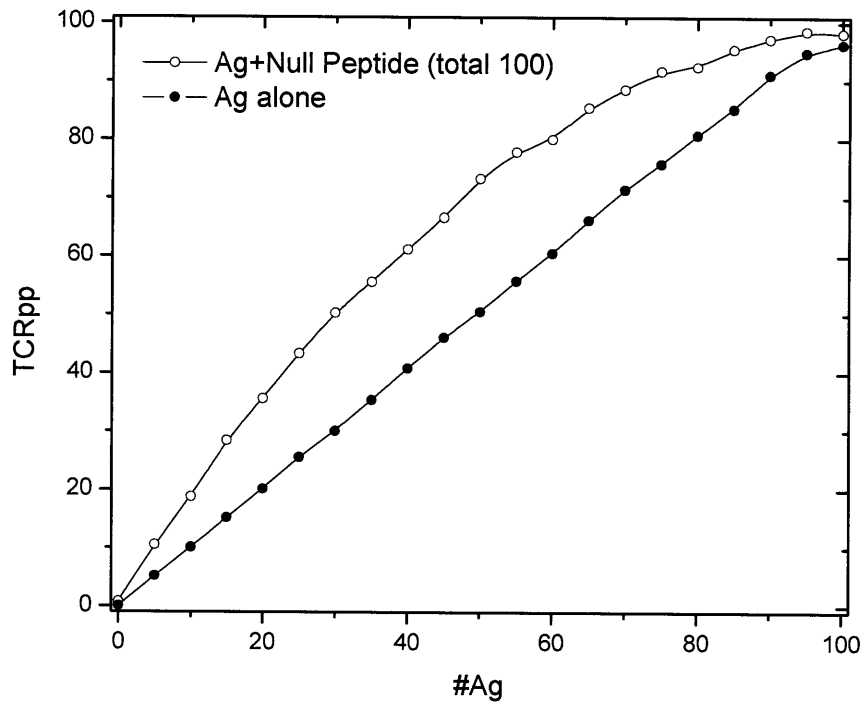
Fig. 6.7 Signaling levels of the agonist peptide alone (closed squares) compared to the mixture of agonist and non-cognate peptide (open circles). A mixture delivers biger phosphorylation due to non-cognate peptides enhancement.
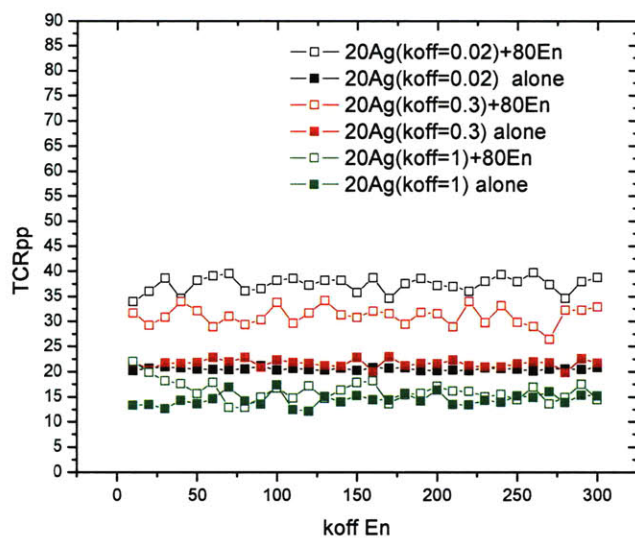
(a)
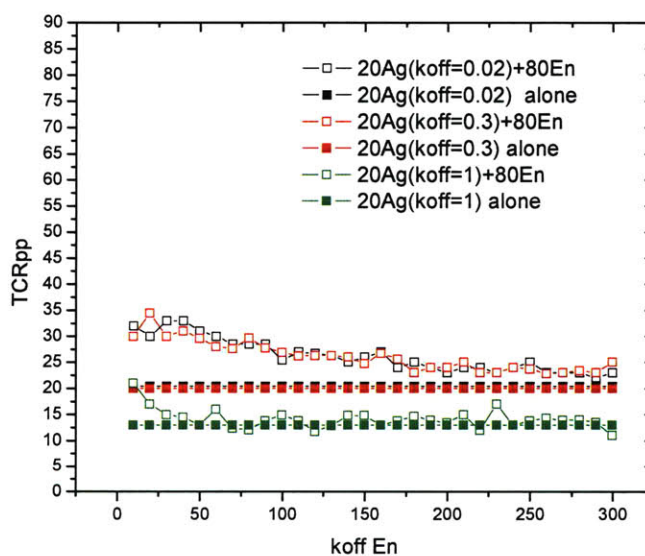
(b)

Fig. 6.8 The TCR phosphorylation stimulated by agonist of indicated k_{off} or mixture of agonist together with non-stimulating peptide (denoted as En, endogenous). x-Axis shows the k_{off} for EnMHC-TCR interaction. Panel (a) corresponds to the high value of the rate of phosphorylation by fully active Lck ($k_{ph,act} = 300$), while panel (b) has low rate of phosphorylation by fully active Lck ($k_{ph,act} = 100$).
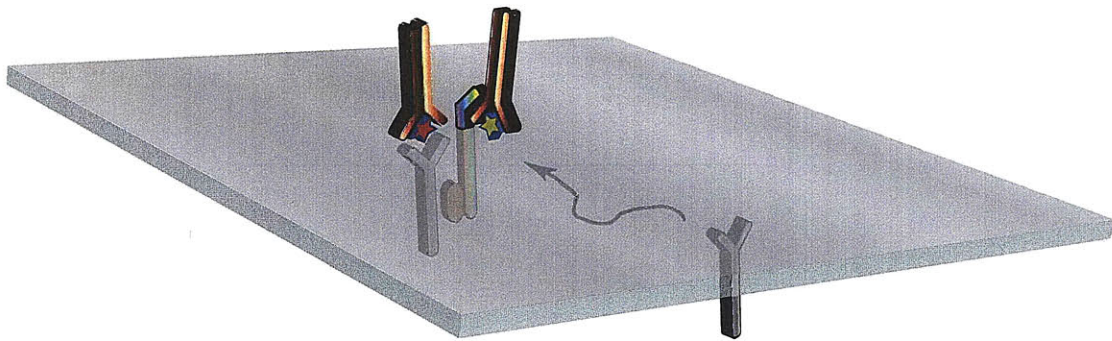
Fig. 6.9 Pictorial description of Model 3. Coreceptor associated Lck binds to the intracellular domain of TCR that is engaged with pepMHC. Due to steric reasons, extracellular domain of coreceptor can not bind to the engaged MHC, and so it binds to the bypassing pepMHC, forming "pseudodimer" like structure. When Lck is fully activated, it can cross-phosphorylated any TCR transiently engaging with the second pepMHC, leading to signal enhancement.
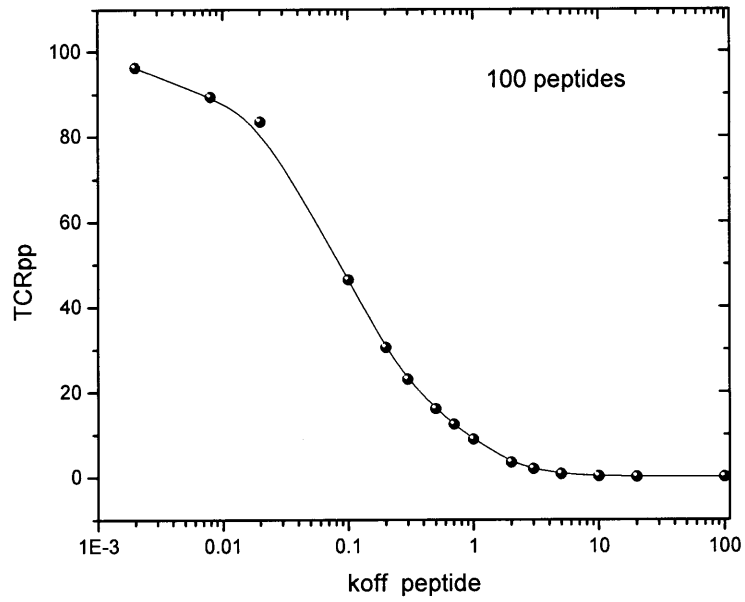
Fig. 6.10 Peptide potency as a function of k_{off} of the peptidesMHC-TCR for the model 3. Simulation of 1 μm^2 of the T-cell/APC contact area. Concentration of TCR is 300 per μm^2, concentration of coreceptor is 100 per μm^2.
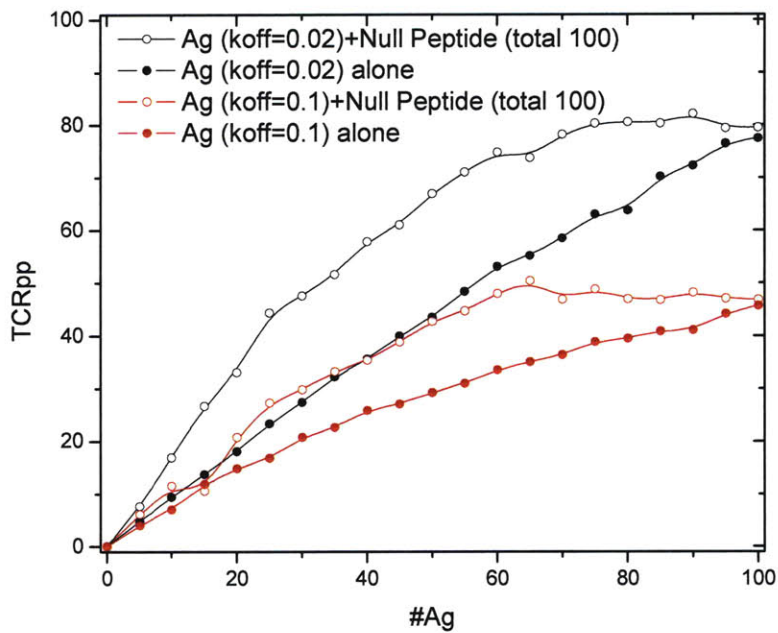
Fig. 6.11 Signaling levels of the agonist peptide alone (closed squares) compared to the mixture of agonist and non-cognate peptide (open circles) for simulation of model 3. A mixture delivers biger phosphorylation due to non-cognate peptides enhancement. Enhancement curves for two different agonist peptides are shown: strong agonist (k_{off}=0.02 s^{-1}) and weak agonist (k_{off}= 0.1 s^{-1})
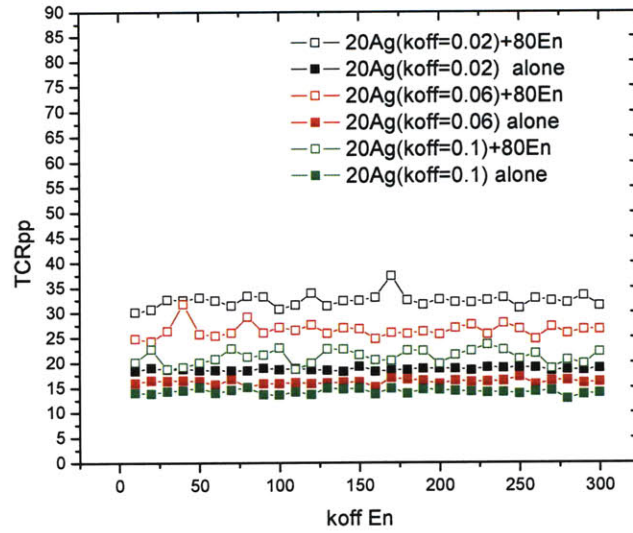
(a)
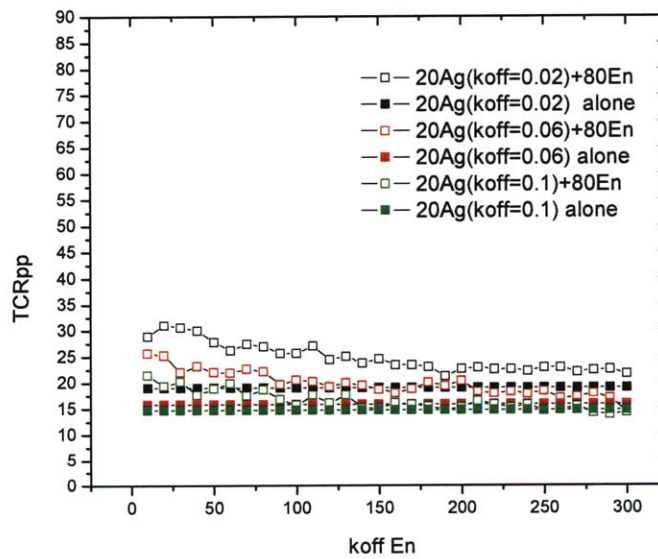
(b)

Fig. 6.12 The TCR phosphorylation in Model 3 when stimulated by mixture of agonist of indicated k_{off} and non-stimulating peptide (denoted as En, endogenous) or agonist alone. x-Axis shows the k_{off} for EnMHC-TCR interaction. Panel (a) corresponds to the high value of the rate of phosphorylation by fully active Lck ($k_{ph,act} = 300$), while panel (b) has low rate of phosphorylation by fully active Lck ($k_{ph,act}=100$).

PART III

Chapter 7

A model for genetic and epigenetic regulatory networks identifies rare pathways for transcription factor induced pluripotency

7.1 Introduction

Cellular states are plastic, and even terminally differentiated cells (e.g., B-cells) can be reprogrammed to pluripotency by ectopic expression of selected transcription factors (Aoi et al., 2008; Jaenisch and Young, 2008; Meissner et al., 2007; Park et al., 2008; Takahashi et al., 2007; Takahashi and Yamanaka, 2006; Wernig et al., 2007). This finding raises the possibility of creating patient-specific stem cells for regenerative medicine (Nishikawa et al., 2008). However, reprogramming efficiencies range from 0.0001 % to 29 % (Huangfu et al., 2008; Takahashi et al., 2007; Takahashi and Yamanaka, 2006; Yu et al., 2007), with most reports showing that successful induction of the pluripotent state is rare even if all required factors are present (Brambrink et al., 2008; Hanna et al., 2008). The genetic and epigenetic regulatory mechanisms that make reprogramming possible, and determine its efficiency, are poorly understood (Jaenisch and Young, 2008). Elucidating these mechanistic principles can help define optimal strategies for reprogramming differentiated cells, and answer fundamental questions regarding how cellular identity is maintained and transformed.

In spite of recent progress, our knowledge of the identities and functions of the genes and proteins involved in regulating the transformation of cellular identity is grossly

incomplete (Jaenisch and Young, 2008; Sridharan and Plath, 2008; Stadtfeld et al., 2008). Thus, it is not yet possible to construct a detailed molecular mechanistic description of how epigenetic modifications and expression of master regulatory genes are controlled. However, ectopic expression of the same transcription factors can reprogram different cell types (Aoi et al., 2008; Hanna et al., 2008; Takahashi and Yamanaka, 2006), and the genetic and epigenetic transformations observed during reprogramming of diverse differentiated cells share many common features (Brambrink et al., 2008; Egli et al., 2008; Jaenisch and Young, 2008; Maherali et al., 2007; Meissner et al., 2008; Mikkelsen et al., 2008; Mikkelsen et al., 2007). These common observations can be the basis for developing a conceptual understanding of the architecture of the genetic and epigenetic networks that regulate transcription factor induced reprogramming and establish cellular identity during differentiation.

We have taken a step toward this goal by developing a computational model that is consistent with, and suggests general mechanistic explanations for, empirical observations of transcription factor induced reprogramming. The model makes experimentally-testable predictions. If validated, descendents of this model could also provide insights into the aberrant de-differentiation events which characterize some of the most malignant cancers.

7.2 Model Development

Elegant theoretical models for the molecular regulatory networks responsible for stem cell renewal and differentiation and the population dynamics of these processes have been created (Cinquin and Demongeot, 2005; Jones and Simons, 2008; Jones et al., 2007; Qu and Ortoleva, 2008; Winkler et al., 2007). Our goal is different. We aim to develop a model for the architecture of coupled epigenetic and genetic networks which describes large changes in cellular identity (e.g., induction of pluripotency by reprogramming factors). Although the general principles of interactions between genetic and epigenetic layers of regulation have been described (Goldberg et al., 2007; Loeffler and Roeder, 2002), no computational model has been developed to study the outcomes of such interactions and their biological consequences. Such a computational model would be a useful complement to experiments in understanding the processes that occur during

reprogramming of differentiated cells, and why reprogramming is rare. Here, we propose, to our knowledge, the first computational model that describes how cellular identity changes by creating a mathematical description of interactions between epigenetic and genetic networks. Our goal is not to describe the details of how specific regulatory proteins interact, but rather, to understand general principles underlying how cellular states evolve upon ectopic expression of certain types of genes. The concise model we have developed explains why reprogramming probability is low, and makes experimentally testable predictions.

Almost all cells in a multi-cellular organism share the same DNA sequence. Yet, different cell types express distinct genes and perform different functions. Epigenetic modifications are major regulators of cell-type specific gene expression. They function by packaging DNA into configurations that allow only some genes to be expressed, while other genes are tightly packed into heterochromatin structures that hinder access of most transcription factors (Henikoff, 2008). Changes in cellular identity during developmental differentiation or transcription factor induced reprogramming require modification of the epigenetic state of the cell. The maintenance and alteration of cellular identity is regulated by a complex set of interactions between developmentally important genes, chromatin modifiers, transcription factors *etc.*, the details of which remain unknown. Toward developing a model for the architecture of these complex regulatory networks we consider only the developmentally important genes. For simplicity, each ensemble of genes responsible for maintenance of a particular cellular identity (e.g., Oct4, Sox2, *etc.*, for pluripotency) is described as a single module (Fig. 7.1a). Theoretical justification for treating genes that control the embryonic stem (ES) cell state as a collective unit exists (Chickarmane et al., 2006). We also carried out some studies with each module consisting of a small number of genes (see Fig. 7.6 and corresponding discussion below).

ES cells can differentiate in to various lineages. Upon further differentiation, cells become more restricted. For example, hematopoetic stem cells can differentiate in to T and B-lymphocytes, but not neural cells. Therefore, in our model, we arrange gene modules in a hierarchy (Fig. 7.1a). Although each cell state can potentially differentiate in to many branches, without loss of generality, we consider two branches to emanate from each cell state. Thus, the cellular states are arranged on a Cayley tree. In our

model, a cell state (Fig. 7.1b) is specified by: i] the state of the epigenome, and ii] the expression levels of master regulatory genes.

Specification and regulation of the epigenome: The epigenome is specified by chromatin states. Histones with positive marks (e.g. H3K4me3) promote transcription, and histones with negative marks (e.g. H3K27me3) repress transcription (Kouzarides, 2007; Orford et al., 2008). Hypermethylated genes are also silent (Fouse et al., 2008; Meissner et al., 2008). Genes associated with both H3K4me3 and H3K27me3 simultaneously (bivalent marks) can recruit promoters, but transcription is suppressed (Bernstein et al., 2006; Efroni et al., 2008; Guenther et al., 2007). Based on these observations, in our model, each developmentally important gene module can adopt one of three possible epigenetic states. It can be silent either due to negative histone marks or DNA methylation (denoted as the "-1" state), marked positively by histone marks (denoted as the "+1" state), and marked bivalently (denoted as the "0" state). From the standpoint of gene expression, each module can be either actively transcribing (denoted as the "+1" state) or not (denoted as the "0" state).

During interphase, DNA with genes packaged in a way characteristic of the cell's identity manages gene transcription and protein synthesis. Before cell division, the chromosomes condense. During telophase at the end of mitosis, the prevailing protein environment could alter the chromatin states of decondensing chromosomes in a daughter cell, thereby modifying the epigenetic state of its DNA (Egli et al., 2008; Orford and Scadden, 2008). We divide the cell cycle in to two parts (Fig. 7.2). During phase one (termed interphase, for ease of reference), the epigenetic state cannot be modified and gene expression is subject to this constraint. In phase two (termed telophase, for ease of reference), the epigenetic state can potentially be altered by the protein environment established during the preceding interphase.

Chromatin state maps show that the ES state is characterized by an unusually large proportion of bivalent chromatin marks on developmentally important genes (Bernstein et al., 2006; Mikkelsen et al., 2007; Pan et al., 2007). Therefore, we define the ES state as one where the gene module controlling this state (expressing Sox2, Oct4, *etc.*) is in the open chromatin state and all other master regulator genes are bivalently marked (Fig. 7.1b, left panel). Since the identities of all master-regulatory genes are not yet

experimentally available, it should be noted that bivalency of *all* master-regulatory modules in the ES state is an assumption that extrapolates available knowledge to yet unidentified modules.

It is known that, as cells differentiate from the ES state, bivalently marked genes remain bivalent, acquire a positive mark, or are silenced by negatively marked histones or methylation (Bernstein et al., 2006; Mikkelsen et al., 2007; Pan et al., 2007). Other than pluripotent ES cells, upon receiving appropriate cues, a cell state can only differentiate in to other states in the same lineage. Upon differentiation from the ES state positive histone marks are removed at an earlier stage compared to silencing of genes by DNA methylation, and reactivation of DNA methylated genes is more difficult than those with negative histone marks. These facts are encapsulated in our model by the following rules regarding how proteins expressed by a particular gene module can modify epigenetic states during telophase (Fig. 7.3b): 1] They favor putting positive marks on the module that expresses them, which enables stable maintenance of cellular identity. 2] They favor putting negative histone marks on the modules regulating the immediate progenitor or an immediate "sibling" in the hierarchy; this hinders differentiation in to cells in competing lineages and accidental de-differentiation to the progenitor. 3] They favor putting bivalent histone marks on the modules that regulate immediate progeny, which keeps cells poised to differentiate. 4] They favor methylation of all modules that regulate cell states in competing lineages or less differentiated states in the same lineage. This has a similar effect as the marking of histones in rule 2.

Rules 1-3 are based on experimental facts, and concern how proteins expressed by a gene module can affect the histone marks of only modules that regulate its immediate precursor, immediate progeny (see Fig. 7.3b), or other states to which its precursor can differentiate (i.e., "nearest neighbors" on the hierarchy of gene modules shown in Fig. 7.1a).

Rule 4 states that proteins expressed by a gene module favor silenced chromatin state of gene modules that are distal from it in the hierarchy by DNA methylation (Fig. 7.3b). Although there are no experimental measurements showing that methylation of unrelated lineages is directly caused by master-regulatory genes of current cell state, this rule is motivated by the global DNA methylation of genes of unrelated lineages observed

upon cell differentiation (Maherali et al., 2007; Meissner et al., 2008) and the fact that global DNA hypomethylation blocks differentiation (Jackson et al., 2004). To further investigate the effect of such long-range interactions, we have perturbed the formulation of rule 4 in different ways. We find that unless long-ranged nature of rule 4 is included, the *in silico* reprogramming trajectories exhibit features which are inconsistent with experimental observations. In particular, stable expression of protein products of the ES master-regulatory module becomes possible within the first reprogramming cycle, in contradiction with the observation that endogenous Oct4 is expressed shortly before completion of reprogramming after at least 12 days of action of reprogramming factors (see, for example Fig. 2 in (Jaenisch and Young, 2008) and references therein). Our computational results are also inconsistent with this observation if we allow proteins expressed by a module to put bivalent marks on all modules that regulate states in the lineage that are below it, rather than just the immediate progeny (rule 3 above).

Specification and regulation of gene expression: In our model, gene expression during interphase is subject to constraints imposed by the epigenetic marks as follows: 1'] if a gene module is positively marked, its expression is favored. Expression of bivalently marked gene modules is not favored, but it is not as strongly suppressed as modules that are negatively marked or DNA methylated (see Eq. 3 in Methods). 2'] Diverse experimental data (Briscoe et al., 2000; Rekhtman et al., 1999) show that, due to effects such as feedback regulation, *etc.*, expression of genes from competing lineages is mutually repressed. For example, GATA-1, erythroid lineage specific gene, and PU-1, transcription factor for genes of myeloid lineage are among the most studied master-regulatory genes. They posses typical properties attributed to the master-regulators in this manuscript: they enhance their own expression (Nishimura et al., 2000; Okuno et al., 2005) and mutually antagonize each others' activity (Cinquin and Demongeot, 2005; Rekhtman et al., 1999; Roeder and Glauche, 2006). We thus impose such mutually repressive interactions to gene modules that regulate directly competing cellular states (i.e., nearest neighbors in the hierarchy in Fig. 7.3a).

Rules 1-4 noted above are meant to describe how the epigenetic state is maintained and how it could evolve due to protein products of signaling events or ectopic

expression of transcription factors. During telophase, there could be a "tug of war" between the epigenetic state preferred by newly expressed proteins and that preferred by proteins expressed in accord with the preceding epigenetic state (Orford and Scadden, 2008). Similarly, rules 1' and 2' could lead to a tug of war between expression of different genes. Our computations reveal possible outcomes of these battles.

The epigenetic modifications during telophase or gene expression patterns during interphase are simulated on a computer using a Monte-Carlo algorithm, with rules 1-4 and 1'-2' represented as effective Hamiltonians (Eqs. 2-3, Methods). We specify the initial epigenetic state of the cell or the proteins that have been expressed in the previous interphase (including signaling products and ectopic expression of transcription factors). If the gene expression pattern is specified, simulation of telophase results in an epigenetic state that becomes the input for simulation of the next interphase, and so on (see Methods).

7.3 Results: Differentiation

ES cells are cultured in specific media (e.g., containing LIF/BMP4 for mouse ES cells) to prevent differentiation (Ying et al., 2008). The medium inhibits a self-induced differentiation pathway. We represent this feature by assuming that proteins expressed by the module regulating the ES state favor putting positive chromatin marks on gene modules regulating immediate progenies if LIF, *etc.* are absent. Simulations of this situation show (Fig. 7.4) that, as in experiments (Jaenisch and Young, 2008), ES cells differentiate randomly to one of their progeny.

Our model exhibits robust differentiation (forward programming) to specific cell states when the appropriate cues are delivered. Appropriate cues are expression of proteins (e.g., signaling products) that become available during interphase. In the next telophase, these proteins favor putting positive histone marks on the gene module regulating the appropriate progeny of the current cellular state (rule 1). Results from our computer simulations demonstrate that our model exhibits high-fidelity responses to such differentiation cues. This is consistent with the experimental observation that overexression of the master-regulatory genes of desired lineage leads to predominant differentiation in that direction (David et al., 2008; David et al., 2009). This result is

relevant because practical use of induced pluripotent cells will involve differentiating them to desired cell types. We also find an exponential decay of the number of progenitor cells (with a signal strength-dependent lifetime), as has been noted before (Johnston et al., 2007).

7.4 Results: Reprogramming

We simulate reprogramming experiments by starting with a terminally differentiated cell state where genes from other lineages, *etc.*, have been epigenetically silenced. Our basic premise is that terminally differentiated cells can reprogram because protein products of the ectopically expressed genes can potentially alter the epigenetic state of the cell as a cell progresses through the telophase. In our low resolution model, we identify genes not by names, but rather by their functional properties. We presume that Klf4 and c-Myc are important ingredients of the reprogramming "cocktail" because they promote progression through the cell cycle, and this provides more opportunities for the other reprogramming factors to perturb the epigenome during telophase. This functional identification of Klf4 and c-Myc makes our model general, and is validated by experiments showing that shutting down p53 abrogates the need for Klf4 and c-Myc for reprogramming (only Oct4 and Sox2 required) precisely because this also allows faster progression through the cell cycle (Banito et al., 2009; Hong et al., 2009; Kawamura et al., 2009; Marion et al., 2009; Utikal et al., 2009). Oct4 and Sox2 have an enormous number of binding targets on the DNA, and are responsible for maintenance of the ES state which likely implies multiple interactions with master-regulatory genes. We therefore identify the ectopic expression of these genes with the function of being highly likely to perturb the epigenome during telophase.

Each gene module in our model corresponds to an ensemble of carefully tuned mutually interacting master-regulatory genes that govern a particular cellular identity. At the moment, not all of the master-regulatory genes of cellular states are experimentally identified, thus we use gene modules to represent these ensembles in a general way. Even though products of ectopically expressed Oct4 and Sox2 have numerous targets (Wilson and Koopman, 2002), it is unlikely that the epigenetic state of many such sets of genes will be simultaneously altered. Thus, in order to mimic the effect of reprogramming

factors, we randomly pick one epigenetically silenced gene module and change its state to correspond to open chromatin. To examine the effects of overexpression of ectopic genes, we also study the consequences of multiple epigenetic transformations at a time (see Fig. 7.6 and discussions below).

Starting with a terminally differentiated state we perturb the epigenome as described above, and then simulate the next gene expression phase where both the module regulating the terminally differentiated state and the one which was transformed to open chromatin status can express proteins according to rules 1'-2' (or Eq. 3). The protein atmosphere thus generated becomes the input to simulation of the next telophase according to rules 1-4 (or Eq. 2). This can then potentially establish a new epigenetic state which becomes input to simulation of the next gene expression phase; i.e., the genetic and epigenetic states are allowed to come to a new balance. Then, the epigenetic state of another randomly picked silent gene module is changed to open chromatin because of the effects of reprogramming factors. This procedure is continued.

We carried out 10, 000 independent replicate simulations of the effects of ectopic expression of reprogramming factors on a differentiated cell in a model with four levels in the hierarchy of cellular states. Results from each simulation describe the fate of a single cell in a population. Only 3 out of 10, 000 "cells" successfully reprogrammed; i.e, as in experiments, reprogramming is rare. The percentage of cells that reprogram depends upon the number of levels in the hierarchy (0.0001 % and 2 % of the cells reprogram successfully for a five-level and three-level hierarchy, respectively). This suggests that reprogramming efficiency should improve for less differentiated cells. This remains to be demonstrated directly in a well-defined lineage such as the hematopoietic system. However, some support for this idea exists. Hanna et al. demonstrated a notable increase in the efficiency of reprogramming B cells upon Pax5 knockdown (17). Loss of Pax5 had been previously shown to cause dedifferentiation of B cells to a common progenitor that upon transplantation allowed T cell development (Nutt, 2008).

We report results for models consisting of 3-, 4- and 5-levels in the hierarchy of gene modules, but in real organisms the depth of the differentiation tree could be as large as tens of levels (Matthew and Brian, 2006). Since our results indicate that reprogramming efficiency decreases quickly with the increase in the depth of the

hierarchy, it is natural to ask why reprogramming is at all feasible. The reason is that master-regulatory genes that regulate closely related states are not mutually exclusive sets of genes. The difference between genes that regulate closely related cellular states can be as small as one or two genes (Nutt, 2008). However, genes that regulate cellular states distal in the hierarchy are not correlated in this way. As our model does not treat correlations between genes that regulate closely related states, in effect, each gene module in our model represents master regulatory genes that control the identity of a number of cellular states that have many master regulatory genes in common. Thus, a 5-level hierarchy in our model might represent a 50-level depth of differentiation in a real organism.

The results reported above were obtained for specific values of parameters (Table 1) which represent rules 1-4 and 1'-2' (Eqs., 2-3 in Methods). Our simulation results are consistent with diverse experimental observations (see Table 2 and discussion below) only if the methylation constraints (rule 4) and mutual repression of expression of gene modules (rule 2') are relatively strong effects (i.e. H>G and J>F, see Table 1 and parameter sensitivity in SI for further details). As long as these two conditions are met, the specific choice of parameter values only alters the quantitative value of the number of successfully reprogrammed cells, but reprogramming to the ES state remains rare.

Our simulation results suggest a mechanistic explanation for why reprogramming is so rare. When reprogramming factors attempt to change cellular identity by altering the epigenetic state of a previously silenced gene module, the probability of success depends upon the position of this module relative to the one that regulates the terminally differentiated state. We find that the position of the module whose epigenetic state is altered can belong to one of three categories (Fig. 7.5a).

Suppose this gene module regulates a cellular identity in a different lineage from the terminally differentiated state. In the next interphase, both modules can express proteins as there are no mutually repressive interactions between them. In the subsequent telophase, proteins expressed by each module would favor epigenetic silencing of the other (rule 4). Expression of proteins characteristic of a cell type from a different lineage does not favor reprogramming because it leads to cell death or arrest in our model. Cell death could be mediated by various mechanisms including genetic instabilities if the two

open gene modules send conflicting instructions to housekeeping genes. Of course, there is also the chance that the cell will be rescued by stochastic expression of some master-regulatory gene, or that the cell will assume an "intermediate" cell state without master regulation that could be viable, but does not reprogram, such as some arrested states(Mikkelsen et al., 2008); finally, there is a possibility that two master regulators will not repress each other in full, but some minuscule amount of expression of both will remain thus, arresting the cell. Within the framework of our model we do not distinguish between these possibilities, and classify cells in all these unusual, dead, or arrested states to be dead/arrested.

The gene module whose epigenetic state is altered by reprogramming factors could be in the same lineage as the differentiated cell, but not be its sibling or progenitor. In the following interphase, this module and the one that regulates the terminally differentiated state can both express proteins. In the subsequent telophase, according to our model, protein products of the gene module regulating the terminally differentiated state will favor epigenetic silencing of the module that was turned on by the action of reprogramming factors (rule 4). But, the opposite is not true because the cellular state regulated by the gene module whose epigenetic state was altered by reprogramming factors could potentially differentiate to the terminally differentiated cell type. Thus, the altered gene module will be silenced again, and the cell remains terminally differentiated.

Reprogramming factors could also change the epigenetic state of a previously silenced gene module which regulates an immediate sibling or the progenitor of the terminally differentiated state. In the subsequent interphase, these two gene modules with open chromatin status will not simultaneously express proteins at high levels. This is because gene modules that are "nearest neighbors" in the hierarchy mutually repress each other (rule 2'). If the dominantly expressed gene module (determined stochastically) is the one which regulates a sibling or the progenitor of the terminally differentiated state, then during the next telophase its products will establish epigenetic marks consistent with a new identity (rule 1). Thus, with a probability determined by stochastic effects, a step toward reprogramming can occur via trans-differentiation or de-differentiation.

These arguments suggest that a step toward reprogramming occurs with significant probability only if the epigenetic state of a gene module regulating a sibling or

progenitor of the differentiated cell is changed to open chromatin status by reprogramming factors. This is a rare event in our simulations where the set of master regulator genes that determine a cellular identity are considered to be one gene module. In reality, this is even less likely because it requires reprogramming factors to orchestrate changes to a set of master regulator genes synchronously. For successful reprogramming to the ES state, a sequence of such rare events must occur in a particular cell. This is because after a step toward reprogramming occurs, the partially reprogrammed cell is subject to all the constraints discussed above. Therefore, although cellular identity is plastic, reprogramming a terminally differentiated cell to the ES state is rare and requires many cell cycles.

Two examples of how states evolve under the influence of reprogramming factors in our simulations are shown in Fig. 7.5b. The first example shows a "cell" that does not successfully reprogram, as after a successful trans-differentiation, ultimately the cell is arrested/dead. In the second example reprogramming to the ES state occurs successfully, and it shows an interesting feature. At an intermediate time point, before the ES state is realized, reprogramming factors have turned on expression of the endogenous gene module that regulates the ES state. But this is transient, as this module is quickly silenced. We find that, unless proteins expressed by each gene module can DNA methylate genes that are distal in the hierarchy of states (rule 4), expression of endogenous genes that regulate the ES state can occur early and prior to the temporal increase in the number of bivalently marked genes observed during reprogramming. In other words, our model recapitulates the observation that endogenous expression of Oct4 and Sox2 is the last step toward reprogramming only if the DNA methylation constraint is "long-ranged". Thus, the model suggests that transient blocking of de novo methyltransferases might allow endogenous expression of Oct 4, Sox2, *etc.*, at intermediate time points. This is consistent with the observation that DNA methyltransferase and histone deacetylase (HDAC) inhibitors, such as valproic acid (VPA), an HDAC inhibitor, improve reprogramming efficiency (Huangfu et al., 2008).

Our model predicts that reprogramming occurs via a sequence of trans-differentiations to immediate siblings or de-differentiations to immediate progenitors in the hierarchy of cellular states. *Note, however, that our results do not imply that pure*

differentiated states will be observed as reprogramming occurs. Oct4, Sox2, *etc.*, have numerous targets, and so genes from unrelated lineages will transiently be expressed during reprogramming to the ES state (22). But, the entire set of master regulatory genes for a cellular state from a different lineage will not be expressed.

We illustrate this point by showing computer simulation results from a model where we consider each gene module to be comprised of three individual genes (Fig. 7.6). Reprogramming factors can attempt to change the epigenetic state of the individual genes randomly as before. However, in this more complex model, if we allow only one gene's epigenetic state to be modified in every telophase, reprogramming becomes so rare that we cannot observe it in a realistic computer simulation time. So, we allowed a larger number of transformations per cycle. Choosing this number to be too large corresponds to overexpression of reprogramming factors, and this severely hinders reprogramming (supplementary information, section 3). For the results shown in Fig. 7.6, we randomly pick 12 genes and change their epigenetic states during each simulated telophase. We assume that the entire set of genes comprising a module must be expressed for its products to regulate the epigenetic or genetic network. This is consistent with combinatorial control of regulation.

Fig. 7.6a shows two examples of *in silico* cells that successfully reprogram to the ES state. Reprogramming takes place via a sequence of trans-differentiation and de-differentiation events wherein the entire set of genes that regulate a progenitor or sibling of the previous cellular state is expressed. But, the intermediate states are not pure differentiated states as some genes from unrelated lineages are also turned on at the same time (as observed in experiments (Mikkelsen et al., 2008)). If the terminally differentiated state in our simulations is analogous to a B cell, our simulations predict that all successfully reprogrammed cells must transit through an impure state where all the genes regulating the hematopoetic stem cell state are turned on (as in Fig. 7.6a).

Although beyond the scope of this work, it would be reasonable to test this prediction by applying a cre-lox based lineage-tracing approach. Using one or more stem/progenitor specific promoters that are inactive in the terminal state (e.g., B cell), in combination with a lox-STOP-lox reporter, one could retrospectively determine whether all the resulting iPS cells are labeled and hence have transiently expressed markers of

earlier stages within the same lineage. An unrelated cell type, such as fibroblasts, should generate unlabeled iPS cells because it would not be expected to transition through hematopoietic progenitor stages and hence serve as an appropriate control.

The results depicted in Fig. 7.6 could also potentially be assessed quantitatively in experiments where the temporal evolution of the gene expression patterns of a number of successfully reprogrammed cells is observed. Consider a state where the master regulator genes corresponding to a particular cellular identity are all expressed. One could then ask: when these genes are subsequently silenced during reprogramming, which complete set of master regulatory genes start expressing proteins? One could ask this question at various times during reprogramming and in various successfully reprogrammed cells. This would enable calculation of the following four point correlation function (C):

$$C(i,j;t,t+\Delta t) = \left\langle \delta_{Si,0}(t+\Delta t)\delta_{Sj,1}(t+\Delta t).\delta_{Si,1}(t)\delta_{Sj,0}(t) \right\rangle \qquad (1)$$

where δ is the Kroenecker delta, t is time, t+Δt is a later instant in time during reprogramming (a cycle in our simulations), i and j are labels of two genes, and S_i is either 1 or 0 depending upon whether the ith gene is expressing proteins or turned off.

Our computer simulations predict (Fig. 7.6b) that, at each stage of reprogramming, the correlation function would have high values for genes from lineages related to the terminally differentiated starting point and low values for genes of unrelated lineages. We hope that this prediction can also be assessed in future experiments. This could involve permanent labeling as mentioned above, or possibly, in the long-term, real-time monitoring of cell state transitions.

7.5 Discussion

To the best of our knowledge, we have developed the first computational model that describes how terminally differentiated cells may be reprogrammed by expression of ectopic genes. This is achieved by a mathematical description of interactions between epigenetic and genetic networks of master-regulatory genes that govern specific cell states. The model also describes differentiation in accord with experiments. Our model describes cellular states as attractors on a generalized landscape of all possible genetic/epigenetic configurations. Cellular states are stable, self-renewing states unless a perturbing signal (either differentiation cue or reprogramming factors are introduced).

As summarized in the table 2, major features of the reprogramming process are explained by our results and the mechanism of reprogramming it suggests. For instance, different cell types can be reprogrammed with the help of the same set of factors(Aoi et al., 2008; Hanna et al., 2008; Maherali et al., 2007) because ectopic expression of genes that have many targets (e.g., Oct4 and Sox2) can perturb the epigenetic state regardless of the identity of the starting differentiated cell type. The importance of fast progression through the cell cycle (due to cMyc, Klf4, or p53 knockdown) is because this offers more opportunities for epigenetic transformations during telophase. The important experimental observation that endogenous Oct4 and Nanog expression (Jaenisch and Young, 2008) occurs just prior to complete reprogramming is also recapitulated by our model. The stochastic nature of the reprogramming process(Hanna et al., 2009) and its low yield (Jaenisch and Young, 2008) are because only a few types of trajectories can lead to successful reprogramming, and they are realized rarely by stochastic perturbation of the epigenome by the reprogramming factors. Our model predicts the nature of these rare trajectories to be those that progress through reprogramming via de-differentiation to closely related cell types (immediate progenitors or siblings in the hierarchy). Ways to directly test this prediction are suggested. However, any feature that involves a specific molecular interaction between specific molecules is not described by our model.

In our current model, we consider states with genes that express proteins with conflicting demands to die/arrest. In reality, some of these situations can give rise to steady states that do not arrest or reprogram (such as the recently studied BIV1, MCV8, *etc.*, cell lines) (Mikkelsen et al., 2008). The ideas emerging from our model are consistent with observations made by manipulating these trapped states.

For example, consider the observation that removing reprogramming factors allows cells from the BIV1 cell line (isolated during reprogramming of B lymphocytes) (Mikkelsen et al., 2008) to reprogram to the ES state. This suggests that overexpression of reprogramming factors prevents these cells from reprogramming to the ES state. Our model suggests that this could be due to two reasons. First, over expression of reprogramming factors (which have many targets) could simultaneously change the epigenetic states of a number of silenced genes to permissive chromatin status. Our simulations of the model shown in Fig. 7.6 with a large number of such simultaneous

transformations (e.g., 22 at a time, rather than 12 at a time used for Fig. 7.5) prevents successful reprogramming because of the large probability of obtaining dead or arrested states. As noted above, one of these states that cannot reprogram could correspond to the BIV1 cells.

Secondly, our model describes how lowering expression of reprogramming factors in BIV1 cells could enable reprogramming. In our simulations, we consider proteins expressed during each interphase to act on the epigenome to reach a new balance which then leads to a corresponding protein expression pattern before another epigenetic transformation can occur due to the action of reprogramming factors. This is analogous to assuming that the reprogramming factors can act to change the epigenetic state of a set of master regulator genes rarely. If reprogramming factors are grossly overexpressed, this would not be true. So, before a new protein expression pattern could be expressed consistent with a newly acquired epigenome (say, de-differentiation to a progenitor), another epigenetic transformation would occur, and the whole cycle would start again. Simulation results showing this effect upon overexpression of reprogramming factors are depicted in Fig. S7.4b. Removing reprogramming factors could potentially allow reprogramming of cells trapped in such an infinite loop.

Our low-resolution model for the architecture of genetic and epigenetic regulatory networks that determine how cellular identities change is consistent with diverse observations (Table 2). In formulating this model, we ruled out many models that were inconsistent with known experimental results, but we cannot rule out all other possible models. Therefore, the predictions of the model (noted earlier) need to be experimentally tested (perhaps in ways that we have suggested) to either falsify it or encourage studying it further. If tested positively, the suggestions emerging from our model regarding ways to enhance reprogramming yields should be further explored. It would also be interesting to study other transcription factor induced cell state conversions (Davis et al., 1987; Xie et al., 2004) within the conceptual and computational framework we have developed for how cellular identity is transformed. In particular, recent results of direct conversion between exocrine and endocrine cells through ectopic expression of three alternative transcription factors (Zhou et al., 2008) should be examined.

It would be interesting to further investigate several assumptions adopted in the model for the lack of specific information about individual master-regulatory modules. For example, maximum expression levels of different master-proteins within different modules could differ, as well as coupling between genetic and epigenetic networks could be different for different modules. Also, we assumed that every simulated cell (as represented by a simulated trajectory) has the same level of expression of reprogramming factors while in reality cells can be transfected in a heterogeneous fashion. Also, the difference in viral integration sites in different cells could lead to the different expression levels of exogeneous genes thus making effect of reprogramming factors heterogeneous across the population. In a sense then, we have studied those cells which have expressed reprogramming factors at levels above a threshold. It would be interesting to further explore the consequences of such heterogeneity. Another avenue for further exploration lies in defining the notion of time during the reprogramming process, in this work cell cycling has been adopted as a measure of time required for reprogramming while in reality cells cycle with non-equal rates determined from some form of cell division rate distribution (simplest form would be an exponential distribution). It would be interesting to see applicability of the 4-point correlation function based analysis for the situation when cell cycling rates are not identical. Finally, de-silencing action of reprogramming factors is assumed to be distributed randomly. It would be interesting to consider situations when de-silencing distribution is not uniform across the hierarchy. It is possible that non-uniform distributions can improve the reprogramming efficiency.

From the standpoint of statistical physics, our model couples a Potts model with short and long-ranged interactions in external fields (Eq. 2) with an Ising model with short-ranged interactions in an external field (Eq. 3). It may be fruitful to develop a deeper field-theoretic understanding of such models.

7.6 Simulation Methods

All simulations are carried out with the help of two hierarchical lattices because two lattices are required to properly describe the cell state as shown in Fig. 7.1b. In the simulation code provided in supplement, we consider 4 levels in the hierarchy (such as the one in Fig. 7.1b). Other possibilities (3 and 5 levels) have been considered also.

The epigenetic lattice has a discrete epigenetic state associated with each node (-1,0,+1). S^{epigen} = -1 corresponds to closed chromatin, S^{epigen} = 0 corresponds to bivalent chromatin and S^{epigen} = +1 corresponds to open chromatin. Genetic lattice describes expression of proteins from master-regulatory modules. It has discrete gene expression states associated with each node (0, +1). S^{gen} = 0 corresponds to the absence of any protein expression from the given gene, S^{gen} = +1 corresponds to the maximum protein expression from the gene.

In order to initialize simulations one has to specify either the epigenetic or genetic state of the lattice (see Fig. 7.7). If we start by specifying the protein expression pattern, computer simulations are carried out to determine the epigenetic state that is realized in telophase. A Monte-Carlo simulation algorithm is used in accord with the following Hamiltonian, with its four terms representing rules 1-4 (see Model development), respectively:

$$H[\{S_i^{ep}\}] = -G\sum_i <S_i^{gen}> S_i^{ep} + G \sum_{i,j \in sibling, progeny\, of\, i} <S_j^{gen}> S_i^{ep} + G \sum_{i,j \in parent\, of\, i} <S_j^{gen}> |S_i^{ep}|$$
$$+ H \sum_{\substack{j, i \notin progeny\, of\, j \\ and\; i \neq j}} (<S_j^{gen}> -a) S_i^{ep}$$

(2)

S_i^{ep} denotes the epigenetic spin state of the ith module, and S_i^{gen} specifies the protein expression level of the ith module. The angular brackets denote the average expression level of the jth module obtained during the preceding interphase, and could include protein products of ectopic genes or signaling events. $|S_i^{ep}|$ represents the absolute value of S_i^{ep}. The quantity G is a positive parameter that represents the strength with which the protein atmosphere can modify the epigenetic state by altering histone marks. H is a positive parameter that represents the strength of the DNA methylation constraint. The quantity, a, is a positive constant that favors values of S_i^{ep} < a if proteins expressed by gene, j, are present. As detailed in the supplementary information (section 2), the results of our simulations are inconsistent with experimental results if H is not greater than G. As long as H > G, our qualitative results do not depend upon the specific values of these parameters. The specific value of a does not affect qualitative results. Results presented in the main text are for a = 0, and G = 25, H = 40 (in units described below).

During simulation of the telophase, the epigenetic state S^{epigen} of each module fluctuates. The output of the telophase simulation is $< S^{epigen}>$, an average of these fluctuating values for each node of the lattice (i.e. for each module). Because we have a discrete representation for the epigenetic marks (+1, 0, or -1) while actually each gene bears multiple marks, using the average allows us to reflect intermediate levels of positive and negative histone marks on a gene. For example, an average value near zero for the epigenetic state of a gene module implies that both positive and negative marks are present on histones associated with it, a value close to one represents an open chromatin state, *etc*.

Average values of epigenetic state serve as input for simulation of interphase. If $<S^{epigen}>$ ~1 (gene is epigenetically available), than it will favor protein expression during the interphase in accord with the rules depicted on Fig. 7.3a. Similarly, if two neighboring states are epigenetically available, only one protein will be expressed due to mutual repression of neighboring master-regulators. Separate Monte Carlo simulations are carried out to establish gene expression patterns during interphase. The following Hamiltonian, with the two terms in it corresponding to rules 1' and 2' (see Model development), respectively, is used:

$$H\left[\{S_i{}^{gen}\}\right] = -F\sum_i (< S_i{}^{ep} > -b)S_i{}^{gen} + J \sum_{i \neq j \in nearest\ neighbors} S_i{}^{gen} S_j{}^{gen} \qquad (3)$$

The angular brackets denote the average value of epigenetic state of the ith module obtained during the preceding telophase. F is a positive constant that represents how strongly a protein is expressed or repressed if it is in open chromatin state or in heterochromatin, respectively. The parameter, b, is a positive constant; protein expression is favored if $<S_i{}^{ep}>$ > b. Note that the form of the first term in Eq. 3 implies that protein expression is more strongly repressed if a gene is packaged in heterochromatin compared to if it is bivalently marked. J represents the strength of mutual repression by other proteins. As detailed in the supplementary information (section 2), our results are inconsistent with experiments if J is not greater than F. As long as J >F, the specific values do not affect qualitative results. As long as the parameter b is larger than the typical size of fluctuations in $<S_i{}^{ep}>$ (~ 0.1), the specific

value of b does not affect qualitative results. Results presented in the main text correspond to b = 0.3, and F = 2000, J = 3000 (for units, see below).

Values of S_i^{gen} fluctuate during this Monte-Carlo procedure. The output of the simulation of the interphase is $<S_i^{gen}>$, which represents the average expression level of the regulatory protein in the interphase. These averages are further used in the next telophase simulation, thus, completing the cycle.

The Monte-Carlo algorithm is standard (Frenkel, 2002): the lattice spins (+1/0/-1 on epigenetic lattice; +1/0 on genetic lattice) are initialized randomly. The move consists of 1) randomly choosing the node on the lattice; 2) randomly deciding on the choice of new value of S_i for this node (i.e. if S_i^{epigen} was 0 then it can become -1 or +1 with equal probability; 3) energy for this configuration is computed according to the appropriate Hamiltonian; 4) attempted changes in state are accepted with probability equal to min [1, exp $\{[-\beta \Delta H \{S_i\}]$]. The parameter, β, is analogous to inverse temperature used in simulation of thermal systems, and sets the scale for the parameters, F, G, H and J. If we pick this effective temperature to be too high ($\beta \ll$ F, G, H, J), the system is disordered; specific cellular identities are not established and the model has no biological significance. We use $\beta = 1$ for results reported in the main text.

A computer code written using the C++ language is provided in the supplement allows calculation of all the results we report. For details regarding the output and input formats see the supplementary information.

Table 7.1: Parameters used to obtain the simulation results reported in the main text. Results do not change qualitatively as long as the parameters lie in the following ranges: H>G, J>F, 0.1<b<0.5 and 0<a<0.6

Parameter of the model	Value of the parameter
Protein action on epigenetic lattice	G=25
Mutual suppression by two proteins	J=3000
Action from epigenetic to genetic lattice	F=2000
Methylation strength	H=40
Minimal protein expression level required to actively affect epigenetic state of the gene	a=0
Minimal epigenetic availability of the gene required to allow protein expression	b=0.3

Table 7.2: Experimental features of reprogramming explained by the proposed model

Experimental reprogramming features explained by the model
Reprogramming takes *at least* 12 days of continuous cell transformation (Brambrink et al., 2008)
Low yield of reprogramming process(Jaenisch and Young, 2008)
Stochastic nature of reprogramming(Hanna et al., 2009)
The fact that the same gene cocktail can reprogram different terminal cell types(Aoi et al., 2008; Hanna et al., 2008)
The fact that immortalizing the cells should improve reprogramming, e.g. by knock-down of p53 gene(Hong et al., 2009; Kawamura et al., 2009; Marion et al., 2009)

7.6 Appendix to Chapter 7

7.6.1. Parameter sensitivity studies

Here we discuss how changes in the six parameters in our model (Eqs. 2 and 3 in the main text) affect the results of our computer simulations. The results shown in the main text correspond to the following parameter values: F=2000; J=3000; G=25; H=40; a=0; b=0.3

The effective temperature used in the simulations is T=1. For any other choice of the temperature, T^{new}, the parameters F,J,G,H have to be linearly scaled; i.e., $F^{new}=F*T^{new}$, *etc*, for the simulation results to be invariant.

The simulations results reported in the main text do not change qualitatively as long as the parameters lie in the following ranges: H>G>>T & J>F>>T, b lies between 0.1 and 0.5, and a between 0 and 0.6. Below, we provide a detailed description of why these parameter ranges are appropriate.

1a. Parameters involved in simulation of the epigenetic network (Eq. 2): H, G, and a.

The parameters, G and H, describe the coupling between genetic and epigenetic networks. For example, G is the strength with which a value of $<S_i^{gen}>$ = 1 (i.e., high protein expression by the ith gene module) favors a value of + 1 for $<S_i^{epigen}>$ (i.e., open chromatin) during simulation of the epigenetic network. H is the strength of the methylation constraint. It determines the strength with which proteins expressed by a gene module favor methylation of another gene module according to rule 4 in the main text. We have carried out simulations with H>G, H~G, and H<G. The simulation results reported in the main text correspond to H>G. The other two parameter regimes lead to results that are inconsistent with experimental findings as described below.

Consider situations where G and H are approximately of the same magnitude, H ~ G. Now, when proteins expressed by two distal gene modules are expressed simultaneously (for example, as a result of reprogramming factors' action), this leads to a tug-of-war which is resolved in several (typically, 2-3) cell cycles. Because H~G, stochastically, only one of the modules will be silenced when cell will achieve epigenetically/genetically balanced state. When H=G and two proteins are expressed at exactly the same level, the methylation (H) and self-support (G) will exactly cancel each

other and in the next cycle two protein will be expressed at exactly the same levels as in the previous cycle. However, due fluctuations, proteins never have *exactly* same expression levels. And the slightest differences in protein levels is amplified in the next cycle because H-G is positive for one protein (hence, this protein is more suppressed) and negative for the other protein (hence, it will be epigenetically more available in the next cycle).Thus, in several cycles balanced epigenetic/genetic state with only one master-protein expressed is achieved. (in the simulations reported in fig. S7.2a we perform epigenetic perturbation every 4 cycles in order to allow for cell state equilibration time). If we allow epigenetic perturbations to be more frequent, in this circumstance, each expressing gene module experiences methylation constraint from multiple other modules which results in silencing of all genes (death/arrest).

Therefore, in this parameter regime, random lineage switches occur during reprogramming and no specific path for reprogramming trajectories can be identified (see Fig. S7.2a). For example, lineage switches can occur from a fully differentiated state to a state in a different lineage that is almost fully reprogrammed. This corresponds to sudden changes in epigenetic patterns of the cell, rather than a continuous evolution of the number of bivalent domains. For G=H, we find that such cross-lineage jumps are dominant in our simulations, occurring 21 times more frequently than gradual evolution of epigenetic patterns in successfully reprogrammed trajectories. Thus, when H~G, "shortcuts" are dominant during simulations of reprogramming, leading to large and rapid "jumps" from a differentiated state to a nearly reprogrammed state. This is inconsistent with experimental observation of timelines of reprogramming with distinct changes in cell appearance and cell markers (Jaenisch and Young, 2008). Note that sudden jumps of the sort we see in our simulations with G~H have been observed upon treating fibroblasts with a drug called AZA (Meissner et al., 2008), which causes global demethylation (i.e., removes the methylation constraint). These large jumps do not lead to reprogramming. Based on these considerations, we believe that our simulation results for G~H are not consistent with experimental observations.

If H<G, because the methylation constraints are weak, we often see stable expression of the endogenous gene module regulating the ES state after one or two cycles, and without transitioning through intermediate states. For example, even if H is

only 3 times smaller than G, 8% of trajectories exhibit stable induction of the gene module regulating the ES state after the very first cycle of epigenetic perturbations. This is inconsistent with experimental results showing that endogenous expression of Oct4, Sox2, *etc.*, never appears before at least day twelve, and that reprogramming is associated with a gradual evolution of cell appearance and markers (1). For these reasons, we focused on the parameter regime H>G, for which the results reported in the main text are qualitatively robust. For example, if H = 60 and G = 35, rather than H = 40 and G = 25 as in the main text's figures, the reprogramming efficiency for a 4-level hierarchy is ~5 out of 10^4 cells and the qualitative pathways followed by successfully reprogrammed cells remains the same as discussed in the main text.

The parameter, a, describes the minimal level of gene expression that allows a particular gene module to exert methylation constraints on other genes. In our simulations, $a=0$, which means that, as long as protein is expressed, it can methylate other gene modules with a strength related to its expression level. Changing the parameter a between 0 and 0.6 does not alter qualitative results. For example, if a = 0.5, rather than 0 as in the main text, the reprogramming efficiency for a 4-level hierarchy of states is ~6 out of 10^4 cells, and the qualitative pathways followed by successfully reprogrammed cells remain the same as discussed in the main text.

1b. Parameters involved in the simulation of the genetic network: F, J and b.

Parameters F and J describe the coupling between the epigenetic and genetic network. Parameter F is the strength with which open chromatin (positive $<S_i^{epigen}>$) favors gene expression (positive S_i^{gen}) during the simulation of the genetic network. Parameter J describes the strength with which proteins expressed by two neighboring gene modules in open epigenetic states mutually repress each other. The choice of F and J is dictated by our assumption that two neighboring genes with open chromatin should with high probability express only one out of two proteins. This is to be consistent with experimental evidence that genes modules responsible for the competing lineages are mutually repressive (Cinquin and Demongeot, 2005). For this to be true, the parameter J needs to be larger than F.

Stochastically, during simulation of the genetic network, transitions can occur from a state where one of two mutually repressive gene modules is expressed to one

where the other is expressed. F and J must be much larger than the effective temperature of the simulations (T) in order to minimize such transitions during the simulation time as this would result in both proteins being simultaneously expressed. If the two gene modules thus expressing proteins are immediate siblings in the hierarchy of states, then rules 3 and 4 in the main text would result in silencing of both gene modules, leading to cell death/arrest. If the two gene modules were progenitor and immediate progeny, then the cell would remain in a terminally differentiated state because of the asymmetry of the methylation constraint for states in the same lineage. Fig. S7.3 shows an example of the latter situation for simulations carried out with values of F and J that are too small. Thus, if F and J are not sufficiently large, reprogramming is not possible. For intermediate values of F and J we find that reprogramming probability decreases (for F=1000 and J=2000 only about 4 cells out of 10^5 cells reprogram successfully), but the qualitative nature of the temporal pathways followed by successfully reprogrammed cells remain the same.

Protein expression by gene module i is favored when the value of $<S_i^{epigen}>$ representing the chromatin state of this gene is greater then b. The value of b has to be more than zero to ensure stability of the stem cell state in the absence of self-induced differentiation (e.g., when LIF, *etc.*, are present in the medium). Typical simulations show that, in the ES state, $<S^{epigen}(\text{ES gene module})> \sim 1$ and $<S^{epigen}(\text{all other gene modules})>$ fluctuates between -0.1 and +0.1. The parameter, b, must be larger than the size of these typical fluctuations in order to prevent spontaneous differentiation. Our qualitative results do not change if b lies between 0.1 and 0.5. For example, if b = 0.4, rather than 0.3 as in the main text, the reprogramming efficiency for a 4-level hierarchy of states is ~6 out of 10^4 cells, and the qualitative pathways followed by successfully reprogrammed cells remain the same as discussed in the main text.

7.6.2. Effect of increasing the frequency of action by reprogramming factors

Fig. S7.4a details the events that occur during one-level de-differentiation in our simulations with parameters that are consistent with experimental observations (see above). In our simulations, it takes two cell cycles to complete a successful de-differentiation or trans-differentiation. During the first cycle, reprogramming factors cause a new perturbation to the epigenome in the telophase and the perturbed epigenome

controls protein expression in the subsequent interphase. The resulting protein expression profile may not be in balance with underlying epigenetic state. During the subsequent telophase (which begins the second cycle), a new balance between the epigenetic and genetic states of the cell is established. Hence, in the subsequent interphase, the protein expression reaches accord with the prevailing epigenetic state. Note that in experiments, it is likely that more time (cell cycles) is required in order to balance the genetic and epigenetic states. For example, in nuclear transfer experiments it takes 12 to 20 cycles to achieve global demethylation of somatic DNA (Jaenisch and Young, 2008; Simonsson and Gurdon, 2004).

If the reprogramming factors are overexpressed, they will have a higher probability of altering the epigenetic states of gene modules more frequently than once in two cell cycles. In this situation, our simulations predict that reprogramming is not possible because cells will never be able to balance their genetic and epigenetic networks, as is shown on Fig. S7.4b. Any endogenous expression of Oct-4 or Nanog will only be observed transiently, because further epigenetic perturbations occur before the ES state can be stably established.

Overexpression of reprogramming factors should also lead to an increase in the number of genes that can be epigenetically altered in every cell cycle. In our simulations, this means that a greater number of genes would be epigenetically modified during each perturbation of the epigenome. The effects of this can be evaluated in a model where each module consists of three individual genes (Fig. 7.6 in main text). If only 12 genes are altered in a cycle (as in main text), 15 out of 100, 000 "cells" reprogram. But if 15, 20, or 22 genes are altered per cycle, out of 100, 000 "cells", only 5, 2, and 0 cells reprogram. Thus, overexpression of reprogramming factors hinders reprogramming.

7.6.3. C++ code that illustrates reprogramming process

C++ code is provided as supplementary file "reprog.cpp". The code allows one observe reprogramming trajectories. (Simply compile and then run the code after compiling).

Simulation shows progression of the cell states starting from terminally differentiated state under the action of reprogramming factors (they are turned on in the second cell cycle after stability of the terminal state by itself is illustrated).

The simulation outputs the average values of the genetic and epigenetic variable at each cycle in accord with the simulation flow chart.

So, for fully differentiated state genetic and epigenetic networks will look like:

GENETIC

```
                    0
          0                   0
      0         0         0         0
  1       0       0       0       0       0       0       0
```

EPIGENETIC

```
                   -1
         -1                  -1
     -1        -1        -1        -1
 +1      -1      -1      -1      -1      -1      -1      -1
```

Terminally differentiated state

The GENETIC diagram indicates that only proteins of "bottom left corner" master-regulatory gene module are expressed. And EPIGENETIC diagram indicates that chromatin open only for "bottom left corner" master-gene, while the rest of master genes are in the closed chromatin state.

Note that this state is stable state of the cell, "attractor" in the space of the cellular states, because it will renew itself, provided that cell is forced to proliferate.

Fully differentiated state will be recognized in the program output as having the following genetic/epigenetic structure:

GENETIC

```
                    1
          0                   0
      0         0         0         0
  0       0       0       0       0       0       0       0
```

EPIGENETIC

```
                        +1
            0                       0
       0           0          0           0
   0      0      0      0      0      0      0      0
```

ES state

Simulation begins in the fully differentiated state of the 4 levels hierarchy of cell states. Beginning from the second cell cycle reprogramming factors start acting. They randomly perturb epigenetic structure of the terminally differentiated cell. In the output one can see that randomly chosen epigenetic module will have its "-1" state flipped to "+1" state. Then, during the next cell cycle epigenetic and genetic will find assume balanced state, which in the majority of cases results in the completely silenced, unviable state such as follows:

GENETIC

```
                        0
            0                       0
       0           0          0           0
   0      0      0      0      0      0      0      0
```

Dead/arrested

EPIGENETIC

```
                        -1
            -1                      -1
       -1          -1         -1          -1
   -1     -1     -1     -1     -1     -1     -1     -1
```

This is indicative of cell arrest/cell death situation.

In order to start new reprogramming trajectory, one can simply restart the compiled code.

In the few cases one will observe one-level de-differentiation and in about 5-10 out of 10,000 cases one will see the complete reprogramming from terminal fully differentiated state to the ES cell state.

7.6.4. Population behavior of the cells

On Fig. S7.5 the diagram depicting transitions between stable cell populations under the action of reprogramming factors is presented. The starting point is terminally differentiated state depicted on the bottom of the picture. Possible transitions from this state are indicated by arrows. Probability to experience a particular transition is indicated by the % value on top of the corresponding arrow. For example, starting from the population of terminally differentiated cells, only ~7 % of cells successfully de-differentiate one level, while 67 % of cells either die or get arrested in the intermediate state. Further, one sees that out of one level de-differentiated cells only ~7% dedifferentiate one level further *etc*, thus giving very low total yield of reprogrammed cells at about 0.04%.

Table S1. Summary of efficiency of reprogramming in different parameter regimes.

	G<H	G~H	G>H
Efficiency of reprogramming (out of 10^4 cells)	4	84	0

Table S2. Dependence of reprogramming efficiency on number of genes perturbed during action of reprogramming factors for the model where each master-regulatory module consists of 3 genes.

Number of genes perturbed during action of reprogramming factors	8	10	12	15	20	22
Efficiency of reprogramming (out of 10^5 cells)	6	10	15	5	2	0

7.7 Bibliography

Abramowitz, M., and Stegun, I.A. (1972). Handbook of mathematical functions with formulas, graphs, and mathematical tables, 10th printing, with corrections. edn (Washington: U.S. Govt. Print. Off.).

Acar, M., Becskei, A., and van Oudenaarden, A. (2005). Enhancement of cellular memory by reducing stochastic transitions. Nature *435*, 228-232.

Allen, R.J., Frenkel, D., and ten Wolde, P.R. (2006). Simulating rare events in equilibrium or nonequilibrium stochastic systems. Journal of Chemical Physics *124*, 024102.

Altan-Bonnet, G., and Germain, R.N. (2005). Modeling T cell antigen discrimination based on feedback control of digital ERK responses. PLoS Biol *3*, e356.

Anikeeva, N., Lebedeva, T., Clapp, A.R., Goldman, E.R., Dustin, M.L., Mattoussi, H., and Sykulev, Y. (2006). Quantum dot/peptide-MHC biosensors reveal strong CD8-dependent cooperation between self and viral antigens that augment the T cell response. Proceedings of the National Academy of Sciences of the United States of America *103*, 16846-16851.

Aoi, T., Yae, K., Nakagawa, M., Ichisaka, T., Okita, K., Takahashi, K., Chiba, T., and Yamanaka, S. (2008). Generation of pluripotent stem cells from adult mouse liver and stomach cells. Science *321*, 699-702.

Artyomov, M.N. (2009). Comment on "reciprocal relations for nonlinear coupled transport". Phys Rev Lett *102*, 149701; discussion 149702.

Artyomov, M.N., Das, J., Kardar, M., and Chakraborty, A.K. (2007a). Purely stochastic binary decisions in cell signaling models without underlying deterministic bistabilities. Proceedings of the National Academy of Sciences of the United States of America *104*, 18958-18963.

Artyomov, M.N., Morozov, A.Y., and Kolomeisky, A.B. (2008). Molecular motors interacting with their own tracks. Physical review *77*, 040901.

Artyomov, M.N., Morozov, A.Y., Pronina, E., and Kolomeisky, A.B. (2007b). Dynamic properties of molecular motors in burnt-bridge models. Journal of Statistical Mechanics: Theory and Experiment *2007*, P08002-P08002.

Banito, A., Rashid, S.T., Acosta, J.C., Li, S., Pereira, C.F., Geti, I., Pinho, S., Silva, J.C., Azuara, V., Walsh, M., *et al.* (2009). Senescence impairs successful reprogramming to pluripotent stem cells. Genes Dev. *23*, 2134-2139.

Bateman Manuscript Project., Bateman, H., Erdélyi, A., and United States. Office of Naval Research. (1953). Higher transcendental functions (New York,: McGraw-Hill).

Berg, O.G., Paulsson, J., and Ehrenberg, M. (2000). Fluctuations and quality of control in biological cells: Zero-order ultrasensitivity reinvestigated. Biophysical Journal *79*, 1228-1236.

Bernstein, B.E., Mikkelsen, T.S., Xie, X.H., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bhalla, U.S., Ram, P.T., and Iyengar, R. (2002). MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. Science *297*, 1018-1023.

Brambrink, T., Foreman, R., Welstead, G.G., Lengner, C.J., Wernig, M., Suh, H., and Jaenisch, R. (2008). Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. Cell Stem Cell *2*, 151-159.

Briscoe, J., Pierani, A., Jessell, T.M., and Ericson, J. (2000). A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube. Cell *101*, 435-445.

Brower, R.C., England, R., Takeshita, T., Kozlowski, S., Margulies, D.H., Berzofsky, J.A., and Delisi, C. (1994). Minimal Requirements for Peptide Mediated Activation of Cd8(+) Ctl. Molecular Immunology *31*, 1285-1293.

Burroughs, N.J., and van der Merwe, P.A. (2007). Stochasticity and spatial heterogeneity in T-cell activation. Immunological reviews *216*, 69-80.

Campi, G., Varma, R., and Dustin, M.L. (2005). Actin and agonist MHC-peptide complex-dependent T cell receptor microclusters as scaffolds for signaling. pp. 1031-1036.

Chickarmane, V., Troein, C., Nuber, U.A., Sauro, H.M., and Peterson, C. (2006). Transcriptional dynamics of the embryonic stem cell switch. Plos Computational Biology *2*, 1080-1092.

Cinquin, O., and Demongeot, J. (2005). High-dimensional switches and the modelling of cellular differentiation. Journal of Theoretical Biology *233*, 391-411.

Craciun, G., Tang, Y.Z., and Feinberg, M. (2006). Understanding bistability in complex enzyme-driven reaction networks. Proceedings of the National Academy of Sciences of the United States of America *103*, 8697-8702.

Danos, V., and Laneve, C. (2004). Formal molecular biology. Theoretical Computer Science *325*, 69-110.

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

David, R., Brenner, C., Stieber, J., Schwarz, F., Brunner, S., Vollmer, M., Mentele, E., Muller-Hocker, J., Kitajima, S., Lickert, H., *et al.* (2008). MesP1 drives vertebrate cardiovascular differentiation through Dkk-1-mediated blockade of Wnt-signalling. Nature cell biology *10*, 338-345.

David, R., Stieber, J., Fischer, E., Brunner, S., Brenner, C., Pfeiler, S., Schwarz, F., and Franz, W.M. (2009). Forward programming of pluripotent stem cells towards distinct cardiovascular cell types. Cardiovascular research *84*, 263-272.

Davis, M.M., Krogsgaard, M., Huse, M., Huppa, J., Lillemeier, B.F., and Li, Q.J. (2007). T cells as a self-referential, sensory organ. Annu Rev Immunol *25*, 681-695.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a Single Transfected Cdna Converts Fibroblasts to Myoblasts. Cell *51*, 987-1000.

DeMond, A.L., Mossman, K.D., Starr, T., Dustin, M.L., and Groves, J.T. (2008). T Cell Receptor Microcluster Transport through Molecular Mazes Reveals Mechanism of Translocation. Biophysical Journal *94*, 3286-3292.

Dushek, O., Mueller, S., Soubies, S., Depoil, D., Caramalho, I., Coombs, D., and Valitutti, S. (2008). Effects of Intracellular Calcium and Actin Cytoskeleton on TCR Mobility Measured by Fluorescence Recovery. PLoS ONE *3*, e3913.

Dustin, M.L. (2009). The Cellular Context of T Cell Signaling. Immunity *30*, 482-492.

Ebert, P.J., Jiang, S., Xie, J., Li, Q.J., and Davis, M.M. (2009). An endogenous positively selecting peptide enhances mature T cell responses and becomes an autoantigen in the absence of microRNA miR-181a. Nature immunology *10*, 1162-1169.

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D.G., Buetow, K.H., *et al.* (2008). Global transcription in pluripotent embryonic stem cells. Cell Stem Cell *2*, 437-447.

Egli, D., Birkhoff, G., and Eggan, K. (2008). Mediators of reprogramming: transcription factors and transitions through mitosis. Nature Reviews Molecular Cell Biology *9*, 505-516.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science *297*, 1183-1186.

Evavold, B.D., Sloan-Lancaster, J., and Allen, P.M. (1994). Antagonism of superantigen-stimulated helper T-cell clones and hybridomas by altered peptide ligand. Proc Natl Acad Sci U S A *91*, 2300-2304.

Faeder, J.R., Blinov, M.L., and Hlavacek, W.S. (2009). Rule-Based Modeling of Biochemical Systems with BioNetGen. Methods in Molecular Biology, 113-167.

Fleming, G.R., and Ratner, M.A. (2008). Grand challenges in basic energy sciences. Physics Today *61*, 28-33.

Fouse, S.D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., Jaenisch, R., and Fan, G.P. (2008). Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3K4/K27 trimethylation. Cell Stem Cell *2*, 160-169.

Frenkel, D., Smit B. (2002). Understanding Molecular Simulations. Academic Press.

Gao, G.F., Rao, Z., and Bell, J.I. (2002). Molecular coordination of alphabeta T-cell receptors and coreceptors CD8 and CD4 in their recognition of peptide-MHC ligands. Trends in immunology *23*, 408-413.

Gardiner, C.W. (2004). Handbook of stochastic methods : for physics, chemistry, and the natural sciences, 3rd edn (Berlin ; New York: Springer).

Gillespie, D.T. (1977). Exact Stochastic Simulation of Coupled Chemical-Reactions. Journal of Physical Chemistry *81*, 2340-2361.

Gillespie, J.H. (2004). Population genetics : a concise guide, 2nd edn (Baltimore, Md.: Johns Hopkins University Press).

Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: A Landscape Takes Shape. Cell *128*, 635-638.

Goldbeter, A., and Koshland, D.E. (1981). An Amplified Sensitivity Arising from Covalent Modification in Biological-Systems. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences *78*, 6840-6844.

Gomez-Uribe, C., Verghese, G.C., and Mirny, L.A. (2007). Operating regimes of signaling cycles: Statics, dynamics, and noise filtering. Plos Computational Biology *3*, 2487-2497.

Gradshtein, I.S., Ryzhik, I.M., and Jeffrey, A. (2000). Table of integrals, series, and products, 6th edn (San Diego: Academic Press).

Grakoui, A., Bromley, S.K., Sumen, C., Davis, M.M., Shaw, A.S., Allen, P.M., and Dustin, M.L. (1999). The immunological synapse: a molecular machine controlling T cell activation. Science (New York, N.Y *285*, 221-227.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.

Hamad, A.R.A., O'Herrin, S.M., Lebowitz, M.S., Srikrishnan, A., Bieler, J., Schneck, J., and Pardoll, D. (1998). Potent T Cell Activation with Dimeric Peptide-Major Histocompatibility Complex Class II Ligand: The Role of CD4 Coreceptor. pp. 1633-1640.

Hanna, J., Markoulaki, S., Schorderet, P., Carey, B.W., Beard, C., Wernig, M., Creyghton, M.P., Steine, E.J., Cassady, J.P., Foreman, R., *et al.* (2008). Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. Cell *133*, 250-264.

Hanna, J., Saha, K., Pando, B., van Zon, J., Lengner, C.J., Creyghton, M.P., van Oudenaarden, A., and Jaenisch, R. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. Nature *advance online publication*.

Hattne, J., Fange, D., and Elf, J. (2005). Stochastic reaction-diffusion simulation with MesoRD. Bioinformatics (Oxford, England) *21*, 2923-2924.

Henikoff, S. (2008). Nucleosome destabilization in the epigenetic regulation of gene expression. Nature Reviews Genetics *9*, 15-26.

Henrickson, S.E., Mempel, T.R., Mazo, I.B., Liu, B., Artyomov, M.N., Zheng, H., Peixoto, A., Flynn, M.P., Senman, B., Junt, T., *et al.* (2008). T cell sensing of antigen dose governs interactive behavior with dendritic cells and sets a threshold for T cell activation. Nature Immunology *9*, 282-291.

Hlavacek, W.S., and Faeder, J.R. (2009). The complexity of cell signaling and the need for a new mechanics. Sci Signal *2*, pe46.

Holler, P.D., and Kranz, D.M. (2003). Quantitative analysis of the contribution of TCR/pepMHC affinity and CD8 to T cell activation. Immunity *18*, 255-264.

Hong, H., Takahashi, K., Ichisaka, T., Aoi, T., Kanagawa, O., Nakagawa, M., Okita, K., and Yamanaka, S. (2009). Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. Nature *460*, 1132-U1195.

Huangfu, D.W., Maehr, R., Guo, W.J., Eijkelenboom, A., Snitow, M., Chen, A.E., and Melton, D.A. (2008). Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. Nature Biotechnology *26*, 795-797.

Huppa, J.B., Davis, M. M. (2009). Nature, in press.

Irvine, D.J., Purbhoo, M.A., Krogsgaard, M., and Davis, M.M. (2002). Direct observation of ligand recognition by T cells. Nature *419*, 845-849.

Jackson, M., Krassowska, A., Gilbert, N., Chevassut, T., Forrester, L., Ansell, J., and Ramsahoye, B. (2004). Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. Molecular and Cellular Biology *24*, 8862-8871.

Jaenisch, R., and Young, R. (2008). Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell *132*, 567-582.

Janeway, C., Murphy, K.P., Travers, P., Walport, M., and Janeway, C. (2008). Janeway's immuno biology (New York: Garland Science).

Johnston, M.D., Edwards, C.M., Bodmer, W.F., Maini, P.K., and Chapman, S.J. (2007). Mathematical modeling of cell population dynamics in the colonic crypt and in colorectal cancer. Proceedings of the National Academy of Sciences of the United States of America *104*, 4008-4013.

Jones, P., and Simons, B.D. (2008). Epidermal homeostasis: do committed progenitors work while stem cells sleep? Nature Reviews Molecular Cell Biology *9*, 82-U89.

Jones, P.H., Simons, B.D., and Watt, F.M. (2007). Sic transit gloria: Farewell to the epidermal transit amplifying cell? Cell Stem Cell *1*, 371-381.

Karmakar, R., and Bose, I. (2007). Positive feedback, stochasticity and genetic competence. Phys Biol *4*, 29-37.

Kawamura, T., Suzuki, J., Wang, Y.V., Menendez, S., Morera, L.B., Raya, A., Wahl, G.M., and Belmonte, J.C.I. (2009). Linking the p53 tumour suppressor pathway to somatic cell reprogramming. Nature *460*, 1140-U1107.

Kepler, T.B., and Elston, T.C. (2001a). Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. Biophysical Journal *81*, 3116-3136.

Kepler, T.B., and Elston, T.C. (2001b). Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. Biophys J *81*, 3116-3136.

Kersh, E.N., Shaw, A.S., and Allen, P.M. (1998). Fidelity of T cell activation through multistep T cell receptor zeta phosphorylation. Science (New York, N.Y *281*, 572-575.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Krogsgaard, M., Juang, J., and Davis, M.M. (2007a). A role for "self" in T-cell activation. Seminars in immunology *19*, 236-244.

Krogsgaard, M., Juang, J., and Davis, M.M. (2007b). A role for "self" in T-cell activation. Seminars in Immunology *19*, 236-244.

Krogsgaard, M., Li, Q.J., Sumen, C., Huppa, J.B., Huse, M., and Davis, M.M. (2005). Agonist/endogenous peptide-MHC heterodimers drive T cell activation and sensitivity. Nature *434*, 238-243.

Lai, K., Robertson, M.J., and Schaffer, D.V. (2004). The Sonic hedgehog signaling system as a bistable genetic switch. Biophysical Journal *86*, 2748-2757.

Levine, J., Kueh, H.Y., and Mirny, L. (2007). Intrinsic fluctuations, robustness, and tunability in signaling cycles. Biophysical Journal *92*, 4473-4481.

Li, Q.J., Chau, J., Ebert, P.J., Sylvester, G., Min, H., Liu, G., Braich, R., Manoharan, M., Soutschek, J., Skare, P., *et al.* (2007). miR-181a is an intrinsic modulator of T cell sensitivity and selection. Cell *129*, 147-161.

Li, Q.J., Dinner, A.R., Qi, S., Irvine, D.J., Huppa, J.B., Davis, M.M., and Chakraborty, A.K. (2004a). CD4 enhances T cell sensitivity to antigen by coordinating Lck accumulation at the immunological synapse. Nat Immunol *5*, 791-799.

Li, Q.J., Dinner, A.R., Qi, S.Y., Irvine, D.J., Huppa, J.B., Davis, M.M., and Chakraborty, A.K. (2004b). CD4 enhances T cell sensitivity to antigen by coordinating Lck accumulation at the immunological synapse. Nature Immunology *5*, 791-799.

Lin, J., and Weiss, A. (2001). T cell receptor signalling. J Cell Sci *114*, 243-244.

Lipshtat, A., Loinger, A., Balaban, N.Q., and Biham, O. (2006). Genetic toggle switch without cooperative binding. Physical review letters *96*, -.

Lis, M., Artyomov, M.N., Devadas, S., and Chakraborty, A.K. (2009). Efficient stochastic simulation of reaction-diffusion processes via direct compilation. Bioinformatics (Oxford, England) *25*, 2289-2291.

Lo, W.L., Felix, N.J., Walters, J.J., Rohrs, H., Gross, M.L., and Allen, P.M. (2009). An endogenous peptide positively selects and augments the activation and survival of peripheral CD4+ T cells. Nature immunology *10*, 1155-1161.

Loeffler, M., and Roeder, I. (2002). Tissue stem cells: definition, plasticity, heterogeneity, self-organization and models--a conceptual approach. Cells, tissues, organs *171*, 8-26.

Loinger, A., Lipshtat, A., Balaban, N.Q., and Biham, O. (2007). Stochastic simulations of genetic switch systems. Physical Review E *75*, -.

Luescher, I.F., Vivier, E., Layer, A., Mahiou, J., Godeau, F., Malissen, B., and Romero, P. (1995). CD8 modulation of T-cell antigen receptor-ligand interactions on living cytotoxic T lymphocytes. Nature *373*, 353-356.

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., *et al.* (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell *1*, 55-70.

Mallaun, M., Naeher, D., Daniels, M.A., Yachi, P.P., Hausmann, B., Luescher, I.F., Gascoigne, N.R.J., and Palmer, E. (2008). The T cell receptor's alpha-chain connecting peptide motif promotes close approximation of the CD8 coreceptor allowing efficient signal initiation. J Immunol *180*, 8211-8221.

Marion, R.M., Strati, K., Li, H., Murga, M., Blanco, R., Ortega, S., Fernandez-Capetillo, O., Serrano, M., and Blasco, M.A. (2009). A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. Nature *460*, 1149-U1119.

Markevich, N.I., Hoek, J.B., and Kholodenko, B.N. (2004). Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. Journal of Cell Biology *164*, 353-359.

Matthew, K.V., and Brian, K.H. (2006). Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest. pp. 425-455.

McAdams, H.H., and Arkin, A. (1997). Stochastic mechanisms in gene expression. Proc Natl Acad Sci U S A *94*, 814-819.

McNeill, L., Salmond, R.J., Cooper, J.C., Carret, C.K., Cassady-Cain, R.L., Roche-Molina, M., Tandon, P., Holmes, N., and Alexander, D.R. (2007). The differential regulation of Lck kinase phosphorylation sites by CD45 is critical for T cell receptor signaling responses. Immunity *27*, 425-437.

McQuarrie, D.A. (1967). Stochastic Approach to chemical kinetics. Journal of Applied Probability *4*, 413-478.

Meissner, A., Mikkelsen, T.S., Gu, H.C., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X.L., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., *et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-U791.

Meissner, A., Wernig, M., and Jaenisch, R. (2007). Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. Nature Biotechnology *25*, 1177-1181.

Mikkelsen, T.S., Hanna, J., Zhang, X.L., Ku, M.C., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49-U41.

Mikkelsen, T.S., Ku, M.C., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-U552.

Miller, C.A., and Beard, D.A. (2008). The effects of reversibility and noise on stochastic phosphorylation cycles and cascades. Biophysical Journal *95*, 2183-2192.

Morozov, A.Y., Pronina, E., Kolomeisky, A.B., and Artyomov, M.N. (2007). Solutions of burnt-bridge models for molecular motor transport. Physical review *75*, 031910.

Nishikawa, S., Goldstein, R.A., and Nierras, C.R. (2008). The promise of human induced pluripotent stem cells for research and therapy. Nature Reviews Molecular Cell Biology *9*, 725-729.

Nishimura, S., Takahashi, S., Kuroha, T., Suwabe, N., Nagasawa, T., Trainor, C., and Yamamoto, M. (2000). A GATA box in the GATA-1 gene hematopoietic enhancer is a critical element in the network of GATA factors and sites that regulate this gene. Molecular and Cellular Biology *20*, 713-723.

Nutt, S.L. (2008). B-cell identity - Commitment is not forever. N. Engl. J. Med. *358*, 82-83.

Okuno, Y., Huang, G., Rosenbauer, F., Evans, E.K., Radomska, H.S., Iwasaki, H., Akashi, K., Moreau-Gachelin, F., Li, Y.L., Zhang, P., *et al.* (2005). Potential autoregulation of transcription factor PU.1 by an upstream reaulatory element. Molecular and Cellular Biology *25*, 2832-2845.

Orford, K., Kharchenko, P., Lai, W., Dao, M.C., Worhunsky, D.J., Ferro, A., Janzen, V., Park, P.J., and Scadden, D.T. (2008). Differential H3K4 methylation identifies developmentally poised hematopoietic genes. Developmental Cell *14*, 798-809.

Orford, K.W., and Scadden, D.T. (2008). Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation. Nature Reviews Genetics *9*, 115-128.

Ozbudak, E.M., Thattai, M., Lim, H.N., Shraiman, B.I., and Van Oudenaarden, A. (2004). Multistability in the lactose utilization network of Escherichia coli. Nature *427*, 737-740.

Palmer, E., and Naeher, D. (2009). Affinity threshold for thymic selection through a T-cell receptor-co-receptor zipper. Nat Rev Immunol *9*, 207-213.

Pan, G.J., Tian, S.L., Nie, J., Yang, C.H., Ruotti, V., Wei, H.R., Jonsdottir, G.A., Stewart, R., and Thomson, J.A. (2007). Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell Stem Cell *1*, 299-312.

Park, I.H., Zhao, R., West, J.A., Yabuuchi, A., Huo, H.G., Ince, T.A., Lerou, P.H., Lensch, M.W., and Daley, G.Q. (2008). Reprogramming of human somatic cells to pluripotency with defined factors. Nature *451*, 141-U141.

Purbhoo, M.A., Irvine, D.J., Huppa, J.B., and Davis, M.M. (2004). T cell killing does not require the formation of a stable mature immunological synapse. Nat Immunol *5*, 524-530.

Qu, K., and Ortoleva, P. (2008). Understanding stem cell differentiation through self-organization theory. Journal of Theoretical Biology *250*, 606-622.

Rekhtman, N., Radparvar, F., Evans, T., and Skoultchi, A.I. (1999). Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. Genes Dev. *13*, 1398-1411.

Reyes, B.M.R., Danese, S., Sans, M., Fiocchi, C., and Levine, A.D. (2005). Redox Equilibrium in Mucosal T Cells Tunes the Intestinal TCR Signaling Threshold. pp. 2158-2166.

Rice, S.H. (2004). Evolutionary theory : mathematical and conceptual foundations (Sunderland, Mass., USA: Sinauer Associates).

Roeder, I., and Glauche, I. (2006). Towards an understanding of lineage specification in hematopoietic stem cells: A mathematical model for the interaction of transcription factors GATA-1 and PU.1. J. Theor. Biol. *241*, 852-865.

Roose, J.P., Mollenauer, M., Ho, M., Kurosaki, T., and Weiss, A. (2007). Unusual interplay of two types of Ras activators, RasGRP and SOS, establishes sensitive and robust Ras activation in lymphocytes. Mol Cell Biol *27*, 2732-2745.

Rosette, C., Werlen, G., Daniels, M.A., Holman, P.O., Alam, S.M., Travers, P.J., Gascoigne, N.R., Palmer, E., and Jameson, S.C. (2001). The impact of duration versus extent of TCR occupancy on T cell activation: a revision of the kinetic proofreading model. Immunity *15*, 59-70.

Samoilov, M., Plyasunov, S., and Arkin, A.P. (2005a). Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations.

Proceedings of the National Academy of Sciences of the United States of America *102*, 2310-2315.

Samoilov, M., Plyasunov, S., and Arkin, A.P. (2005b). Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations. Proc Natl Acad Sci U S A *102*, 2310-2315.

Samoilov, M.S., and Arkin, A.P. (2006). Deviant effects in molecular reaction pathways. Nature Biotechnology *24*, 1235-1240.

Sasai, M., and Wolynes, P.G. (2003). Stochastic gene expression as a many-body problem. Proceedings of the National Academy of Sciences of the United States of America *100*, 2374-2379.

Simonsson, S., and Gurdon, J. (2004). DNA demethylation is necessary for the epigenetic reprogramming of somatic cell nuclei. Nature Cell Biology *6*, 984-990.

Sridharan, R., and Plath, K. (2008). Illuminating the black box of reprogramming. Cell Stem Cell *2*, 295-297.

Stadtfeld, M., Maherali, N., Breault, D.T., and Hochedlinger, K. (2008). Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. Cell Stem Cell *2*, 230-240.

Starr, T.K., Jameson, S.C., and Hogquist, K.A. (2003). Positive and negative selection of T cells. Annu Rev Immunol *21*, 139-176.

Stefanova, I., Hemmer, B., Vergelli, M., Martin, R., Biddison, W.E., and Germain, R.N. (2003). TCR ligand discrimination is enforced by competing ERK positive and SHP-1 negative feedback pathways. Nat Immunol *4*, 248-254.

Stone, J.D., Chervin, A.S., and Kranz, D.M. (2009). T-cell receptor binding affinities and kinetics: impact on T-cell activity and specificity. Immunology *126*, 165-176.

Surh, C.D., and Sprent, J. (2008). Homeostasis of Naive and Memory T Cells. Immunity *29*, 848-862.

Sykulev, Y., Joo, M., Vturina, I., Tsomides, T.J., and Eisen, H.N. (1996). Evidence that a single peptide-MHC complex on a target cell can elicit a cytolytic T cell response. Immunity *4*, 565-571.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861-872.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.

Utikal, J., Polo, J.M., Stadtfeld, M., Maherali, N., Kulalert, W., Walsh, R.M., Khalil, A., Rheinwald, J.G., and Hochedlinger, K. (2009). Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. Nature *460*, 1145-U1112.

van Bergen, J., Kooy, Y., and Koning, F. (2001). CD4-independent T cells impair TCR triggering of CD4-dependent T cells: a putative mechanism for T cell affinity maturation. European Journal of Immunology *31*, 646-652.

Weinberger, L.S., Burnett, J.C., Toettcher, J.E., Arkin, A.P., and Schaffer, D.V. (2005). Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. Cell *122*, 169-182.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M.C., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature *448*, 318-U312.

Wilson, M., and Koopman, P. (2002). Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. Current Opinion in Genetics & Development *12*, 441-446.

Winkler, D., Halley, J., and Burden, F. (2007). Modelling stem cell properties and fate decisions. Drugs of the Future *32*, 93-93.

Wooldridge, L., van den Berg, H.A., Glick, M., Gostick, E., Laugel, B., Hutchinson, S.L., Milicic, A., Brenchley, J.M., Douek, D.C., Price, D.A., and Sewell, A.K. (2005). Interaction between the CD8 coreceptor and major histocompatibility complex class I stabilizes T cell receptor-antigen complexes at the cell surface. Journal of Biological Chemistry *280*, 27491-27501.

Wyer, J.R., Willcox, B.E., Gao, G.F., Gerth, U.C., Davis, S.J., Bell, J.I., van der Merwe, P.A., and Jakobsen, B.K. (1999). T cell receptor and coreceptor CD8 alphaalpha bind peptide-MHC independently and with distinct kinetics. Immunity *10*, 219-225.

Wylie, D.C., Das, J., and Chakraborty, A.K. (2007a). Sensitivity of T cells to antigen and antagonism emerges from differential regulation of the same molecular signaling module. Proceedings of the National Academy of Sciences of the United States of America *104*, 5533-5538.

Wylie, D.C., Das, J., and Chakraborty, A.K. (2007b). Sensitivity of T cells to antigen and antagonism emerges from differential regulation of the same molecular signaling module. Proceedings of the National Academy of Sciences of the United States of America *104*, 5533-5538.

Xie, H.F., Ye, M., Feng, R., and Graf, T. (2004). Stepwise reprogramming of B cells into macrophages. Cell *117*, 663-676.

Xiong, W., and Ferrell, J.E., Jr. (2003). A positive-feedback-based bistable 'memory module' that governs a cell fate decision. Nature *426*, 460-465.

Xiong, Y., Kern, P., Chang, H., and Reinherz, E. (2001). T Cell Receptor Binding to a pMHCII Ligand Is Kinetically Distinct from and Independent of CD4. The Journal of biological chemistry *276*, 5659-5667.

Yachi, P.P., Ampudia, J., Gascoigne, N.R., and Zal, T. (2005a). Nonstimulatory peptides contribute to antigen-induced CD8-T cell receptor interaction at the immunological synapse. Nature immunology *6*, 785-792.

Yachi, P.P., Ampudia, J., Gascoigne, N.R.J., and Zal, T. (2005b). Nonstimulatory peptides contribute to antigen-induced CD8-T cell receptor interaction at the immunological synapse. Nature Immunology *6*, 785-792.

Yachi, P.P., Ampudia, J., Zal, T., and Gascoigne, N.R.J. (2006). Altered Peptide Ligands Induce Delayed CD8-T Cell Receptor Interaction--a Role for CD8 in Distinguishing Antigen Quality. Immunity *25*, 203-211.

Yachi, P.P., Lotz, C., Ampudia, J., and Gascoigne, N.R. (2007). T cell activation enhancement by endogenous pMHC acts for both weak and strong agonists but varies with differentiation state. The Journal of experimental medicine *204*, 2747-2757.

Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519-U515.

Yu, J.Y., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., *et al.* (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917-1920.

Zamoyska, R. (2007). Why is there so much CD45 on T cells? Immunity *27*, 421-423.

Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J., and Melton, D.A. (2008). Cellular alchemy: direct reprogramming of pancreatic cells into insulin-producing cells. Regenerative Medicine *3*, 789-789.
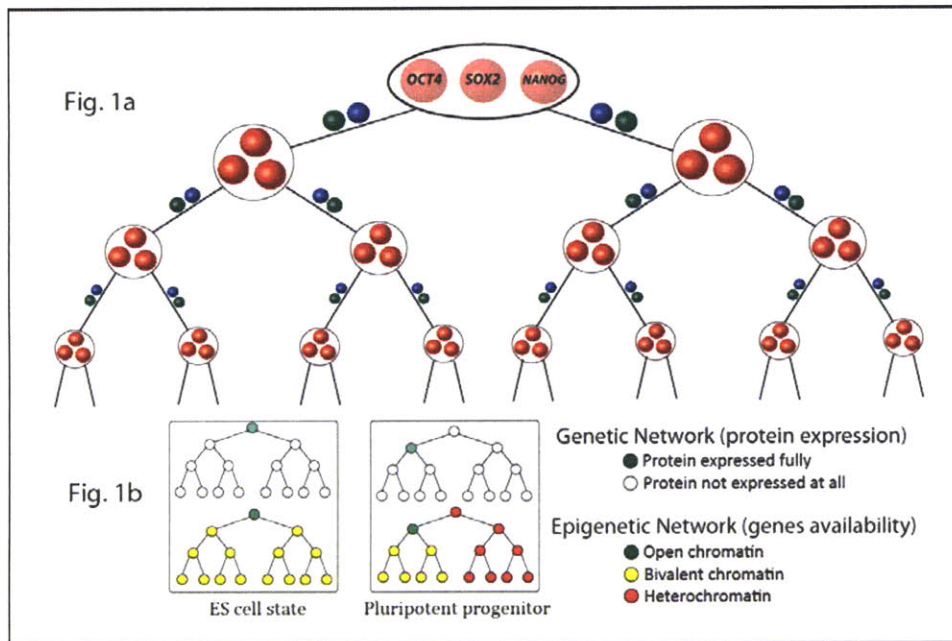
7.8 Figures for Chapter 7

Fig. 7.1 Specification of the genetic and epigenetic states that describe cellular states

(a)   Only the master-regulatory genes that govern cell state are arranged in a hierarchy (house keeping, stress-response and many other genes are not considered). Each node of the hierarchy represents an ensemble of master-regulatory genes that govern a particular cellular state. For example, genes in the top node are known master-regulators of the embryonic stem cell state (e.g. Oct4, Sox2, Nanog). When a cell is in the ES state, only these three genes will be expressed while other genes will not. Similarly, when a cell is fully differentiated, genes in one of the bottom modules will be expressed but not any other gene in the network. Each master-regulatory ensemble can contain many genes, only three are shown in each node.

(b)   Fig. 1a has been coarse-grained such that only master-regulatory modules (nodes in fig. 1a) are shown. Cellular identity is determined by both epigenetic (chromatin marks, DNA methylation) and genetic (expression profile) states. Examples of two states (ES state and "left" pluripotent progenitor) are shown. For each example, two lattices are needed to describe the state of gene expression and the epigenome: top lattice reflects the expression levels of master-regulatory proteins in the ES/progenitor state and bottom lattice reflects the epigenetic state of master-regulatory genes in the ES/progenitor state.
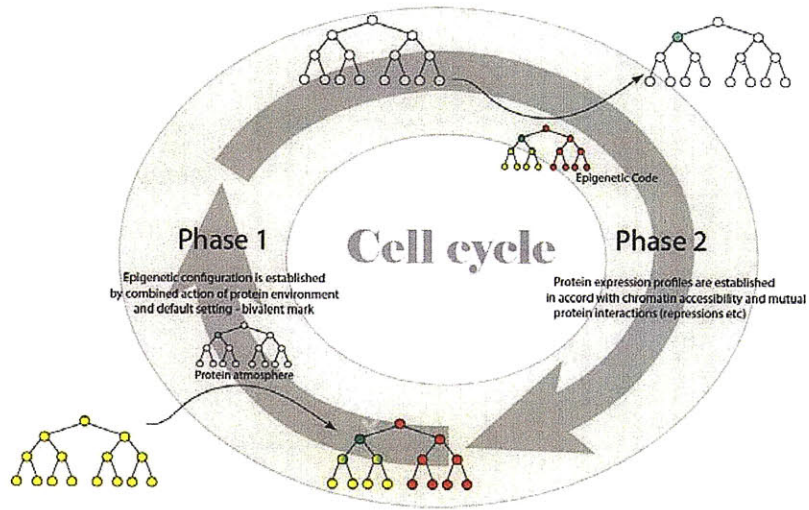
Fig. 7.2 Simplified model for progression through the cell cycle. The cell cycle is divided into two generalized phases: called interphase and telophase for simplicity. Gene expression occurs during the interphase, while cell division and associated processes occur in the telophase. In the interphase gene expression profile is governed by the stable epigenetic marks on the master-regulatory genes. In the telophase, however, protein environment can change the epigenetic marks of the master-regulaory genes, particularly when DNA is decondensing after cell division. Differentiation signals (newly expressed proteins) determine future epigenetic marks created during telophase due to the action of the new protein environment. The color code representing genetic and epigenetic states is the same as in Fig.1.
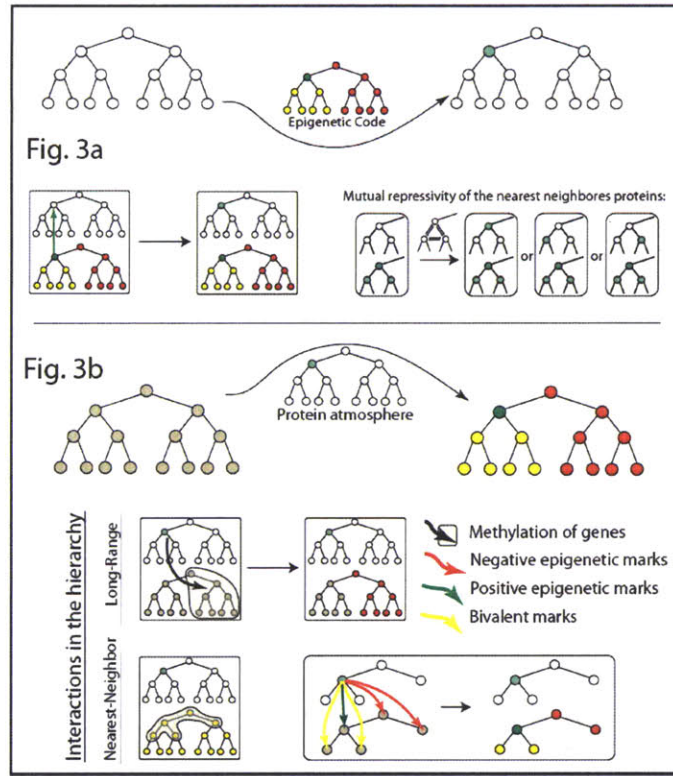
Fig. 3a

Fig. 3b

Fig. 7.3 (a) During interphase, gene expression profiles of master-regulatory modules are established. Gene expression is influenced by epigenetic marking of the corresponding gene and interactions between expressed proteins. Two rules reflect this in our simulation: 1) when master-regulatory gene is in epigenetically marked positively, it favors expression of the corresponding protein; 2) when two (three) neighboring genes are in epigenetically open states, they all favor expression of corresponding proteins, but due to their mutually repressive action (see text) only one of two(three) genes are expressed. Which gene is expressed is chosen stochastically. The color code representing genetic and epigenetic states is the same as in Fig.1

(b) During the telophase, the protein environment can alter the epigenetic marks on the master-regulatory genes. Epigenetic marks on both neighboring and distant genes in the hierarchy can be altered. Long-range effect is typically mediated through DNA methylation which epigenetically silences all of the master-regulatory genes of unrelated lineages and also ancestral states (see text). Short-range interactions affect nearest-neighbors differentially: progenies master-regulatory genes are preferentially put into bivalent states while progenitor and competing lineage modules are epigenetically silenced. The color code representing genetic and epigenetic states is the same as in Fig.1
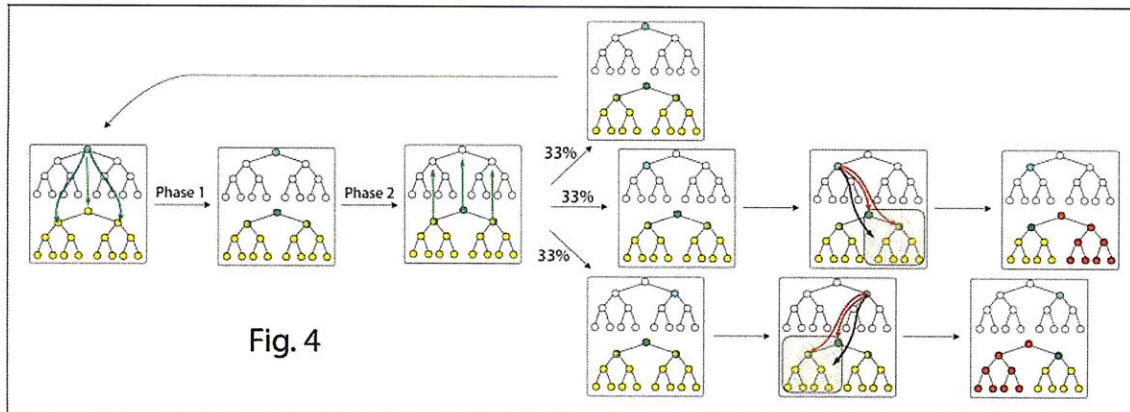
Fig. 4

Fig. 7.4: Changing cellular identity during self-initiated differentiation of the ES cell-state is shown in detail. Process begins with cell division where regulatory modules of progenies are put into epigenetically open states. In phase 2 only one of the three neighboring proteins can be actually expressed in accord with Fig. 3a. Thus, one of three possibilities is realized: self-renewal, and differentiation to the "left" or "right" lineages. In the absence of external stimuli, in our simulations, there is an equal chance to observe each outcome. Simulations are performed with parameter values F=2000; J=3000; G=25; H=40; a=0; b=0.3. The color code representing genetic and epigenetic states is the same as in Fig.1.

Fig. 5a

Dead/arrested state

De-differentiated state

Transdifferentiated state

No harm done

Arrested state
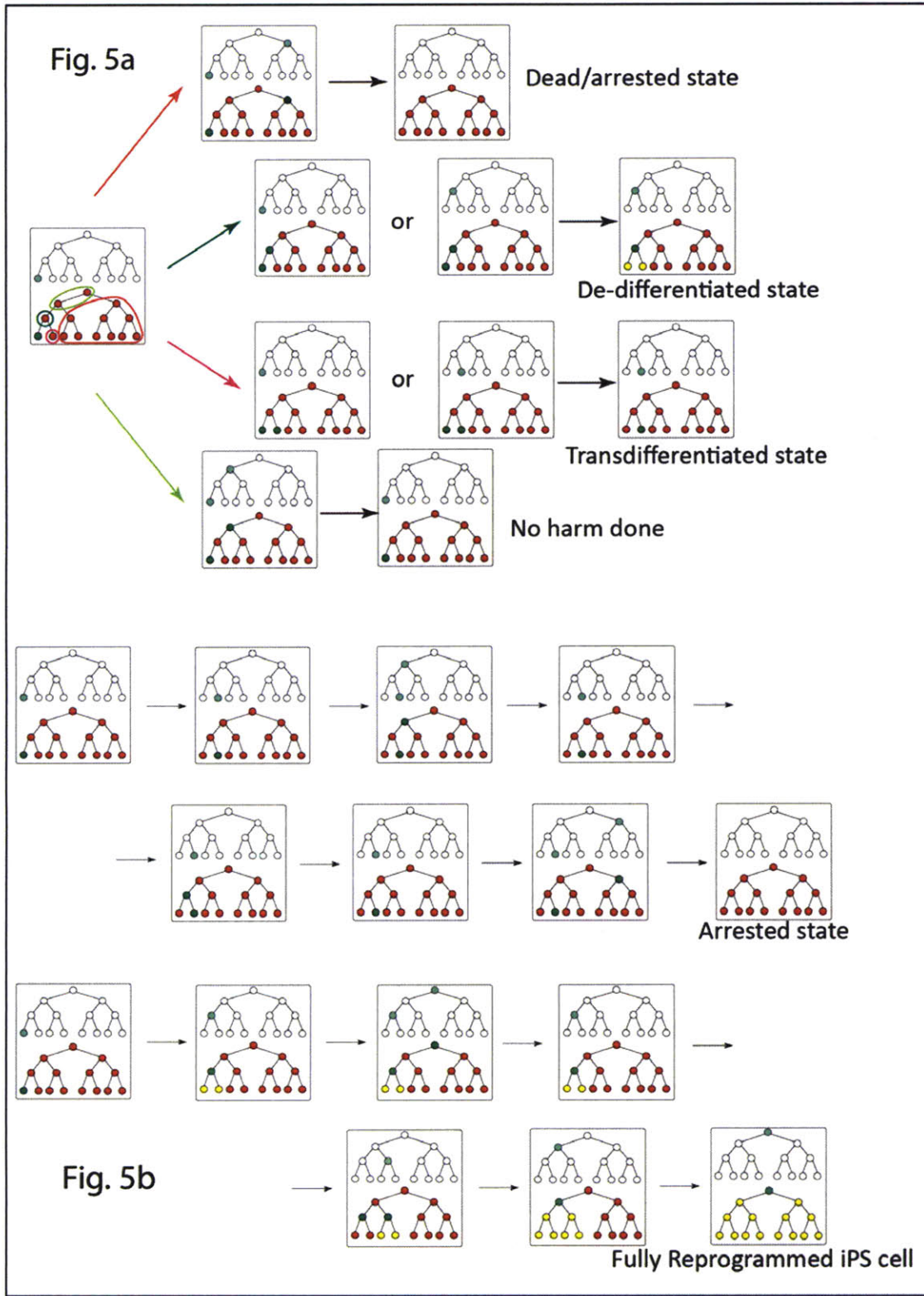
Fig. 5b

Fully Reprogrammed iPS cell

Fig. 7.5: Reprogramming is a consequence of random perturbation of epigenetic state of the cell. In our model, reprogramming factors can change the epigenetic state of randomly chosen regulatory modules (for reasons, see text).

(a) Starting from a fully differentiated state, reprogramming factors can perturb any of the remaining 14 positions (for the case of a 4-level hierarchy). Four outcomes are possible depending on the perturbation site: death/arrest, trans-differentiation, de-differentiation or return to the initial cellular state. These outcomes are determined by simulating the system in accord with the rules described in the text and Figs. 2 - 3. The color code representing genetic and epigenetic states is the same as in Fig.1

(b) Examples of real trajectories observed in simulations illustrating different temporal evolution of epigenetic and genetic states. Complete cell reprogramming appears as a consequence of several successful de-differentiation events as seen in the second example trajectory. Simulations are performed with parameter values F=2000; J=3000; G=25; H=40; a=0; b=0.3. The color code representing genetic and epigenetic states is the same as in Fig.1
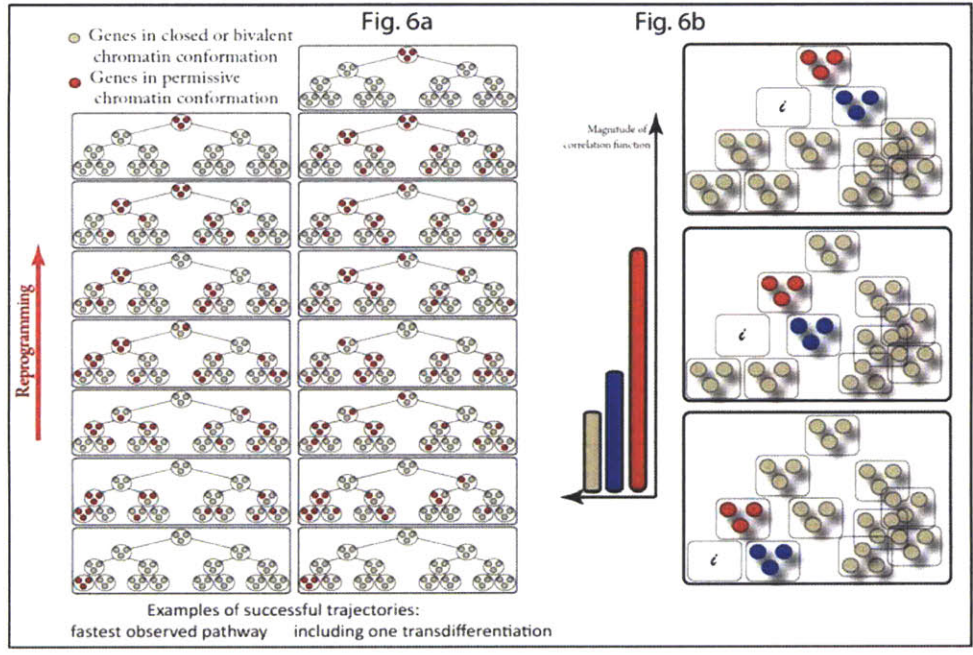
Fig.7.6 (a) Simulations of a model where each gene module regulating a cellular identity consists of three different genes. In this (similar to the previous) model, individual genes do not interact with each other. Rather modules interact with each other when all of the proteins in a module are expressed. Since reprogramming factors change the epigenetic state of randomly chosen individual genes, several (here: at least three) genes have to be changed to open chromatin status at the same time in order to allow a whole module to be able to express proteins. Examples of simulated trajectories show activation of genes of unrelated lineages during successful reprogramming. Simulations are performed with parameter values F=2000; J=3000; G=25; H=40; a=0; b=0.3.

(b) If population averaged expressions of genes during reprogramming can be measured, one can compute a 4-point correlation function (see Eq. 1). This correlation function describes the probability of activation of a given gene after the master regulatory gene module, *i*, was silenced. Then all the genes can be grouped in three groups as our simulation indicates. Thus, the genes defining the most likely paths to reprogramming can be identified as the ones with the highest magnitude of this correlation function. The correlation function was computed by averaging over all successfully reprogrammed trajectories. The colors correspond to the magnitude of the correlation function (as shown on the left)
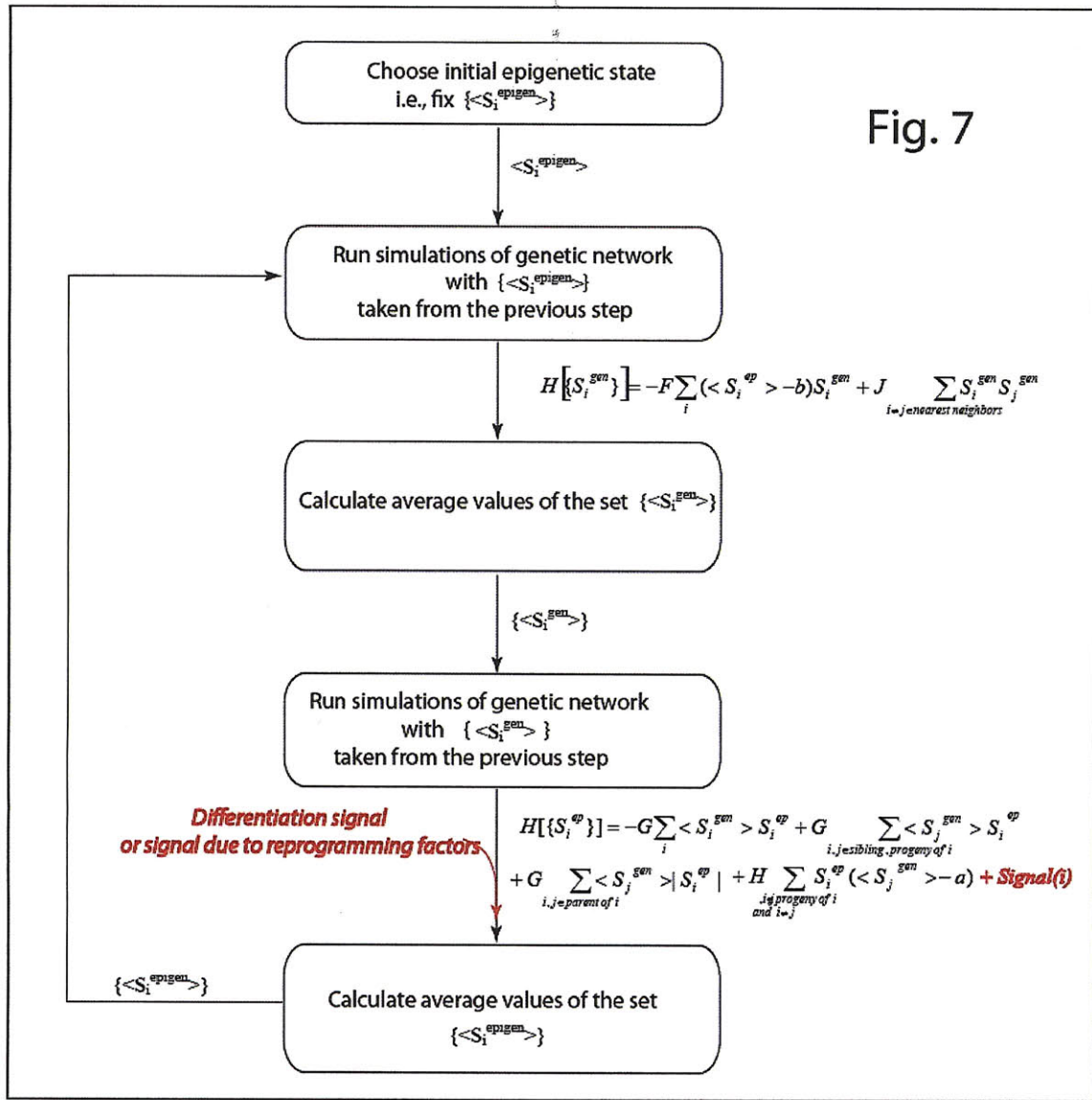
Fig. 7

$$H\left[\{S_i^{gen}\}\right] = -F\sum_i (<S_i^{ep}> - b)S_i^{gen} + J\sum_{i,j \, nearest \, neighbors} S_i^{gen}S_j^{gen}$$

*Differentiation signal
or signal due to reprogramming factors*

$$H[\{S_i^{ep}\}] = -G\sum_i <S_i^{gen}>S_i^{ep} + G\sum_{i,j \, sibling \, .progeny \, of \, i} <S_j^{gen}>S_i^{ep}$$
$$+G\sum_{i,j \, parent \, of \, i} <S_j^{gen}>|S_i^{ep}| + H\sum_{\substack{i,j \, progeny \, of \, i \\ and \, i \neq j}} S_i^{ep}(<S_j^{gen}> - a) \, + Signal(i)$$

Fig.7.7: Flow chart of the simulation procedure. The simulation essentially mimics progression through the cell cycle in accord with Fig. 2. In each phase of the cell cycle, interactions within and between genetic and epigenetic lattices are enforced through the Hamiltonians of Eq. 2 and 3. Mathematical structure and choice of parameters are such that rules depicted in Fig.3 are obeyed. For analysis of sensitivity to parameter variations see supplementary information.
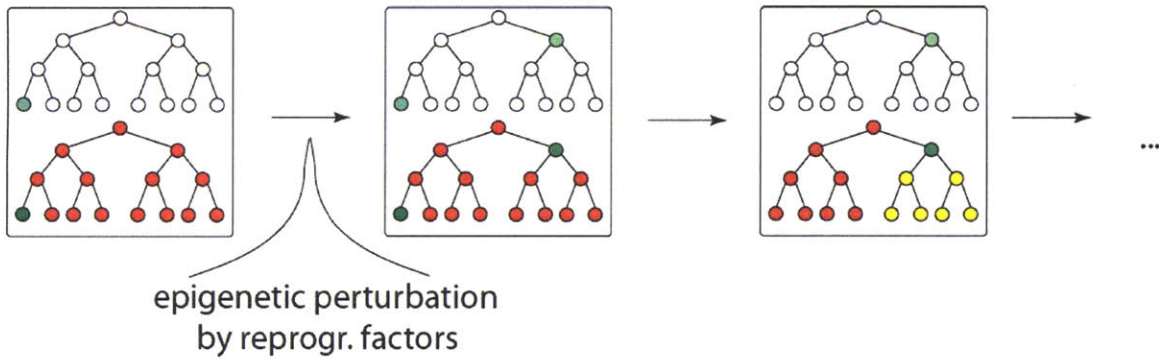
Supplementary figures

epigenetic perturbation
by reprogr. factors

Fig. S7.1a: Example of a part of a simulation trajectory for H=G illustrating large and fast cross-lineage jumps from a terminally differentiated to an almost reprogrammed state. As noted in section 2.a, such jumps are dominant when H~G. (This simulation was carried out with F=2000; J=3000; G=25; H=25; a=0; b=0.3). The color code representing genetic and epigenetic states is the same as in Fig. 7.1.

epigenetic perturbation   action of methylation
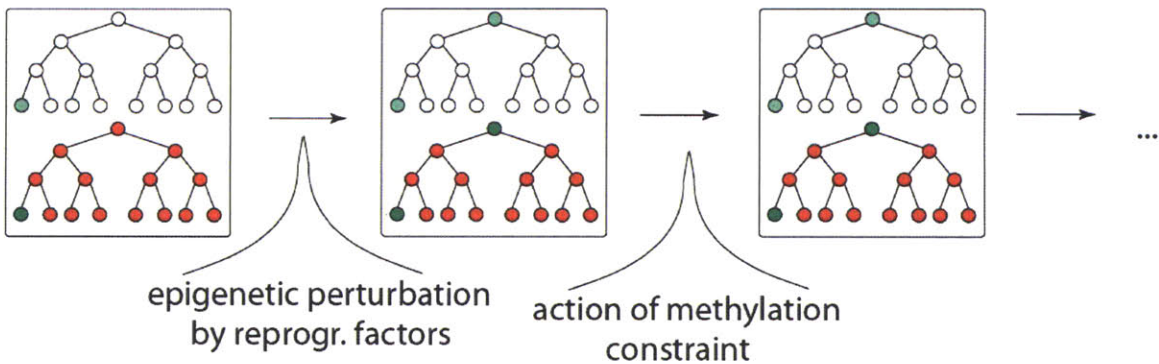by reprogr. factors   constraint

Fig. S7.1b: Example of a part of a simulation trajectory for G>H (in this case G>3H). The gene module regulating the ES state is turned on stably during the first epigenetic perturbation. This simulation was carried out with F=2000; J=3000; G=25; H=8; a=0; b=0.3. The color code representing genetic and epigenetic states is the same as in Fig. 7.1
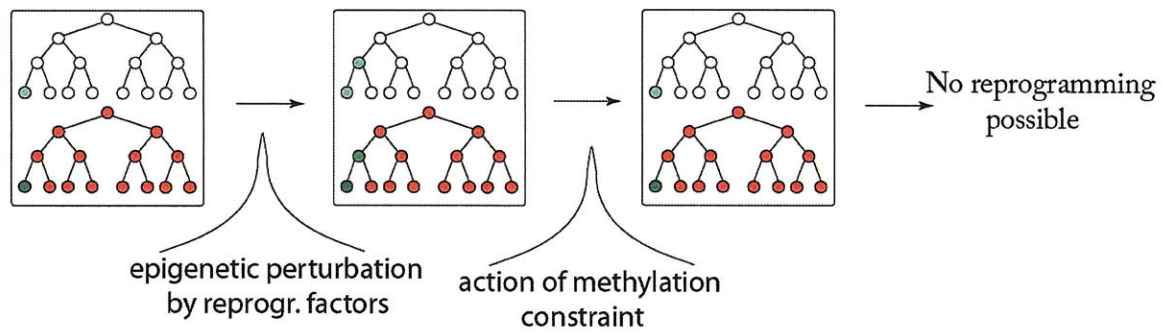
Fig. S7.2: Example of a part of a simulation trajectory with small values of F and J. The simulation was carried out with F=200, J=300, G=25; H=40; a=0; b=0.3. The color code representing genetic and epigenetic states is the same as in Fig. 7.1.
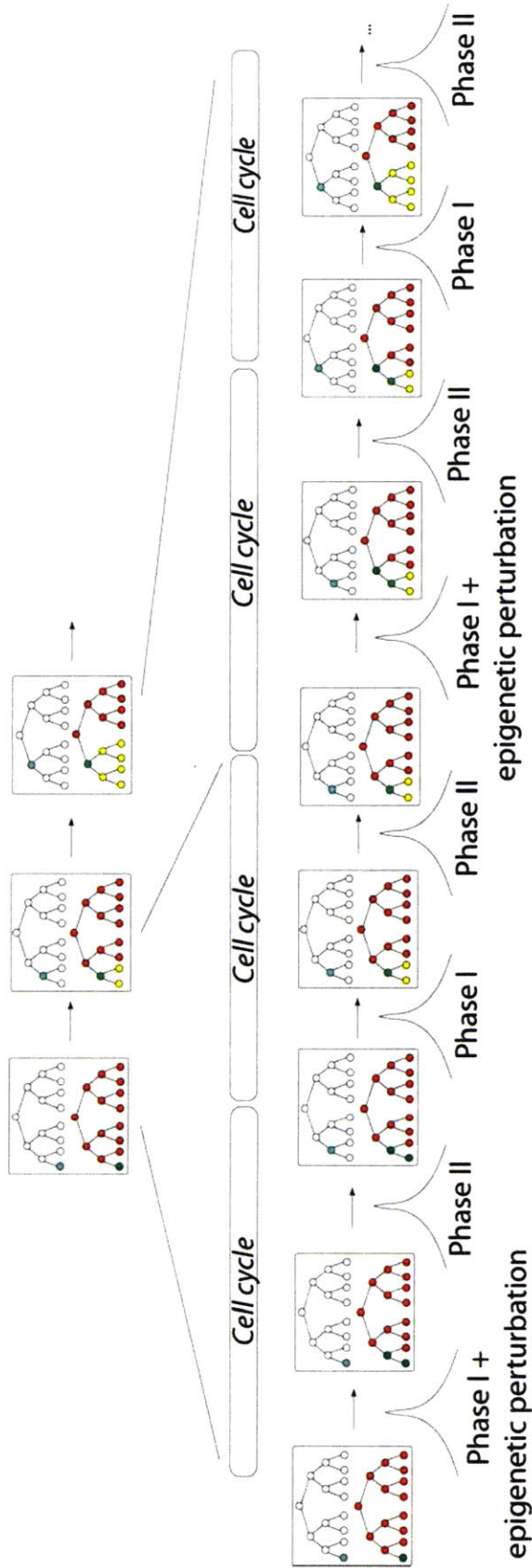
Fig. S7.3a: Example of a typical trajectory obtained from our simulations showing that it takes several cell cycles (in our simulations, 2 cycles) to achieve one level of reprogramming in a stable manner. This simulation was performed with parameter values: F=2000; J=3000; G=25; H=40; *a*=0; *b*=0.3. The color code representing genetic and epigenetic states is the same as in Fig.7.1.
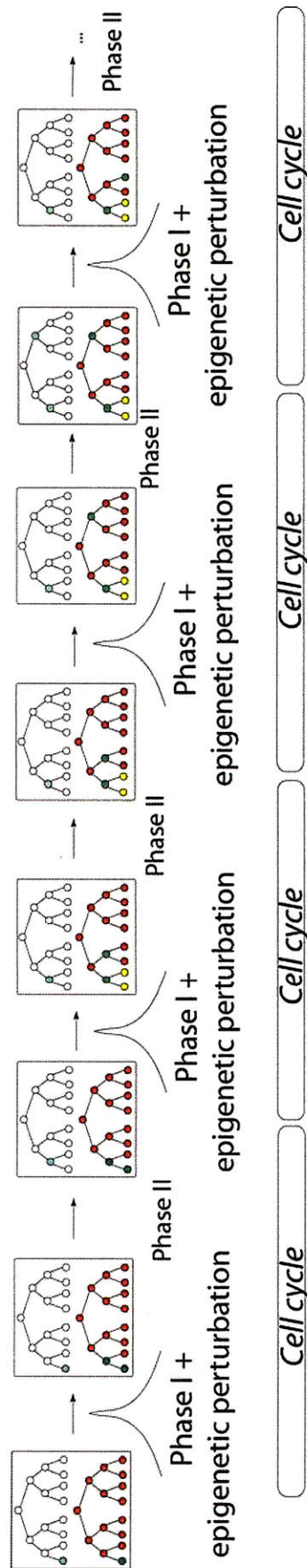
**Fig. S7.3b**: Example of a typical trajectory from our simulations where, because reprogramming factors are overexpressed, they can act at every cycle, as opposed to the every two cycles (as in panel (a)). This renders reprogramming impossible for reasons described in the main text. This simulation was performed with parameter values: F=2000; J=3000; G=25; H=40; a=0; b=0.3.
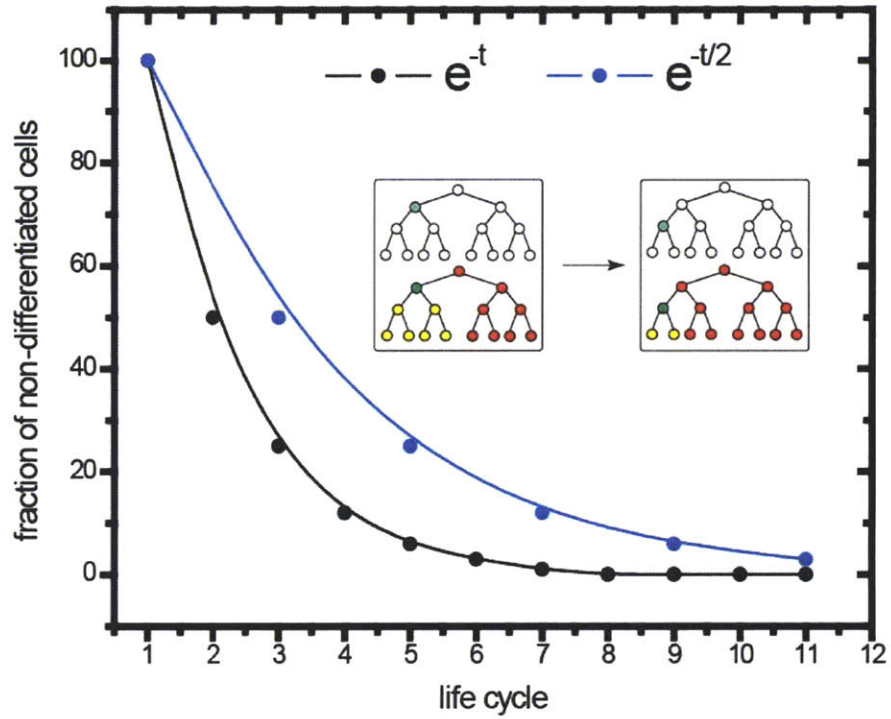
Fig.S7.4: Dynamics of cell differentiation upon receiving cues of different strength. Our simulations show that the progenitor cells differentiate in accord with first order kinetics, with the lifetime of progenitor cells depending on the signal strength. The blue curve describes the behavior of a cell population which received a signal that is twice as weak as the population represents by the black line. Simulations are performed with parameter values F=2000; J=3000; G=25; H=40; a=0; b=0.3.

Fig.S7.5: The diagram depicting transitions between stable cell populations under the action of reprogramming factors. The starting point is terminally differentiated state depicted on the bottom of the picture. Possible transitions from this state are indicated by arrows. Probability to experience a particular transition is indicated by the % value on top of the corresponding arrow.

Chapter 8

Concluding remarks

In conclusion, I would like to reiterate the premise of this thesis that was stated in Chapter 1. Biology offers a broad spectrum of problems ranging from fundamental studies of chemical networks and application of stochastic chemical networks to signal processing all the way to the application of coarse-grained statistical mechanical model for phenomenological description of cellular behavior. Some of the problems have been explored in this thesis.

Innumerable amount of problems was not covered in this thesis and it is my hope as a future professional scientist to explore connections between post-translational modifications and changes in transcriptional program of the cell in more details. Novel chip-based experimental technics allow one to trace expression of thousands of genes at a time. The wealth of connections and interdependences in the transcriptional networks create enormous problem when trying to recover underlying network structure based on the expression data alone. Additional level of complexity is put in due to non-linearity of "interactions" between different genes since these interactions are mediated by post-translational networks such as the ones studied in this thesis.

Unified view of transcriptional and post-translational networks, even on the crude coarse-grained level could provide a new paradigm for looking at the biological network. The need for such a novel paradigm is nicely referred to as "Need for a New Mechanics" by Bill Hlavacek and Jim Faeder (Hlavacek and Faeder, 2009)[*]

---

[*]Hlavacek, W.S., and Faeder, J.R. (2009). The complexity of cell signaling and the need for a new mechanics. Sci Signal 2, pe46.

Curriculum Vitae

Maksym Artomov

please note professional name spelling: **Maxim N. Artyomov**

Massachusetts Institute of Technology
Room E19-538
77 Massachusetts Ave
Cambridge, MA 02139   USA
Work phone:1-617-252-1744
e-mail: maxart@mit.edu

born May 30, 1982

PhD Thesis Title: **Stochastic Processes in Biological Systems: Selected Problems**

Education:

- 2009 (Dec.1) …   post-doctoral researcher at Broad Institute
- 2005-2009   graduate student at Massachusetts Institute of Technology, Chemistry Department.
- 2004-2005   University of Chicago, **M.Sc.**
- 1999-2004   Moscow State University, **B.Sc.** *with honorary Gold Medal*

Awards:

- 2006   Teaching Award: for excellence in teaching (as teaching assistant in undergraduate Thermodynamics and Kinetics class)
- 1999   *Silver Medal* in International Chemistry Olympiad among high-school students (Bangkok, Thailand)

Teaching and Work Experience:

- Summer 2009   – Internship in Vertex Pharmaceuticals (Cambridge, MA)
  Target identification in signalling pathways affected by IBD
- Fall'07   – TA in the Statistical Mechanics with Biological Applications
  (5.70, graduate course, MIT, Chemistry Dep)
- Spring'06   – TA in the Chemical Thermodynamics and Kinetics course
  (5.60, advanced undergraduate course, MIT, Chemistry Dep)
- Fall'05   – TA in the Chemical Thermodynamics and Kinetics course
  (5.60, advanced undergraduate course, MIT, Chemistry Dep)
- Fall'04/Spr'05   – TA in the General Chemistry course
  (first-year undergraduate course, U.Chicago)

Research interests:

<u>Computational biologys and systems biology</u>: stochastic modeling of signaling pathways, modeling of biophysical measurements (biacore/SPR) in complex systems: heterogeneous binding of multimeric ligands.

<u>Biological Applications of the Statistical Mechanics</u>: molecular motors, statistical models in immunology, stochastic fluctuations in chemical networks.

<u>Non-Equilibrium Statistical Mechanics</u>: Non-Equilibrium Steady States, Stochastic Processes, Transport Phenomena, Stochastic Fluctuations in Chemical Networks.

<u>Equilibrium Statistical Mechanics</u>: Polymers, Systems with longe-range interactions.

Publications:

1. Stone J., **Artyomov M.N.**, Eisen H.N., Chakraborty A.K., Kranz D.
   MHC Tetramer staining and activation assay revisited
   Manuscript in preparation (2009)

2. **Artyomov M.N.**, Lis M., Devadas S., Chakraborty A.K.
   Mechanisms of self-peptide mediated signal enhancement during early T-cell activation
   To be submitted to Journal of Immunology (2009)

3. Mouquet H., Scheid J.F., Zoller M., Ott R., Shukair S., **Artyomov M.N.**, John Pietzsch J., Connors M., Pereyra F., Walker B.D., Ho1D., Wyatt R.T., Mascola J.R., Wilson1 P.C., Seaman1 M.S., Eisen H.N., Chakraborty A.K., Hope T., Ravetch J.V., Wardemann L, Nussenzweig M.C.
   Polyreactivity increases the affinity of anti-HIV antibodies by heteroligation
   Submitted to Nature (2009)

4. **Artyomov M.N.**, Eisen H.N., Chakraborty A.K.
   Method for determining the effect of heteroligation on the stability of multivalent interactions: application to neutralizing antibodies to HIV antigens
   Submitted to Journal of Immunological Methods (2009)

5. **Artyomov M.N.**, Morozov A.Yu., Kolomeisky A.B.
   Dynamics of Reversible Burnt-Bridge Models of Molecular Motors
   Submitted to Condensed Matter Physcis (2009)

6. **Artyomov M.N.**, Lis M., Devadas S., Chakraborty A.K.
   Role of coreceptor in early T-cell signaling events
   Submitted to Immunity (2009)

7. **Artyomov M.N.**, Meissner A., Chakraborty A.K.
   A model for genetic and epigenetic regulatory networks identifies rare pathways for transctiption factor induced pluripotency
   Submitted to Plos Comp Biology (2009)

8. **Artyomov M.N.**, Mathur M.S., Samoilov M., Chakraborty A.K.
   Bimodality in monostable reversible chemical networks
   Journal of Chemical Physics, 131, 195103 (2009)

9. Lis M., **Artyomov M.N.**, Devadas S., Chakraborty A.K.
    Efficient Stochastic Simulation of Reaction-Diffusion Processes via Direct Compilation
    Bioinformatics 25(17), 2289-91(2009)

10. **Artyomov M.N.**
    On the higher order Onsager Reciprocity relations
    Physical Review Letters (C), 102, 149701 (2009)

11. **Artyomov M.N.**, Morozov A.Yu., Kolomeisky A.B.
    Molecular Motors Interacing with their own tracks
    Physical Review E, 77, 040901(R) (2008)

12. Henrickson S.E., Mempel T.R., Mazo I.B., Liu B., **Artyomov M.N.**, Zheng H, Peixoto A, Flynn M.P., Senman B, Junt T., Wong H.C., Chakraborty A.K., von Andrian U.H.
    T cell sensing of antigen dose governs interactive behavior with dendritic cells and sets a threshold for T cell activation
    Nature Immunology, 9(3), 282-291 (2008)

13. **Artyomov M.N.**, Das J., Kardar M., Chakraborty A.K.
    Purely stochastic binary decisions in cell signaling models without underlying deterministic bistabilities
    Proceedings of the National Academy of Sciences of USA, 104, 18958-18963 (2007)

    • See coverage of this work in **Physics Today**, July 2008, page 28.

14. **Artyomov M.N.**, Morozov A.Yu., Pronina E., Kolomeisky A.B.
    Dynamic properties of molecular motors in Burnt-Bridge Motors
    Journal of Statistical Mechanics – Theory and Experiment, P08002 (2007)

15. Morozov A.Yu., Pronina E., Kolomeisky A.B., **Artyomov M.N.**
    Solution of burnt-bridge models for molecular motor transport
    Physical Review E, 75(3) 031910 (2007)

16. **Artyomov M.N.**, Freed K.F.
    Actin polymerization under pressure: A theoretical study
    Journal of Chemical Physics, 126(2) 024908 (2007)

17. **Artyomov M.N.**, Freed K.F.
    Compressible models of equilibrium polymerization
    Journal of Chemical Physics, 123(19) 194906 (2005)

18. **Artyomov M.N.**, Kolomeisky A.B.
    Thermodynamics and phase transitions of electrolytes on lattices with different discretization parameters
    Molecular Physics, 103(21-23) 2863-2872 (2005)

19. **Artyomov M.N.**, Kobelev V., Kolomeisky A.B.
    Lattice models of ionic systems with charge asymmetry
    Journal of Chemical Physics, 118(14) 6394-6402 (2003)

20. Durov V.A., **Artemov M.N.**, Shilov I.Y. (name spelled incorrectly in the translated version)
    Molecular structure and dielectric properties of associated liquids: Models of comblike associates with branches of arbitrary length
    Russian Journal of Physical Chemistry, 79(12) 1955-1966 (2005)

21. Durov V.A., **Artemov M.N.** (name spelled incorrectly in the translated version)
    Molecular structure and dielectric properties of associated liquids. The simplest model of comb association
    Russian Journal of Physical Chemistry, 79(2) 207-214 (2005)

Presentations:

- **May 10, 2009** – **Oral Presentation** at AAI 2009 (imunology) Meeting in Seattle:
  Role of the Coreceptor in Early T-cell Signaling

- **Mar 19, 2009** – **Oral Presentation** at APS March Meeting in Pittsburgh:
  Reaction-Diffusion vs Reactions in cell signalling

- **Mar 16, 2009** – **Oral Presentation** at APS March Meeting in Pittsburgh:
  Stochastic bimodality without underlying bistability

- **Jan 11, 2009** – **Poster Presentation** at Mini-Stat-Mech Meeting, UC Berkeley:
  Mechanism of somatic cells reprogramming to induced embryonic cells

- **Aug 20, 2008** – **Oral Presentation** at ACS Fall Meeting in Philadelphia:
  Molecular Motors Interacting with Their own Tracks

- **Aug 1, 2008** – **Oral Presentation** at Les Houches Summer School Seminar:
  Molecular Motors Interacting with Their own Tracks

- **Jul 8, 2007** – **Poster Presentation** at IUPAP StatPhys23 Conference, Genova, Italy:
  Purely stochastic binary decisions in cell signaling models without underlying deterministic bistabilities

Professional Activities:

- **Jun-Aug 2009** – **Internship in Vertex Pharmaceuticals** (Cambridge, MA)
  study of signalling pathways affected by IBD

- **Aug 2008** – **Les Houches Theoretical Physics Summer School**
  on Long-Range Interactions and Non-Equilibrium Phenomena

- **Sep 2007** – Institute Poincare' **Trimester on Non-Equilibrium Statistical Physics**,
  attended Month I: Non-Equilibrium Steady States, IHP, Paris, France

- **Jan 2007** – D.Frenkel' &B.Smit' **Understanding Molecular Simulations Tutorial**
  University of Amsterdam, Netherlands

- **Jul 2006** – **Boulder Summer School in Condensed Matter and Materials Physics**: Physics of Soft Matter, Complex Fluids and Biological Materials. Boulder, CO, USA

References:

Arup K. Chakraborty

Robert T. Haslam Professor of Chemical Engineering
Professor of Chemistry, Professor of Biological Engineering
Department of Chemical Engineering
Massachusetts Institute of Technology
Room E19-502C
77 Massachusetts Ave
Cambridge MA 02139 USA
Phone: 617.253.3890
Fax: 617-258-5766
Email: arupc@mit.edu
Web: http://web.mit.edu/akcgroup

Mehran Kardar

Professor of Physics
Department of Physics
Massachusetts Institute of Technology
Room 6C-315
77 Massachusetts Ave
Cambridge MA 02139 USA
Phone: (617) 253-3259
Fax: (617) 253-2562
Email: kardar@mit.edu
Web: http://web.mit.edu/kardar/www/

Anatoly B. Kolomeisky

Associate Professor
Rice University
Department of Chemistry, Department of Chemical and Biomolecular Engineering
6100 Main Street
Houston, TX 77005-1892
Office: Biology 306
tel. (713) 348-5672
fax. (713) 348-5155
E-mail: tolya@rice.edu
http://python.rice.edu/~kolomeisky/