ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

A MATHEMATICAL MODEL FOR $\alpha\beta$ T CELL DIFFERENTIATION IN THE THYMUS

M.Sc. THESIS

Emrah ŞİMŞEK

Department of Physics Engineering

Physics Engineering Programme

JUNE 2012

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

A MATHEMATICAL MODEL FOR $\alpha\beta$ T CELL DIFFERENTIATION IN THE THYMUS

M.Sc. THESIS

Emrah ŞİMŞEK
(509091114)

Department of Physics Engineering

Physics Engineering Programme

Thesis Supervisor: Prof. Dr. Sondan DURUKANOĞLU FEYİZ

JUNE 2012

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

TİMUSTA $\alpha\beta$ T HÜCRESİ FARKLILAŞMASI İÇİN BİR MATEMATİKSEL MODEL

YÜKSEK LİSANS TEZİ

Emrah ŞİMŞEK
(509091114)

Fizik Mühendisliği Anabilim Dalı

Fizik Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Sondan DURUKANOĞLU FEYİZ

HAZİRAN 2012

Emrah ŞİMŞEK, a **M.Sc.** student of ITU **Graduate School of Science, Engineering and Technology** student ID 509091114 successfully defended the **thesis** entitled **"A MATHEMATICAL MODEL FOR $\alpha\beta$ T CELL DIFFERENTIATION IN THE THYMUS"**, which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Sondan DURUKANOĞLU FEYİZ** Sabancı University	.....
Jury Members :	**Prof. Dr. Sondan DURUKANOĞLU FEYİZ** Sabancı University	.....
	Assoc. Prof. Dr. Haluk ÖZBEK Istanbul Technical University	.....
	Assoc. Prof. Dr. Batu ERMAN Sabancı University	.....

Date of Submission : **4 May 2012**
Date of Defense : **7 June 2012**

To my mother,

FOREWORD

I am very grateful to my supervisor Professor Sondan DURUKANOĞLU FEYİZ for her extraordinary help on my academic career and personal development. I especially thank to Dr. Batu ERMAN for his critical commentaries on my studies from a respectable experimental biologist's point of view. I thank my precious colleagues Onur PUSULUK and Berkin MALKOÇ for fruitful discussions during my master studies, and Dr. Cem SERVANTİE for his critical reading the draft. I also would like to thank Mahmut ELBİSTAN for our funny and ingenuous office-mateship and all members of ITU Physics Engineering Department for their sincere friendships.

At last but not least, I would like to express my gratitude to my family for always being with me.

June 2012

Emrah ŞİMŞEK
(Physicist)

TABLE OF CONTENTS

Page

ABBREVIATIONS

APC : Antigen Presenting Cell
CD4 SP : CD4 Single Positive
CD8 SP : CD8 Single Positive
DN : Double Negative
DNA : Deoxyribonucleic Acid
DP : Double Positive
Gata3 : Symbol Of The Gene That Encodes GATA3
GATA3 : GATA-binding Protein 3
HD : Helper Deficient
IL-7 : Interleukin 7
IL-7R : Interleukin 7 Receptor
IL-7Rα : One Of The Two Subunits Composing A Functional Interleukin 7 Receptor
mRNA : Messenger Ribonucleic Acid
MHC : Major Histocompatibility Complex
NK : Natural Killer
ODE : Ordinary Differential Equation
RNA : Ribonucleic Acid
SQDSM : Standardized Qualitative Dynamical Systems Method
STAT : Signal Transducer and Activator Of Transcription
ThPOK : Zinc Finger Protein T-helper-Inducing POZ/Kruppel-like Factor
TCR : T Cell Antigen Receptor
TCRβ : One Of The Two Subunits Composing A T Cell Antigen Receptor
WT : Wild Type
Zbtb7b : Symbol Of The Gene That Encodes ThPOK

LIST OF TABLES

Page

LIST OF FIGURES

Page

A MATHEMATICAL MODEL
FOR
$\alpha\beta$ T CELL DIFFERENTIATION IN THE THYMUS

SUMMARY

As being a relatively new and an intricate research area of life sciences, immunology is a still evolving subject in which scientists from many different disciplines like biology, medicine, physics, chemistry, mathematics, computer sciences, etc. are joint to gain a deeper understanding on how the immune system reveals its functions.

In most mammalian species, the immune system can be mainly subcategorized into three levels of defense against pathogens: natural barriers, innate immunity and adaptive immunity. Natural barriers are the first line of defense that has to be penetrated by pathogens in order to cause disease and it exists in almost all living organisms. Any invader that penetrates the natural barriers is greeted by innate immune system which is the second line of defense. Innate immunity operates relatively quick, reacts to a variety of usual pathogenic organisms and it has not specific elements against to any particular pathogen. It also activates and controls the adaptive immunity. Almost all organisms get along just fine with only natural barriers and the innate immune system to defend them. However, in the vertebrates, the innate responses call into play the third level of defense: 'adaptive' immunity which has specifically equipped soldiers to cope with almost any foes. Moreover, players of the innate and the adaptive immune systems usually work together to eradicate pathogens. The main factors distinguishing the innate immunity and the adaptive immunity are *timing* and *specificity* of the response against to a pathogenic attack. Both of the innate and the adaptive immune responses depend upon the activities of white blood cells (called as *leukocytes*), which are originated from bone marrow-derived hematopoietic stem cells. Adaptive immune responses are provided by white blood cells called *lymphocytes* being subdivided into two classes as *antibody responses* and *cell mediated responses*, which are carried out by B- and T-cells, respectively. T cells develop in the *thymus*, and B cells, in mammals, develop in the *bone marrow* in adults or in the liver in fetuses.

Pluripotent progenitors of T lymphocytes are produced in the bone marrow like all the other hematopoietic cells, and migrate to the thymus gland for differentiating and eventually committing to different T cell subsets: *cytotoxic*, *helper* and *regulatory (suppressor)* T cells.

Thymic population of T cells is mainly composed of $\alpha\beta$ subset and $\alpha\beta$ thymocytes commit to either *helper T cells* or *cytotoxic T cells* at mature stage. Differentiation process leads to exclusive expression of CD4 and CD8 proteins on the surfaces of helper and cytotoxic T cells, respectively. These *coreceptor* proteins have indispensible roles in the TCR signaling events that modulate cell fate decisions. An immature thymocyte entering into the thymus undergoes the sequential stages of double negative (DN)- $CD4^{-}CD8^{-}$, double positive (DP)- $CD4^{+}CD8^{+}$, Intermediate-

$CD4^+CD8^{low}$ to become either a $CD4^+CD8^-$ helper or a $CD4^-CD8^+$ cytotoxic mature T cell.

The study of genetic regulatory systems has received a major impetus from the recent development of experimental techniques by which spatio-temporal expression levels of genes to be measured. Together with these still developing high throughput experimental tools, it is indispensable employing theoretical models and computer simulations in order to elicit structure and dynamics of the genetic regulatory network that underlies the CD4/CD8 fate decision.

In theoretical biology, the conventional technique in building a regulatory network model for a cell differentiation process is to define different attractors (*or* equilibrium states) in the landscape picture corresponding to different cell types. With this motivation, we aim to build a mathematical model which qualitatively describes differentiation of $\alpha\beta$ thymocytes, particularly beyond the Intermediate stage, as a dynamical sytem. Hence, we form a regulatory network model of 8 components and 13 regulatory interactions among them, using environmental cues and regulatory proteins that are implied to have important roles on the phenomenon in the literature.

To convert our model into a dynamical representation, we adopt a standardized qualitative dynamical systems method which is an ordinary differential equation formalism in nature. In the method, state of each node in a regulatory network can be updated in time by taking into account the regulatory effects by the others and itself with some specified parameters, namely strengths of activations, inhibitions, steepness of the response curves and decay rates. But, in biology it is very ubiqitous that a regulatory event can only occur in the co-existence of two or more regulatory elements and the method fails to mimic such events. Thus, we further contribute to the method by adding (only second order) co-regulatory terms.

By utilizing the improved method, we obtain a set of 8 nonlinear ODEs, each one describing the time derivative of an independent variable in the network. Since there is no reliable kinetic data yet, we choose parametric values for the equations to be not favoring any specific interaction or decay and to make values of the variables Boolean-like at equilibrium states. Then, first, we explore the fixed points of the system utilizing *fsolve* optimization toolbox and *ODE45* system solver of MATLAB. All biologically meaningful fixed points are named Intermediate, CD4 SP and CD8 SP attractors depending on the activation patterns for the components. Second, we investigate the effects of TCR and IL7 signalings onto CD4/CD8 fate decision *in silico*: TCR signals with long duration lead to differentiation into CD4 SP whereas IL7 signals with short duration cannot secure the CD8 lineage alone. Finally, we check the results of salient component overexpression/knockout experiments in computer simulations and capture good agreement with experimental observations in the literature (except for some cases).

Further studies are needed to extend our model to one that describes the whole picture of "DP to SP" transition in which coreceptor proteins have feedback effects in TCR signaling events.

TİMUSTA $\alpha\beta$ T HÜCRESİ FARKLILAŞMASI İÇİN BİR MATEMATİKSEL MODEL

ÖZET

Yaşam bilimlerinin yeni ve karmaşık bir araştırma alanı olarak immünoloji, biyoloji, tıp, fizik, kimya, matematik, bilgisayar bilimleri gibi pek çok farklı disiplinden bilim insanlarının bir araya gelip, bağışıklık sisteminin fonksiyonlarını nasıl ortaya koyduğunu anlamaya yönelik çalıştığı ve hala gelişen bir konudur.

Çoğu memeli türlerinde bağışıklık sistemi patojenlere karşı verilen savaşta üç farklı savunma hattı olarak gruplandırılabilir: doğal bariyerler, doğuştan bağışıklık ve edinilen bağışıklık. Doğal bariyerler, hastalığa sebep olabilmek için patojenler tarafından geçilmesi gereken, savunmanın birinci hattıdır. Bu bariyerler neredeyse tüm yaşayan organizmalarda bulunur. Doğal bariyerlerden sızan bir istilacı, savunmanın ikinci hattı olan doğuştan bağışıklık sistemi tarafından karşılanır. Bu bağışıklık tipi 'doğuştan' olarak adlandırılır, çünkü neredeyse tüm canlılarda doğal olarak bulunur. Doğuştan bağışıklık göreceli olarak çabuk çalışır, çok sayıda ve çeşitli genel patojenlere karşı tepki gösterir ve herhangi bir patojene karşı özel savunma elemanlarına sahip değildir. Aynı zamanda edinilen bağışıklık sistemini etkinleştirir ve kontrol eder. Neredeyse tüm canlılar sadece doğal bariyerler ve doğuştan bağışıklık sistemleri tarafından savunularak yaşamlarını sürdürebilir. Bununla beraber, omurgalılarda doğuştan bağışıklık tepkileri savunmanın üçüncü bir hattını oyuna davet eder: neredeyse tüm düşmanlarla baş etmek üzere özel askerlere sahip olan 'edinilen' bağışıklık. Çoğu zaman, doğuştan ve edinilen bağışıklık sistemlerinin elemanları patojenlerin kökünü kazımak için birlikte çalışır. Doğuştan ve edinilen bağışıklık sistemlerini birbirinden ayıran ana faktörler, bir patojene karşı verilen tepkinin *zamanlaması* ve *özgünlüğü*dür. Doğuştan ve edinilen bağışıklık tepkilerinin her ikisi de, kemik iliğinde üretilen kan kök hücreleri kökenli beyaz kan hücrelerinin (lökositler) etkenliklerine bağlıdır. Sırasıyla B- ve T- hücreleri tarafından yürütülen, antikor tepkileri ve hücre ortamlı tepkiler olarak iki alt sınıfa ayrılan edinilen bağışıklık tepkileri, lenfositler olarak adlandırılan beyaz kan hücreleri tarafından sağlanır.

T ve B hücreleri isimlerini geliştikleri organlardan alırlar. T hücreleri *timus*ta gelişir. Memelilerin B hücreleri, yetişkinlerde *kemik iliği*nde ve ceninlerde karaciğerde gelişir. Aslında B ve T hücreleri, köken olarak, aynı *genel lenfoid öncül hücreler*den iki kola ayrılırlar.

T lenfositlerinin çok potansiyelli öncülleri, diğer bütün kan kökenli hücreler gibi, kemik iliğinde üretilir, farklılaşmak üzere timus bezine göç eder ve nihayetinde farklı özellikteki T hücresi alt gruplarına katılır: *katil (sitotoksik)*, *yardımcı* ve *düzenleyici (baskılayıcı)* T hücreleri. Etkin (efektör) bir katil T hücresi enfekte hücreyi, enfekte hücrenin yüzeyinde MHC sınıf I molekülleri tarafından sunulan kendisine has olan antijeni tanıdığında, doğrudan öldürür. Öte yandan, etkin bir yardımcı T hücresi,

enfekte hücrenin yüzeyinde MHC sınıf II molekülleri tarafından sunulan kendisine has olan antijeni tanıdığında, uyarıcı moleküller salgılama yoluyla, makrofajlar, B hücreleri ve katil T hücreleri gibi diğer bağışıklık sistemi elemanlarını göreve çağırır. Düzenleyici T hücrelerinin bağışıklık sistemindeki rolü ise tam olarak saptanamamış olmasına rağmen, dallantılı (dendritik) hücrelerin, yardımcı ve katil T hücrelerinin fonksiyonlarını düzenlediklerine inanılmaktadır.

T hücrelerinin timustaki popülasyonu genel olarak, olgun aşamada yardımcı ve katil T hücreleri olarak iki alt gruba ayrılan $\alpha\beta$ T hücreleri grubundan oluşur. Farklılaşma süreci, yardımcı ve katil T hücrelerinin yüzeylerinde, sırasıyla, yalnızca CD4 ve CD8 proteinlerinin ifade edilmesine yol açar. Bu *müşterek-almaç (koreseptör)* proteinleri hücre kader kararlarını ayarlayan T hücresi almacı sinyalleşmesi olaylarında vazgeçilmez rollere sahiptir.

Alışılageldiği üzere, timusta olgunlaşmakta olan T hücrelerinin (timositlerin) gelişimsel aşamaları CD4 ve CD8 proteinlerinin ayrımcı (diferansiyel) ifade edilmesi ile tanımlanır: Timusa yeni giren olgunlaşmamış bir timosit, bir $CD4^{+}CD8^{-}$ yardımcı veya bir $CD4^{-}CD8^{+}$ katil T hücresi olmak için, birbirini izleyen çift negatif (ÇN)- $CD4^{-}CD8^{-}$, çift pozitif (ÇP)- $CD4^{+}CD8^{+}$, Ortanca- $CD4^{+}CD8^{az}$ aşamalarından geçer. (Burada, farklı aşamaları gösteren bu semboller CD4 ve CD8 proteinlerinin hücre yüzeyinde bulunup bulunmadığını anlatır. $CD4^{+}CD8^{az}$ ile simgelenen Ortanca aşamada ise CD8 proteini az da olsa hücrenin yüzeyinde bulunur.)

ÇN aşamasındaki bir timositin yüzeyinde T hücresi β almacı olarak adlandırılan öncül bir T hücresi antijen almacı tipi bulunur. Bu almaç uyarıldığında ÇN timosit β *seçilimi* olarak adlandırılan süreci yaşayarak ÇP aşamasına geçer. ÇP aşaması, CD4 ve CD8 proteinlerinin her ikisinin de hücre yüzeyinde yüksek miktarlarda bulunduğu ve aynı zamanda tam fonksiyonlu bir T hücresi antijen almacının timositlerin yüzeyinde ilk defa ortaya çıktığı aşamadır. ÇP aşamasındaki bir timosit antijen uyarımı alırsa, ölmekten kurtulmuş olur (*pozitif seçilim*) ve nihayetinde ya yardımcı ya da katil bir T hücresi olarak olgunlaşmasını tamamlar.

Bugün CD4/CD8 soy seçimini en iyi açıklayan model olarak *kinetik sinyalleşme modeli* yaygın biçimde kabul görmektedir. Kinetik sinyalleşme modelinde, kısa süreli T hücresi almacı sinyalleri CD8'e farklılaşma yolağına neden olurken uzun süreli T hücresi almacı sinyalleri CD4 soyuna farklılaşmanın sürücüsüdür. Eğer Ortanca aşamada T hücresi almacı sinyalleri kesilirse, interlökin 7 almacı CD8 T hücrelerine farklılaşmayı destekleyen interlökin 7 sitokinlerini alabilir. Kinetik sinyalleşme modelinin özgün iki prensibi şunlardır: *i*) Pozitif seçilim ve bir olgun hücre grubu kaderininin seçimi aynı T hücresi antijen almacı sinyalleşmesi ile tetiklenen, eş zamanlı olaylar olmanın aksine, ayrık ve ardışık olaylardır. *ii*) ÇP aşamasından sonra müşterek-almaç proteinlerinden herhangi birisinin üretiminin durdurulması tersinemez bir olay değildir. Yani süreç içersinde üretimi durdurulan bir müşterek-almaç proteini (CD4 veya CD8), daha sonra tekrar üretilmeye başlanabilir.

Genetik düzenleyici sistemlerin çalışılması, genlerin ifade edilme düzeyleri hakkında uzay-zaman bilgisini ölçebilen en son deneysel tekniklerin gelişimi ile büyük bir ivme kazanmıştır. Bu halen gelişmekte olan yüksek işlem hacimli deneysel araçlarla birlikte, CD4/CD8 kader kararının altında yatan genetik düzenleyici ağın yapısını

ve dinamiğini meydana çıkarmak üzere, teorik modeller ve bilgisayar benzetimleri kullanmak kaçınılmazdır.

Doğayı modellemek için kullanılan en geleneksel araç *diferansiyel denklemler*dir ve genetik düzenleyici ağları modelleme çalışmalarında da yaygın biçimde kullanılmaktadırlar. Adi diferansiyel denklemler formalizmi gerçek dünyadaki dinamik sistemler üzerine çalışmak üzere yaygın bir modelleme aracıdır. Bu formalizmini kullanmak gen düzenleyici ağların içersinde yer alan RNA, proteinler vb. düzenleyici elemanların derişimlerini zamana bağımlı ve negatif olmayan gerçel sayılar olan değişkenler ile modellemeye imkan verir.

Hücre farklılaşmasının, farklı hücre tiplerinin matematiksel olarak değişik çekici noktalar (*yani* denge durumları) olarak tanımlanabildiği, peyzaj (lendskeyp) betimlemeleri çerçevesinde ele alınabildiği düzenleyici ağlar oluşturmak teorik biyoloji çalışmalarında geleneksel bir usüldür. Buradan hareket ederek, $\alpha\beta$ timositlerin özel olarak Ortanca aşamasının ötesine farkılaşmalarını dinamik ve nitel bir temsil ile ele alan bir matematiksel model oluşturmayı amaçlamaktayız. Bu yüzden, 8 farklı öğe ve bunlar arasındaki 13 farklı etkileşmenin bir düzenleyici ağ modelini oluşturduk. Bu kaba modelimiz, literatürde problem üzerine önemli rollerinin bulunduğu gösterilen çevresel işaretler ve düzenleyici proteinleri içermektedir.

Timusta T hücrelerinin farklılaşmalarına ilişkin olarak gözleme dayalı bilginin çokluğuna rağmen, dinamik temsillerde kullanılmak üzere devinsel (kinetik) veri ve gerçek derişim değerlerini elde etme çalışmaları halen emekleme dönemindedir. Bu yüzden, modelimizi dinamik bir temsile çevirmek için, bir ölçünlenmiş (standardize) nitel dinamik sistemler yöntemini benimsedik. Bu yöntemde, düzenleyici ağ içersindeki her bir elemanın durumu, diğer elemanlardan (ve hatta kendisinden) dolayı üzerine etkiyen düzenleyici etkileri hesaba katarak, zaman içersinde güncellenir. Fakat, biyolojide bir düzenleyici olayın gerçekleşmesinin ancak ve ancak iki veya daha fazla müşterek-düzenleyici (ko-regülatör) elemanın eş zamanlı varlığı altında olabilmesi sıklıkla karşılaşılan bir durumdur. Bu yüzden, yönteme müşterek-düzenleyici (sadece ikinci dereceden) terimleri ekleyerek geliştirdik.

Bu geliştirilmiş yöntemi kullanarak, her biri ağdaki farklı bir bağımsız değişkenin zamana göre değişimini tanımlayan 8 tane doğrusal olmayan adi diferansiyel denklemden oluşan bir denklem seti elde ettik. Henüz tam anlamıyla güvenilir devinsel veri mevcut olmadığından, denklemler sistemimiz için gerekli olan parametreleri herhangi bir etkileşimi ya da bozunma olayını özellikle desteklemeyecek ve değişkenlerin denge durumlarındayken alacakları etkenlik değerleri Boole değişkenleri gibi (yaklaşık olarak 0 ve 1) olacak şekilde seçtik. İlk önce, MATLAB' ın *fsolve* optimizasyon araç çubuğunu ve *ODE45* adi diferansiyel denklem sistemi çözücüsünü kullanarak sistemin sabit noktalarını ortaya koyduk ve bu sabit noktaların kararlılıklarını inceledik. Elde edilen biyolojik olarak anlamlı olan sabit noktalar ağdaki bileşenlerin etkinlik düzeyi motiflerine göre Ortanca, CD4 ve CD8 çekici noktaları olarak adlandırıldılar. İkinci olarak, modelimize göre T hücresi almacı ve interlökin 7 sinyalleşmelerinin CD4/CD8 kader seçimi üzerine olan etkilerini inceledik. Modele göre, uzun süreli T hücresi almacı sinyalleşmeleri, kinetik sinyalleşme modelinin de kabul ettiği gibi, CD4 kaderine yönelime neden olur ve sadece kısa süreli interlökin 7 sinyalleşmeleri CD8 kaderine farkılaşmanın belirleyicisi

olamaz. Son olarak, ağdaki elemanların literatürde verili olan belli başlı aşırı ifade olunma/devre dışı bırakılma deneylerinin sonuçlarını bilgisayar benzetimlerimizde kontrol ettik. Birkaç durum dışında gözlemsel bulgularla iyi bir mutabakat sağladığımızı gördük.

İlerisi için modelimizi, müşterek-almaç proteinlerinin T hücresi almacı sinyalleşme olaylarında geri besleme rolüne sahip olacağı, Çift Pozitif aşamasından olgun aşamaya geçişin tam bir resmini verebilecek şekilde, genişletmek üzere yeni çalışmalara ihtiyaç vardır.

1. INTRODUCTION

T lymphocytes (or thymocytes) are originated from the bone marrow like all the other hematopoietic cell types and migrate to the thymus gland where they mature into T cells. The progenitor T cells which are able to show an appropriate $\alpha\beta$ T cell antigen receptor on their surfaces in the thymus, mainly differentiate into the helper T cell ($CD4^+$ single positive), cytotoxic T cell ($CD8^+$ single positive) lineages. Although there has been an increasing surge in obtaining experimental data to determine the underlying molecular and genetic mechanisms in the differentiation of baby T cells into mature ones, a mathematical model describing the dynamic nature of a network specific to this differentiation is hardly in the scope of simulations. In this thesis, we construct a network model of 8 components and 13 regulatory interactions that are mostly important for understanding the differentiation mechanism and dynamics by utilizing a comprehensive scanning of the $\alpha\beta$ T cell literature. We treat our model as a continuous dynamical system by using a standardized qualitative dynamical systems method of Luis Mendoza and Ioannis Xenarios (2006) which operates essentially based on a set of ODEs. (Details will be given in Section 4.1.3.1). This method can be used not only to deal with such a cell differentiation problem but also to investigate all kinds of regulatory network problems having poor stoichiometric and kinetic data. We further improve the method by adding second order regulatory input terms. (Details will be given in Section 5.1).

Each node in the network represents a normalized value in the closed interval $[0,1]$ of activation level of a particular transcription factor protein, a cell signaling mediatory protein, a cytokine or a gene at any time t. In addition to capturing functional capabilities of the system without knowledge of any kinetic parameters or real concentrations, the adapted method can easily be updated by possible upcoming data of future works based on the advantage of usage of a normalized activation level value for each node rather than a certain concentration and has the ability to operate as both a continuous formalism and a discrete one by simply changing only a single parameter.

After we present an overview of the immune system in the second chapter, in the third chapter we give some detailed information about differentiation of thymocytes developing in the thymus. In the fourth chapter, we mention some fundamental concepts concerning mathematical and network modeling of biological regulatory systems, and introduce the adopted mathematical formalism for getting closer to the computer simulations of our network. In the fifth (and final) chapter we construct our regulatory network model, and formulate a set of ODEs which gives a qualitative description on dynamics of T cell differentiation in the thymus. We obtain good agreement between steady state patterns of our mathematical model and activation patterns belonging to thymocyte populations at distinct stages of differentiation, i.e. progenitors and their offsprings. Furthermore, in Section 5.3.3, we introduce salient *in silico* perturbations on the topology of the network which can lead to blockade of one or both of the two possible mature subsets or lineage redirection of thymocytes differentiating into either CD4 SP or CD8 SP fates in computer simulations. We also conclude our results by comparing with experimental ones as long as it is possible and make some recommendations for future research which would help to reveal the underlying mechanism of the differentiation process.

2. A BRIEF REVIEW OF THE IMMUNE SYSTEM

2.1 History

As being a relatively new and an intricate research area of life sciences, immunology is a still evolving subject in which scientists from many different disciplines like biology, medicine, physics, chemistry, mathematics, computer sciences, etc. are joint to gain a deeper understanding on how the immune system reveals its functions.

The origin of modern immunology is commonly ascribed to Edward Jenner who discovered in 1776 that cowpox (or vaccinia), brought protection against human smallpox, which was a widespread fatal disease of the era. The term 'vaccination' refers to inoculation of healthy individuals with weakened disease-causing agents to provide protection from the disease. It was named after Jenner's procedure using vaccinia. When Jenner introduced vaccination he knew nothing of the infectious agents that cause disease. Then, Robert Koch proved that infectious diseases are caused by microorganisms called *pathogens* (such as viruses, bacteria, pathogenic fungi, parasites, etc.), and each one of them is responsible for a particular disease, or *pathology* [4].

Such discoveries in 19th century, stimulated the extension of Jenner's strategy of vaccination to other diseases. In the 1880s, Louis Pasteur excogitated a vaccine against cholera in chickens, and brought forth a rabies vaccine that achieved a striking success upon its first trial in a boy bitten by a rabid dog. These practical triumphs led to investigations on the mechanisms of protection and to the development of the science of immunology. In 1890, Emil von Behring and Shibasaburo Kitasato discovered that the blood serum[1] of vaccinated individuals contained substances which they called *antibodies* that specifically bound to a particular pathogenic fragment.

[1] Clear yellowish fluid component of the blood including neither blood cells such as white and red blood cells nor clotting factors. It is obtained upon seperating whole blood into its solid and liquid components after it has been coagulated.

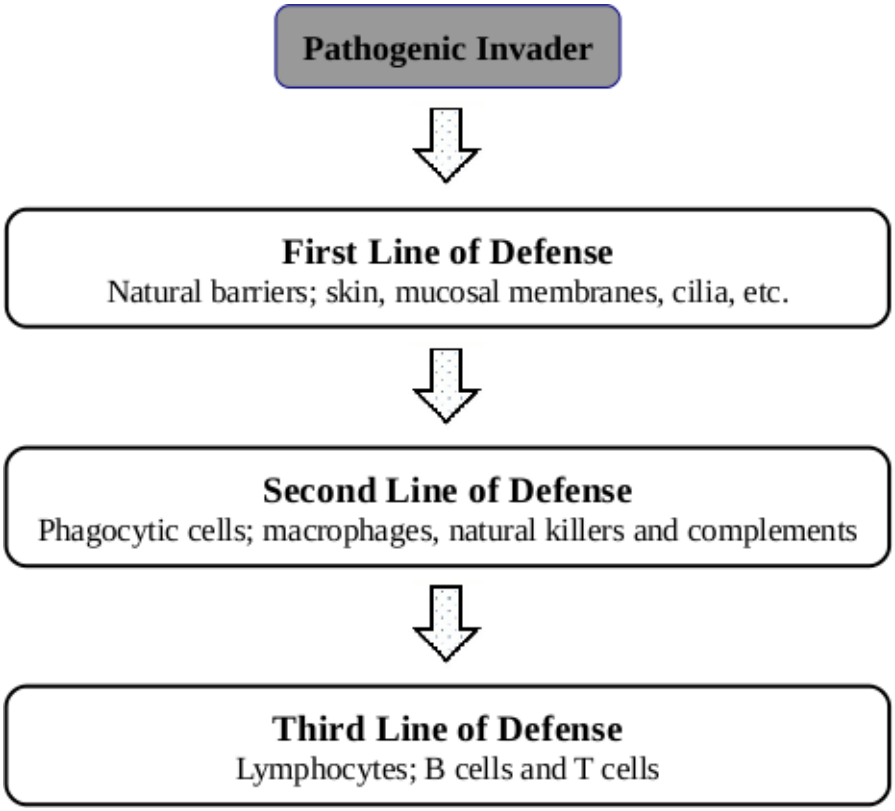

Figure 2.1: An overall aspect of the immune system.

Indeed, it quickly came out that specific antibodies can be induced against a vast range of pathogenic fragments. Such fragments are known as *antigens* because they can stimulate the *gen*eration of *anti*bodies [4].

2.2 The Immune System

In livings, the magnificent orchestra composed of several types of cells, tissues and organs which are responsible to immune functions is referred as 'the immune system'. The immune system is, therefore, a 'network' of a large number of components which interact with each other through many different ways. In most mammalian species, it can be mainly subcategorized into three levels of defense against pathogens: natural barriers, the innate immunity and the adaptive immunity as sketched in Fig.2.1.

Natural barriers are the first line of defense that has to be penetrated by pathogens in order to cause disease and it exists in almost all living organisms. The main factors distinguishing the innate immunity and the adaptive immunity are *timing* and

specificity of the response against to a pathogenic attack. In practice, there are alot of interactions between them and sometimes natural barriers are counted as a preceding subpart of the innate immunity. Both of the innate and the adaptive immune responses depend upon the activities of white blood cells (called as *leukocytes*), which originate from bone marrow-derived hematopoietic stem cells. Since these stem cells can give rise to all of the different types of blood cells, they are referred as *pluripotent progenitor cells*.

2.2.1 Natural Barriers

As forming the first level of defense comprising several natural barriers such as mechanical, chemical and biological barriers, they can protect almost any organism from infection. Pathogenic agents must first breach natural barriers to cause trouble. The outer line of defense mainly operates through skin, cilia, mucous membranes of digestive, respiratory, and reproductive tracts, etc. and provides a challenging media in order to drive back intruders [1].

2.2.2 Innate Immunity

Any invader that penetrates the natural barriers is greeted by the innate immune system which is the second line of defense. This type of immunity is called 'innate' because it is a type of defense that almost all livings naturally have [1]. The innate immunity operates relatively quick (a typical battle with an invader takes a few days), reacts to a variety of usual pathogenic organisms and it has not specific elements against to any particular pathogen. It also activates and controls the adaptive immunity. *Complement proteins*, *professional phagocytes*, and *natural killers* are the most important players of the innate team [1].

2.2.2.1 The complement system

Over twenty different proteins present at high concentrations in blood and in tissues 'complement' the killing of pathogens by antibodies. Any invader having a surface with a spare hydroxy or amino group can be bounded by these complement proteins.

Figure 2.2: Electron micrograph of a macrophage [1].

The complement system has also the ability to alarm other immune system players by reacting very fast in response to a pathogenic attack [1].

2.2.2.2 Professional phagocytes

Professional phagocytes make their living mainly by eating, which is their 'professional' job. The most important ones are *macrophages* and *neutrophils* [1].

Macrophages

A Russian immunologist Elie Metchnikoff discovered that many microorganisms could be eaten by phagocytic cells, which he called *macrophages*[2]. Macrophages are available to struggle against a wide range of pathogens without requiring prior exposure and are the cardinal player in the team of the innate immune system [4]. While a macrophage is eating its meal, the meal is first engulfed in a pouch (vesicle) called 'phagosome'. This vesicle is then taken inside the macrophage and fuses with another vesicle called 'lysosome' which contains powerful chemicals and enzymes to destroy the food. The whole process is called 'phagocytosis'. Indeed, a macrophage is a very versatile cell since it functions as a garbage collector by eating almost everything that it comes across, as an antigen presenting cell[3], or as a vicious killer-depending on its activation level [1].

[2]Etymologically, macro refers to *large* and phage means *eater*, henceforth the term macrophage stands for *big eater*.

[3]Cells that display foreign antigen complexes with major histocompatibility complexes (MHCs) on their surfaces. These cells ingest and process antigens and present them to T-cells via interactions between their MHCs and T cell receptors on the surface of T cells.

Neutrophils

Neutrophils make up about 70% of the white blood cells in circulation, and about 100 billion of these cells are produced each day in the bone marrow. Neutrophils live for a very short time. In contrast to macrophages, neutrophils do not act as antigen presenters- they are only professional killers [1].

2.2.2.3 Natural killer cells

This has been a difficult cell population to be studied by researchers, because there are different kinds of NK cells with somewhat different properties. They can kill tumor cells, virus-infected cells, bacteria, parasites, and fungi [1].

2.2.3 Adaptive Immunity

Almost all livings get along just fine with only natural barriers and the innate immune system to defend them. However, in the vertebrates, the innate responses call into play the third level of defense: 'adaptive' immunity which has specifically equipped soldiers to cope with almost any foes. Moreover, players of the innate and the adaptive immune systems usually work together to eradicate pathogens [1, 2]. A specific immune response, such as the production of antibodies against a particular pathogen, is known as an adaptive immune response, because it occurs during the lifetime of an individual as an adaptation to infection with that pathogen [4]. A person who experienced an exposure to smallpox virus and could get rid of the infection, for example, is protected against smallpox by the adaptive immune system for the rest of his or her life, although not against any other viruses, such as those that cause mumps or measles. An adaptive immune response bestows, in general, lifelong protection against reinfection with the same pathogen [4]. While the phagocytic cells of the innate immune team can deal with a wide range of usual pathogens without requiring a prior exposure, antibodies of the adaptive system are produced only after infection. The adaptive system has also an immunological memory meaning that a living's response to the second exposure of a particular pathogen is earlier and stronger than that of its first exposure to the same pathogen. The antibodies present in a given person therefore directly reflects the infections to which he or she has been exposed [4].

Adaptive immune responses eliminate or destroy invaders and any toxic molecules they produce. Since these responses are very destructive, it is important that they are directed only against foreign molecules and not against molecules of the host organism. The adaptive immune system uses multiple mechanisms to avoid damaging responses against self molecules. Occasionally, however, these mechanisms fail, and the system turns against the host, causing *autoimmune diseases*, which can be fatal [2].

Adaptive immune responses are provided by white blood cells called **lymphocytes** being subdivided into two classes as *antibody responses* and *cell mediated responses*, which are carried out by B- and T-cells, respectively.

T cells and B cells derive their names from the organs in which they develop. T cells develop in the *thymus*, and B cells, in mammals, develop in the *bone marrow* in adults or the liver in fetuses. In fact, both T and B cells are originally bifurcated from the same *common lymphoid progenitor cells*. The common lymphoid progenitor cells themselves derive from multipotential *hematopoietic stem cells* being located primarily in hematopoietic tissues-mainly the liver in fetuses and the bone marrow in adults, which give rise to all blood cell populations, including red blood cells, white blood cells, and platelets (thrombocytes) [2].

2.2.3.1 B cells

In *antibody responses*, B cells are activated to secrete antibodies, which are essentially proteins called immunoglobulins. The antibodies circulate in the bloodstream and permeate the other body fluids, where they bind specifically to the antigen that stimulated their production. Binding of antibody inactivates viruses and microbial toxins by blocking their ability to bind to receptors on target cells. Antibody binding also marks invading pathogens for destruction, mainly by forming a link between cell surface proteins of pathogens and professional phagocytes to make it easier for phagocytic cells of the innate immune system to ingest them [2] as depicted in Fig.2.3.

2.2.3.2 T cells

In *T cell-mediated immune responses*, activated T cells react directly against a foreign antigen that is presented to them on the surface of a host cell, which is therefore

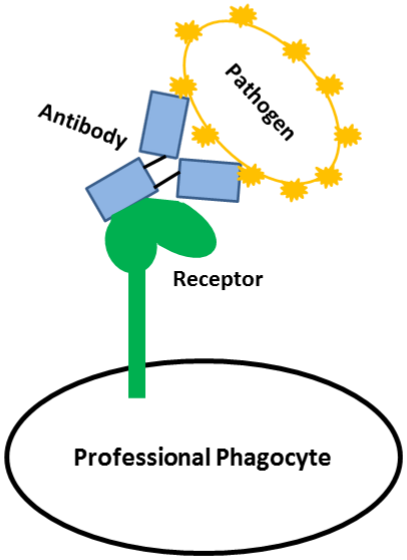

Figure 2.3: Antibodies secreted by B cells form a link between pathogenic agents and professional phagocytes.

referred to as an *antigen-presenting cell*. Remarkably, T cells which can detect pathogens on host cells either kill the infected cells or help other cells to wipe the invaders out. A T cell named as a *killer* (or *cytotoxic*) T cell, for example, might kill a virus infected host cell that has viral antigens on its surface, thereby eliminating the infected cell before the virus has had a chance to replicate. In other cases, the T cell called as *helper* T cell produces signal molecules that either activate macrophages to destroy the microbes that they have phagocytosed or invoke B cells to make antibodies against the microbes [2].

T and B cells become morphologically distinguishable from each other only after they have been activated by antigen. Resting T and B cells look very similar, even in an electron microscope. Both are small, only marginally bigger than red blood cells, and contain little cytoplasm (shown on the left in Fig.2.4). After activation by an antigen, both proliferate and mature into *effector cells*. Effector B cells secrete antibodies. In their most mature form, called *plasma cells*, they are filled with an extensive rough endoplasmic reticulum that is busily making antibodies (shown in the middle in Fig.2.4). In contrast, effector T cells contain very little endoplasmic

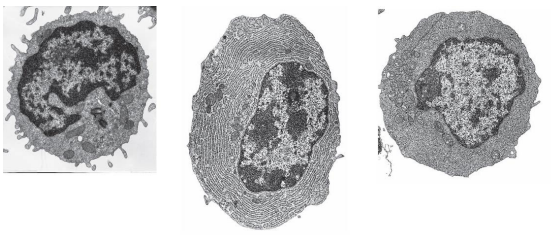

Figure 2.4: Micrographs of B and T cells [2].

reticulum (shown on the right in Fig.2.4) and do not secrete antibodies; instead, they secrete a variety of signal proteins called *cytokines*, which act as local mediators [2].

Whereas B cells can act over long distances by secreting antibodies that are widely distributed by the bloodstream, T cells can migrate to distant sites, but, once there, they act only locally on neighboring cells [2].

T cells must be stimulated by antigens via (T cell antigen receptors) TCRs on their surfaces to either proliferate or differentiate into effector cells. The stimulation can only occur when the antigen is displayed on the surface of *antigen-presenting cells (APCs)*, e.g. stromal cells in the thymus. Whereas B cells recognize intact antigenic proteins, T cells can recognize antigenic protein fragments (peptides) that have been partly degraded inside the antigen-presenting cell. In order to present antigens to TCRs, some protein complexes called as *MHCs; major histocompatibility complexes* are specialized to bind to the peptides and carry them to surface of the APCs where T lymphocytes can recognize them [2].

To briefly summarize their roles in the protection mechanism against invaders, it can be said that T cells survey the inside of cells while B cells survey the outside of the cells.

3. DEVELOPMENT AND DIFFERENTIATION OF T CELLS

3.1 Introduction

As being fundamental units of life, cells sense their environments via proteins on their surfaces, called *receptors*, and fulfill biological functions such as movement, secretion, growth, proliferation, differentiation, etc. in response to environmental cues. To convey a specific message inside the cell, a particular receptor must encounter its specific protein, named *ligand*. Such message delivery event is referred to as *signaling*. Once a receptor bounds to its particular ligand in adequate circumstances, it becomes stimulated and activated, promoting intracellular signaling pathways through interacting proteins in the cytosolic domain. At the end of signaling pathways, some proteins, termed *transcription factors*[1], translocate into the nucleus of the cell, in order to regulate expression of *ad hoc* genes to reveal the biological functions which are relevant to the incoming stimulus as sketched in Fig.3.1.

T cells are originated from a single stem cell that differentiates into several subsets of cells with specialized and exclusive functions. In such cellular differentiation processes, each offspring of a progenitor can still differentiate further until it adopts a specific cell fate. Every step of cellular differentiation leads to an increased specialization and molecular complexity.

Like all the other hematopoietic cells, pluripotent progenitors of T lymphycoytes are produced in the bone marrow, and migrate to the thymus gland to differentiate and eventually commit to different T cell subsets: *cytotoxic*, *helper* and *regulatory (suppressor)* T cells. An effector cytotoxic T cell directly kills the infected cell once it recognizes its particular antigen presented by *MHC class I* molecules on the surface

[1] In molecular biology and genetics, a transcription factor (sometimes called a sequence-specific DNA-binding factor) is a protein that binds to specific DNA sequences, thereby controlling the flow (*or* transcription) of genetic information from DNA to mRNA. Transcription factors perform this function alone or with other proteins in a complex, by promoting (as an *activator*), or blocking (as a *repressor*) the recruitment of RNA polymerase (the enzyme that performs the transcription of genetic information from DNA to RNA) to specific genes [5].

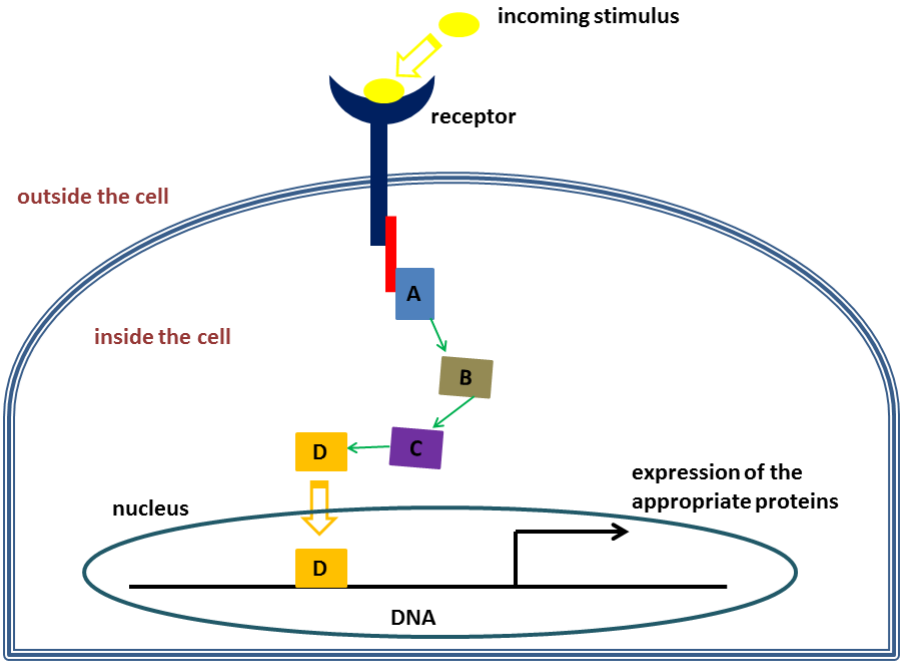

Figure 3.1: A cell does sense its environment via receptors and regulates its behavior in response to incoming stimulus.

of the target cell. An effector helper T cell, on the other hand, calls for the other immune system players such as macrophages, B cells and cytotoxic T cells through secreting stimulatory molecules once it recognizes its particular antigen presented by an *MHC class II* molecule on the surface of the infected cell. Although the functions of regulatory T cells in the immune system are not well established, they are believed to downregulate the function of helper T cells, cytotoxic T cells, and dendritic cells [4].

3.2 Formation of Helper and Cytotoxic Lineage ($\alpha\beta$) T cells

Differentiation of thymocytes in the thymus highly depends on intrathymic stimulations orchestrated by their TCRs. As thymocytes differentiate, they can express either $\alpha\beta$ TCRs or $\gamma\delta$ TCRs on their surfaces in the thymus. Let us remind that the $\alpha\beta$ and $\gamma\delta$ subsets having different functionalities are originally bifurcated from common progenitors.

Thymic population of T cells is mainly composed of $\alpha\beta$ subset that is subdivided into two fates at the mature stage: *helper T cells* and *cytotoxic T cells*. Differentiation

process leads to exclusive expression of CD4 and CD8 proteins on the surfaces of helper and cytotoxic T cells, respectively. These *coreceptor* proteins have indispensible roles in the signaling events that modulate cell fate decisions.

Conventionally, the developmental stages of the maturing thymocytes in the thymus are defined by differential expression of CD4 and CD8 coreceptors: An immature thymocyte entering into the thymus undergoes the sequential stages of DN- $CD4^{-}CD8^{-}$, DP- $CD4^{+}CD8^{+}$, Intermediate- $CD4^{+}CD8^{low}$ to become either a $CD4^{+}$ helper or a $CD8^{+}$ cytotoxic mature T cell. The earliest is the DN stage in which a thymocyte does not express neither TCR nor CD4/CD8 proteins. When DN thymocytes successfully rearrange the genes encoding the TCRβ chain they express pre-TCRs. Next, DN thymocyte goes through a β *selection* process when it is stimulated by its pre-TCRs to become a DP thymocyte. It is this stage at which a fully functional $\alpha\beta$ TCR is firstly expressed in the developmental pathway. DP thymocytes are also unique among intrathymic populations in that they express both CD4 and CD8 coreceptors and are unresponsive to the other survival signals of IL-7 [6]. Only a minority of thymocytes receiving signals through adequate TCR-MHC class I/II-CD8/4 interactions can escape from death and differentiate beyond the DP stage. This vital signaling event is termed *positive selection*. While TCR-MHC class II interactions (MHC class II-restriction) requires CD4 coreceptor proteins, CD8 coreceptors are needed for TCR-MHC class I (MHC class I-restriction) interactions to promote the signaling cascade.

Cellular signals, environmental cues and transcription factors involved in the expression of one or the other coreceptors in the process have extensively been studied for more than 25 years. All classical models share a set of fundamental principles: *i*) positive selection and fate decision are simultaneous events induced by the same TCR signaling cascades, *ii*) termination of one or the other coreceptor is irreversible and indicates commitment to the opposite coreceptor's lineage [6]. Contrary to these principles, with the discovery of *helper-deficient (HD)* mice, a specific strain with exclusive deficiency of CD4 SP helper lineage T cells [7], an intermediate stage (phenotypically $CD4^{+}CD8^{low}$ and transcriptionally $Cd4^{+}Cd8^{-}$) in

which, they initially terminate transcription of CD8 coreceptor proteins even when they are maturing into CD8 SP T cells, was identified [8].

Then, *kinetic signaling model* was proposed [8]. In the model the ultimate lineage choice of positively selected DP (Intermediate) thymocytes is determined by duration of TCR signals and exposure of the thymocytes to IL-7 cytokines. In addition, thymocytes at intermediate stage are defined as the last common progenitors of both CD4 SP helper and CD8 SP cytotoxic T cells in the new model. (Further discussions about the kinetic signaling model is given in Section 3.2.2.).

3.2.1 Classical models of CD4/CD8 lineage choice

Despite the experimental fails several times [6, 8, 9], the two classical models of CD4/CD8 lineage choice are *stochastic selection model* and *signal instructive models*. There is a striking concordance between the specificity of TCR expressed on the surface of T cells and the type of coreceptor expressed by T cells. These models were proposed to explain the mechanism of this concordance. The reader can refer to Singer et al. (2008) for more information about *classical models* of T cell differentiation.

3.2.1.1 Stochastic selection model

According to stochastic selection model, if DP thymocytes receive a signal through a TCR interacting with either an MHC class I or an MHC class II molecule, they randomly terminate expression of one or the other coreceptor with half probabilities. Then, only thymocytes continuing to express coreceptors matching with the MHC specificity of their TCRs can survive and differentiate into mature T cells. The remaining 'mismatched' thymocytes, that have a TCR specific to MHC class I but express CD4 coreceptor or express MHC class II specific TCR but express CD8 coreceptor, die by apoptosis [6].

3.2.1.2 Signal instructive models

Instructive models propose that engagement of TCR by MHC class I or MHC class II ligands results in qualitatively (duration of signal) or quantitatively (strength of signal)

distinct TCR signals that directly dictate the lineage choice of a positively selected thymocyte [9].

i) Strength of signal instructional model : This model postulates that engagement of TCR by MHC class I or MHC class II ligands leads to quantitatively weaker or stronger TCR signals that directly promotes differentiation of DP thymocytes to CD8 SP or CD4 SP lineages, respectively. The differences in TCR signaling strength are surmised to be caused by weaker or stronger affinity of the cytosolic tails of CD8 and CD4 for the key TCR signaling factor LCK, respectively (as cited in [9]).

ii) Duration of signal instructional model : This model implies that engagement of TCR by an MHC class II ligand results with a signal of long duration while engagement of TCR by an MHC class I ligand leads to a signal of shorter duration, and these signals instruct differentiation of DP thymocytes into $CD8^+$ and $CD4^+$ lineages, respectively (as cited in [6]).

3.2.2 From today's perspective: Kinetic signaling model

Kinetic signaling model is widely accepted to give the best explanation of CD4/CD8 lineage choice today. This model incorporates some unrefuted principles of the classical models and new premises based on more recent experimental observations. In kinetic signaling model, TCR signals of long duration may drive differentiation into CD4 SP lineage while TCR signals with shorter duration lead to CD8 SP differentiation pathway. If TCR signals cease at the Intermediate stage, IL-7R can receive IL-7 cytokines promoting to differentiation into $CD8^+$ T cells and thus inducing *coreceptor reversal* (as cited in [6]). Since in all positively selected thymocytes the production of CD8 coreceptor proteins is decreased, CD8-dependent MHC class I-restricted TCR signals may cease in time leading to derepression of IL-7 signaling that induces *coreceptor reversal* [8]. On the other hand, continuing expression of CD4 proteins at $CD4^+CD8^{low}$ stage yield persistent MHC class II-restricted TCR signaling and thus result in $CD4^+$ lineage choice.

In the kinetic signaling model, positive selection and lineage commitment are sequential events rather than being induced simultaneously by the same TCR signals

and the last bipotent precursors regarding in developmental order are intermediate thymocytes in which *Cd8* gene is transcriptionally terminated.

4. MATHEMATICAL AND NETWORK MODELING OF BIOLOGICAL REGULATORY SYSTEMS

4.1 Introduction

Proteins, encoded by genes, function as transcription factors that can bind to regulatory sites of genes, as enzymes catalyzing metabolic reactions, or as components of signal transduction pathways. In an organism, with minor exceptions, all cells contain the same genetic material. This means that, distinct functions of cells in an organism are attributed by genetic regulatory programs determining which genes are expressed, when and where in the organism, and to which extent. Such genetic regulatory programs are essentially structured by networks of regulatory interactions between DNA, RNA, proteins and small molecules [10].

As being core units of life, cells determine their behaviors like growth, move, proliferation, differentiation, etc. through such regulatory networks usually forced by environmental cues. The study of genetic regulatory systems has received a major impetus from the recent development of experimental techniques by which spatio-temporal expression levels of genes to be measured (as cited in [10]). Together with these still developing high throughput experimental tools, it is indispensable to employ theoretical models and computer simulations for eliciting structure and dynamics of genetic regulatory networks. Especially in health sciences, the quantitative models supported by recent improvements of single cell/molecule experimentation techniques would lead to much more reliable predictions on dynamics of real world problems, in particular encountered in health sciences.

Although *ordinary and partial differential equations* are the most conventional mathematical tools to investigate the genetic regulatory networks, *Boolean networks*, and *stochastic master equations* are some other formalims. The *directed graph* technique is, on the other hand, a visual represention of network models.

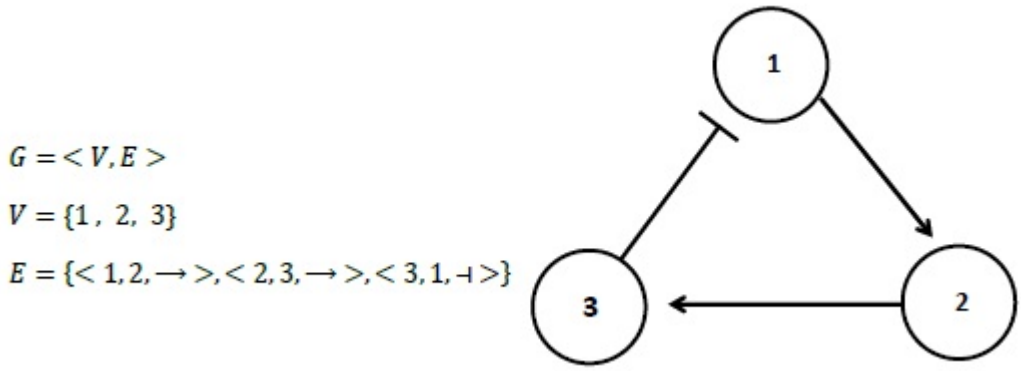

Figure 4.1: Directed graph of a representative regulatory network.

4.1.1 Directed graphs

The simplest way to represent a genetic regulatory network is with a *directed graph*. Such graphs can make biologically relevant predictions about behavior of regulatory systems by applying a number of operations on them. For example, a search for paths between two components may reveal missing regulatory interactions among them or an ignorance of a component or a link may provide clues about its redundancy in the network [10].

A directed graph G is a tuple $\langle V, E \rangle$ of a set of nodes (V) and a set of edges (E). A directed edge is also represented as a tuple $\langle i, j \rangle$ of vertices, where i denotes the head and j the tail of the edge. The nodes in a directed graph may correspond to genes or any other elements of interest in the regulatory system, while the edges represent interactions among them. Defining a directed edge as a tuple $\langle i, j, s \rangle$, with s equal to + or -, denotes whether i is activated or inhibited by j. For activation/inhibiton, the frequent choice is $\rightarrow/\dashv$ [10]. In Fig.4.1 a directed graph representation of a simple regulatory network of three genes is shown.

4.1.2 Boolean networks

The activation state of a gene or any other element in a regulatory system, termed as a *node*, can be approximated by a Boolean variable[1] which is defined as active (*on*, 1) or

[1]Boolean logic is a binary calculus of truth values, named after George Boole who first developed this algebra in the 1840s. It essentially operates based on logical operations *conjunction* ($\vee$), *disjunction* ($\wedge$), and *negation* ($\dashv$). Possible values of variables are conventionally represented by " 0 and 1" to sake for computational simplicity.

inactive (*off*, 0). For instance, a gene encodes its specific product when it is at 'on-state' while there is no production when the gene is at the 'off-state' . Interactions between nodes can be represented by Boolean functions (rules) which are specifically written for each individual nodes. Let the vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ represent the state of a regulatory system taken as a Boolean network of n elements. Since each x_i can take one of the two possible values, the state space of the system consists of 2^n different states. A graph depicting the possible states of the regulatory system and transitions between them is referred to as *state transition graph* and is useful to represent the dynamics of the system. As an example, Fig.4.2.b shows the state transition graph for the network given by Fig.4.2.a (Here, A, B and C are three elements having regulatory effects on each other.). According to the defined Boolean rules, the system always converges to the state (000) regardless of its initial state.

Boolean formalism is discrete both in space and time. The state of a node x_i at time $t+1$ is computed based on the state of the entire network at time t as given by (**4.1**),

$$x_i(t+1) = f_i(\mathbf{x}(t)), \qquad i = 1, 2, \ldots, n \tag{**4.1**}$$

When all nodes in a Boolean network are simultaneously updated, it is referred to *synchronous* updating that characterizes a fully deterministic dynamics for the system: each Boolean state of the system will always converge to a single steady state (named a *point attractor*) or steady cycle (*dynamic attractor*) through only a single *trajectory*. In the biological context of cell differentiation, these end-points correspond to the mature cell types [11]. In the state space of the system, the states which are not part of an attractor are called as *transient states*. An attractor and the transient states leading to the attractor form together a *basin of attraction* as sketched in the landscape picture shown in Fig.4.3. Such landscape pictures aim to depict different states of a cell by different positions on a two dimensional plane. The third dimension corresponds to the (free) energy of a thermodynamical system for which lower positions refer to more stable states for the cell. In fact, the depressions indicate stable solutions to the set of mathematical equations that describe the dynamics of the system. In contrast, when updating the system *asynchronously* (as cited in [11]), only the state value of a single

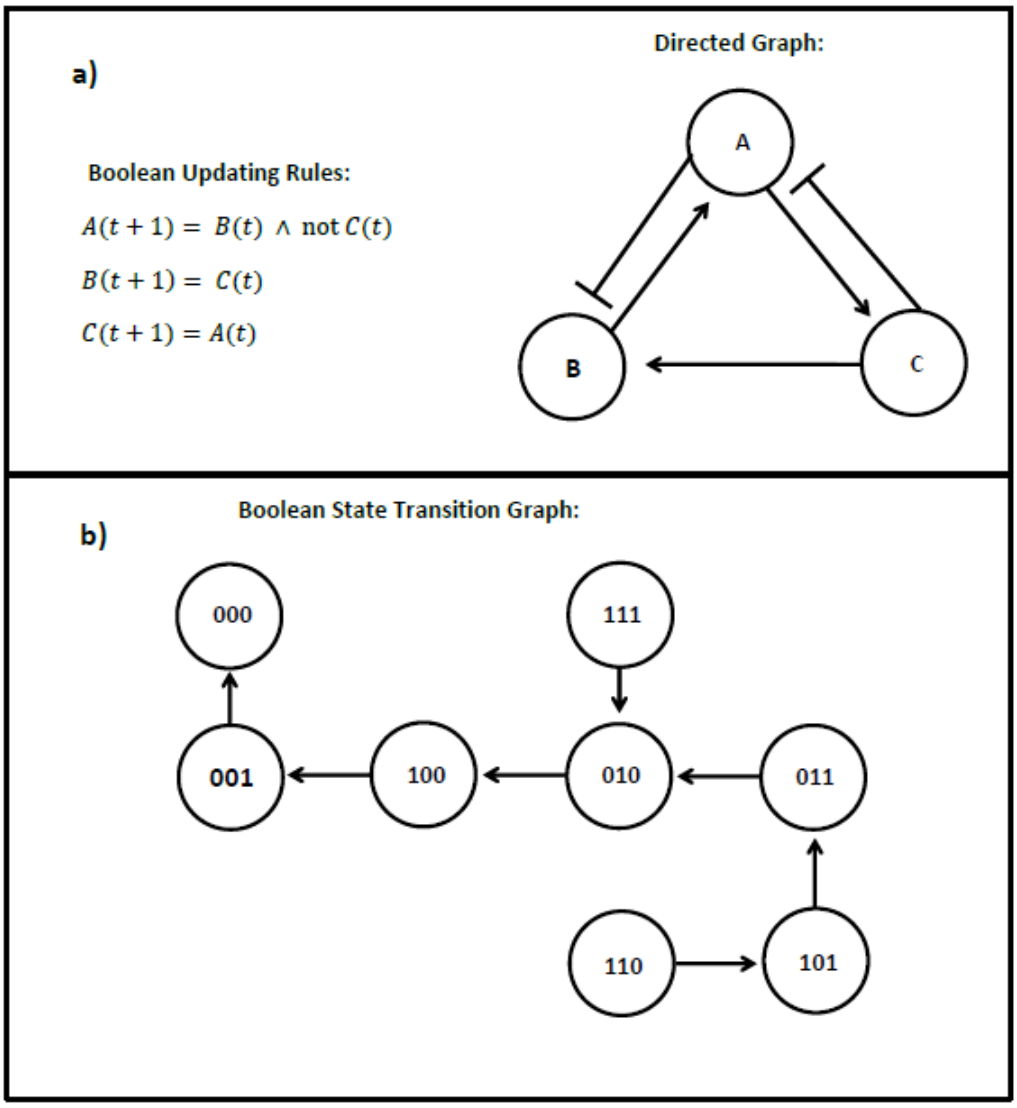

Figure 4.2: A simple regulatory network and its state transition graph under synchronous Boolean updating.

element is changed at each step. In this case, multiple trajectories following the same initial state are possible.

Although the Boolean formalism cannot mimic continuous changes of concentrations of elements of a regulatory network or give time information when regulatory events occur, it allows one to investigate easily the functional capabilities of the system without knowledge of any kinetic parameters even for very large networks and provide only a coarse-grained description of the network behavior.

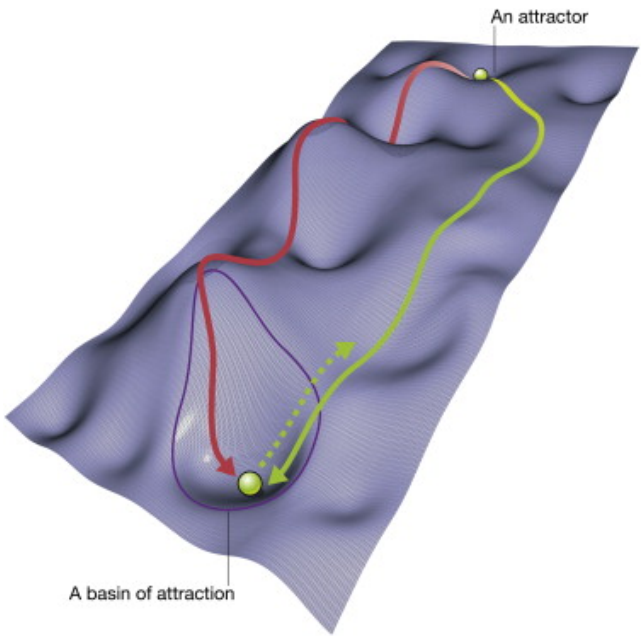

Figure 4.3: Landscape picture of cell differentiation [3].

4.1.3 Nonlinear ordinary differential equations

Ordinary differential equation (ODE) formalism is a widespread modeling tool for studying dynamical systems in the real world. Using ODE formalism allows one to model concentrations of regulatory elements such as RNA, proteins, etc. in gene regulatory networks using variables which are time-dependent and non-negative real numbers. As being essentially a biochemical process, gene regulation is defined by *rate equations* giving the rate of production of any element of the system as a function of current state of the entire system at any time, more specifically current concentrations of the regulatory inputs to the element. In mathematical terms, the rate equation for concentration value of node i at time t_0 is given by (**4.2**)

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = f_i(\mathbf{x}(t_0)), \qquad i = 1,\ 2,\ \ldots\ ,\ n, \tag{4.2}$$

and its concentration value at a later time t_1 is calculated by (**4.3**)

$$x_i(t_1) = x_i(t_0) + \int_{t_0}^{t_1} \frac{\mathrm{d}x_i}{\mathrm{d}t}\,\mathrm{d}t, \qquad i = 1,\ 2,\ \ldots\ ,\ n, \tag{4.3}$$

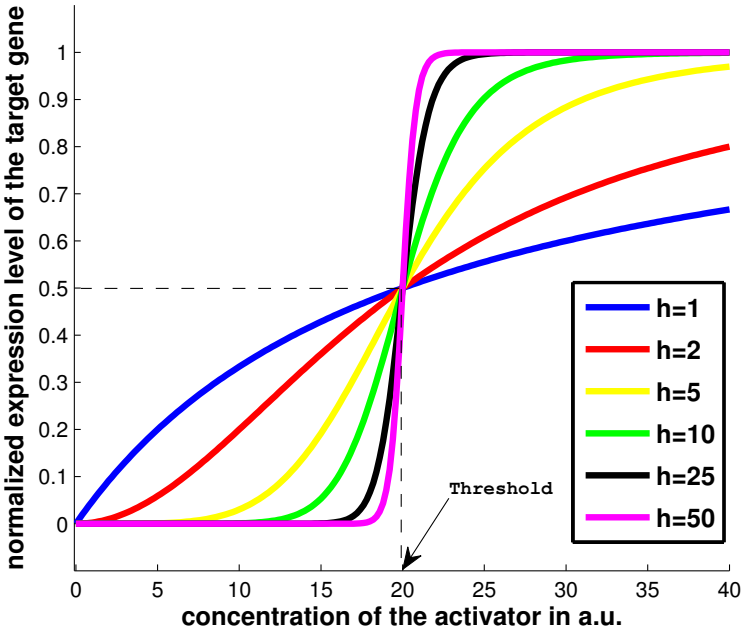

Figure 4.4: Activatory Hill regulation function

where $\mathbf{x} = [x_1, \ \ldots \ , \ x_n]' \geq \mathbf{0}$ is the state vector of the entire system consisting of concentrations of each elements in the regulatory network and regulation functions f_i' s are usually nonlinear functions of the state variables.

One of the most used form of the nonlinear functions for studying gene regulation is *Hill function*. If x_j is an activator of any target gene, the corresponding Hill regulation function is then defined by (**4.4**)

$$H^+(x_j, V_j, h) = \frac{x_j^h}{x_j^h + V_j^h}, \tag{4.4}$$

with $V_j > 0$, the *threshold* for the regulatory influence of x_j on a target gene, and $h > 0$, *steepness* parameter of the response of the target gene.

This function can take values varying in a continuous interval of $[0, 1]$ and increases as $x_j \to \infty$, so that an increase in x_j is reflected as an increase in the expression level of the target gene (*activation*) (See Fig.4.4). In order to express that increasing x_j decreases the expression level of the target gene (*inhibition*), the regulation function $H^+(x_j, V_j, h)$ is subsitituted by $H^-(x_j, V_j, h) = 1 - H^+(x_j, V_j, h)$. For $h > 1$, Hill curves have a sigmoid shape, in agreement with experimental evidence (as cited in [10]).

Here, for larger values of h response curve becomes step-like making the variables of the system Boolean-like at equilibrium states.

4.1.3.1 Standardized qualitative dynamical systems method

The Standardized Qualitative Dynamical Systems Method (SQDSM) was developed by Mendoza, L. and Xenarios, I. in 2006 [12]. It is essentially a nonlinear ODE modeling method that functions basically with the same approximations of Hill functions. The method has the ability to deterministicaly compute time evolution of a given regulatory network. In this method, the state variable, x_i, of an element at any time is determined by total input, ω_i, to it at previous time. The mathematical definition of the method is given by (**4.5**),

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \frac{-e^{0.5h} + e^{-h(\omega_i - 0.5)}}{(1 - e^{0.5h})(1 + e^{-h(\omega_i - 0.5)})} - \gamma_i x_i, \qquad i = 1,\ 2,\ \ldots\ ,\ n \qquad \textbf{(4.5)}$$

$$\omega_i = \begin{cases} \left(\dfrac{1+\sum_p \alpha_{pi}}{\sum_p \alpha_{pi}}\right)\left(\dfrac{\sum_p \alpha_{pi} x_{pi}^{ac}}{1+\sum_p \alpha_{pi} x_{pi}^{ac}}\right)\left(1 - \left(\dfrac{1+\sum_m \beta_{mi}}{\sum_m \beta_{mi}}\right)\left(\dfrac{\sum_m \beta_{mi} x_{mi}^{in}}{1+\sum_m \beta_{mi} x_{mi}^{in}}\right)\right) & \text{(a)} \\ \left(\dfrac{1+\sum_p \alpha_{pi}}{\sum_p \alpha_{pi}}\right)\left(\dfrac{\sum_p \alpha_{pi} x_{pi}^{ac}}{1+\sum_p \alpha_{p,} x_{p,}^{ac}}\right) & \text{(b)} \\ \left(1 - \left(\dfrac{1+\sum_m \beta_{mi}}{\sum_m \beta_{mi}}\right)\left(\dfrac{\sum_m \beta_{mi} x_{mi}^{in}}{1+\sum_m \beta_{mi} x_{mi}^{in}}\right)\right) & \text{(c)} \end{cases}$$

$$0 \le x_i \le 1 \quad 0 \le \omega_i \le 1 \quad h, \alpha_{pi}, \beta_{mi} > 0 \quad \gamma_i \ge 1$$

where $\{x_p^{ac}\}$ is the set of positive regulators acting on x_i, $\{x_m^{in}\}$ is the set of negative regulators of x_i. (a) is used if x_i has both positive regulators and negative regulators, (b) is used if x_i has only positive regulators, and (c) is used if x_i has only negative regulators. Finally, if x_i has no regulatory inputs then ω_i is taken as 0.

SQDSM requires specification of several parameters; strengths of activations ($\alpha'\ s$), strengths of inhibitions ($\beta'\ s$), decay rates ($\gamma'\ s$), and steepness of response curves ($h'\ s$). To keep x_i' s in the closed interval $[0,1]$, α's, β's and h are taken as any positive real numbers and γ's are taken as greater than or equal to 1. In the method, the decay rate of an element causes to inactivation it sooner or later, unless it has an activator. This is valid even if the corresponding element has no inhibitors.

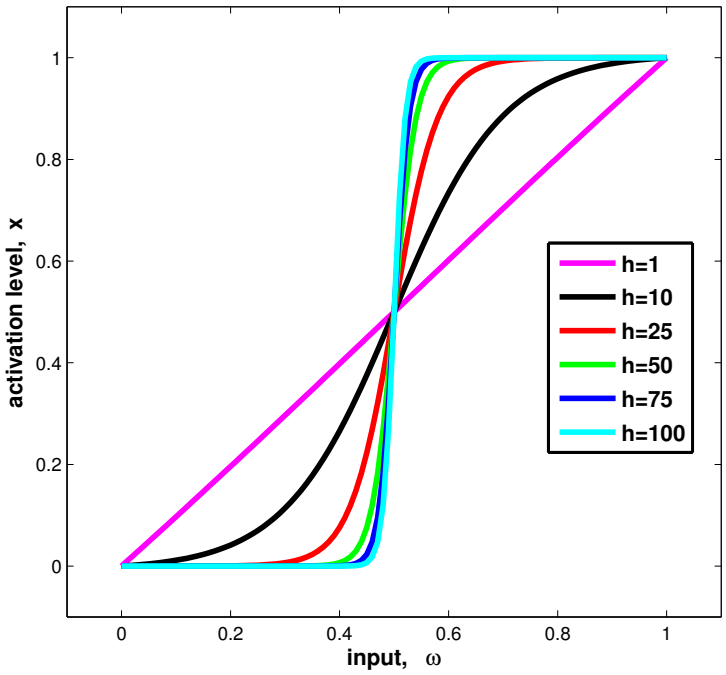

Figure 4.5: The change in the activation value of an element with respect to total input to it under different h choices.

As shown in Fig.4.5 and also pointed out in the context of Hill functions, larger the value of h steeper the response curve. Therefore, SQDSM operates Boolean-like for large values of h giving digital response curves and thus making the equilibrium solutions comparable with the ones obtained from (synchronous) Boolean studies of the same structure.

The total regulatory input to an element due to different strengths of activations by a single activator and inhibitions by a single inhibitor are shown in Fig.4.6 and Fig.4.7, respectively. As it can be easily seen, the total input to the element having only one activator becomes more sharply increasing when the strength of the activation (alpha) is increased. In the case of the element having only one inhibitor, the total input to it becomes more sharply decreasing by increasing the inhibition strength (beta).

The change in the activation level of an element due to the effect of a single activatory input by choosing different activation strengths (alpha) and a single inhibitory input having different inhibition strengths (beta) are depicted in Fig.4.8 and Fig.4.9, respectively. The activation level of the target element becomes more digitally

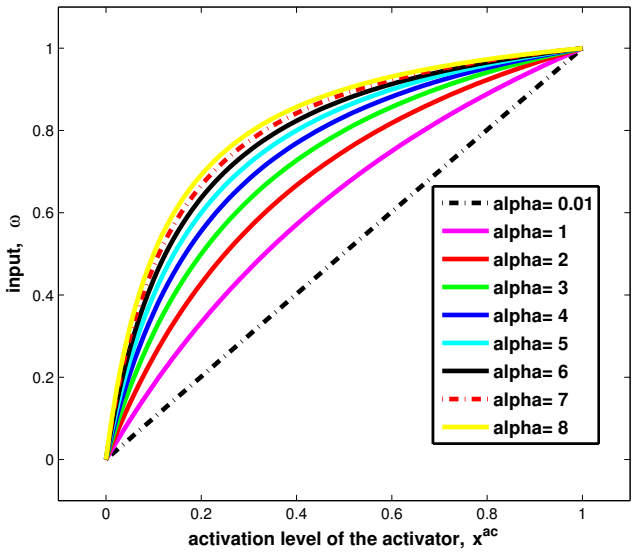

Figure 4.6: The total input to an element having only one activator as a function of the activation level of the activator with different activation strengths.

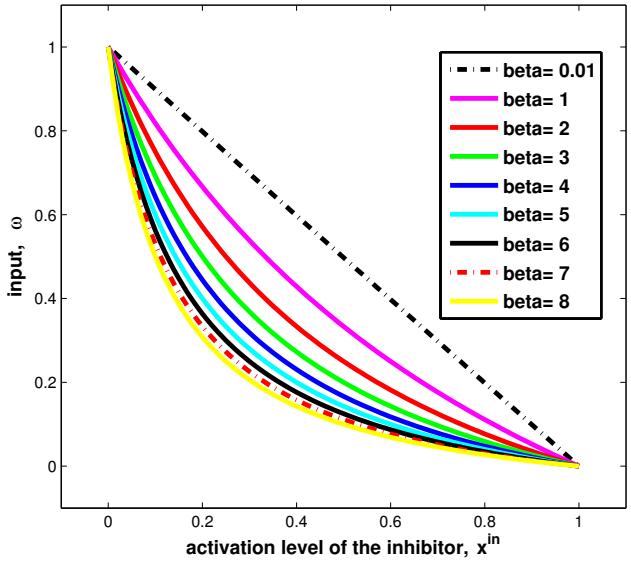

Figure 4.7: The total input to an element having only one inhibitor as a function of the activation level of the inhibitor with different inhibition strengths.

regulated when the strength of the regulation due to the corresponding regulator is increased (for $h = 10$).

The change in the total input and activation level value of an element in the case of co-existence of an activator and an inhibitor acting on it are shown in Fig.4.10 and

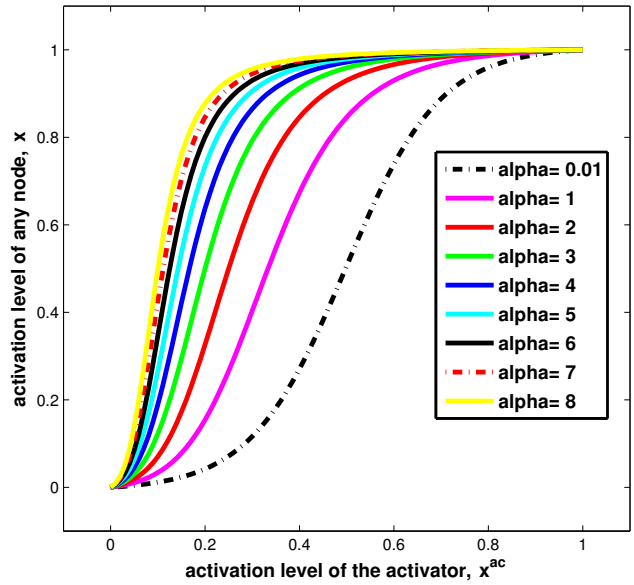

Figure 4.8: The activation level value of an element having only one activator as a function of the activation level of the activator with different activation strengths.

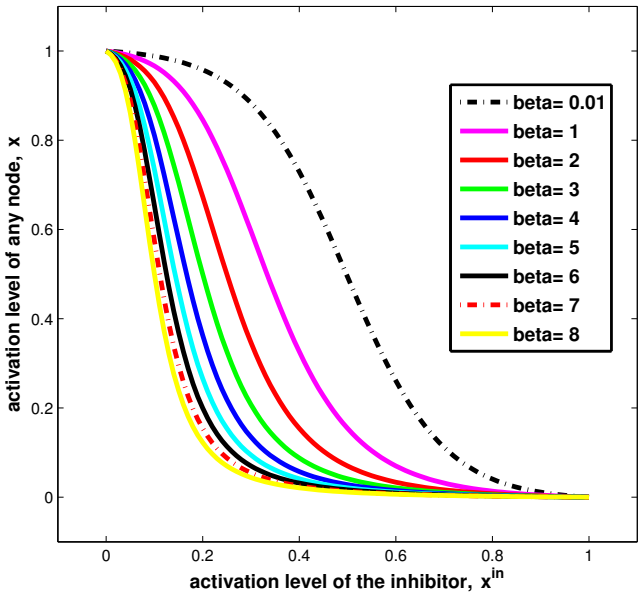

Figure 4.9: The activation level value of an element having only one inhibitor as a function of the activation level of the inhibitor with different inhibition strengths.

in Fig.4.11, respectively. As it is clearly seen from these figures, the target element can only become activated if the inhibitor is not at its maximum level of activation.

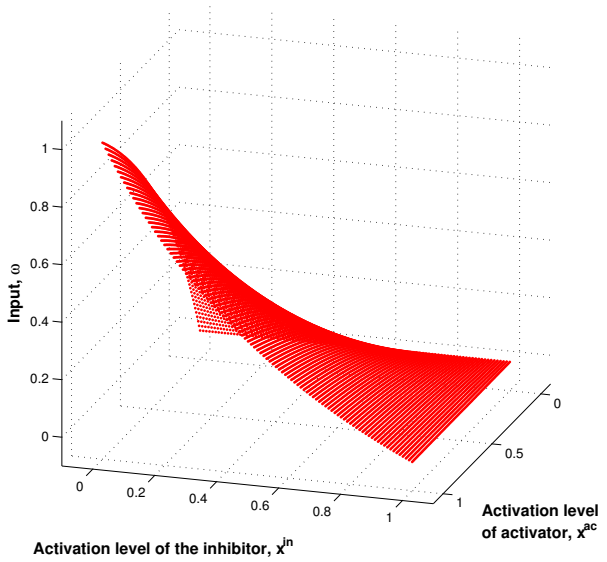

Figure 4.10: Total regulatory input to any node which has an activatory and an inhibitory element as well.

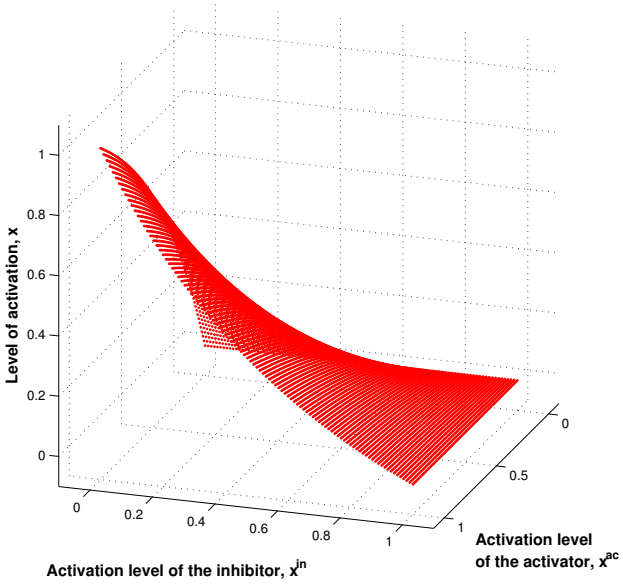

Figure 4.11: Activation level value of any node which has an activatory and an inhibitory element as well ($h = 1$).

It can be also seen that the target element can be fully activated when the inhibitor is inactivated and the activator is simultaneously at its maximum level of activation.

Although analytical solutions to the rate equations (**4.2**) are not generally possible due to the nonlinearity of f_i' s, at least some qualitative properties of the solutions such as the number, pattern and the stability of the steady states (*fixed points*) can be established using numerical and computational tools.

A big trouble challenging the numerical analysis of regulatory systems by utilizing ODEs is the lack of measurements of kinetic parameters for the rate equations. Therefore, in most cases the parametric values had to chosen such that the models are able to reproduce the observed qualitative behavior. For almost last fifteen years, availability of gene expression measurements have increasingly grown allowing the required kinetic data to be extracted from time series data of state variables of a system of interest. A probable interesting property of network models, termed *robustness*, states that the network structure determines the stability character of the system rather than the exact values of the parameters when essential properties of the system quite unresponsive to variations in the parametric values [10].

4.1.3.2 Linear stability analysis of ODE systems

Let each x_i stands for the state variable for element i of a system with $i = 1,\ 2,\ \ldots\ n$, and their changes with time are computed with f_i functions as given by (**4.2**). If state of entire system is represented by $\mathbf{x} = [x_1,\ \ldots\ ,\ x_n]' \geq \mathbf{0}$ and $f_i(\mathbf{x}_*) = 0$ for each i, then $\mathbf{x}_*$ is called a fixed point of the system.

Now let $\vec{\eta}(t) = \mathbf{x}(t) - \mathbf{x}_*$ be a small perturbation away from $\mathbf{x}_*$ at time t. Time evolution of this $\vec{\eta}(t)$ determines the stability character of the fixed point $\mathbf{x}_*$. Thus time derivative of $\vec{\eta}$ is required to be calculated and it is essentially same as with the calculation of time derivative of $\mathbf{x}$ at any time, as given by (**4.6**)

$$\dot{\vec{\eta}} = \frac{\mathrm{d}}{\mathrm{d}t}(\vec{\eta}) = \frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{x} - \mathbf{x}_*) = \dot{\mathbf{x}} \tag{4.6}$$

(here $\mathbf{x}_*$ is constant). Thus $\dot{\vec{\eta}} = \dot{\mathbf{x}} = f(\mathbf{x}) = f(\mathbf{x}_* + \vec{\eta})$. For $n = 1$, using Taylor's expansion we obtain

$$f(x_* + \eta) = f(x_*) + \eta f'(x_*) + O(\eta^2), \tag{4.7}$$

where $O(\eta^2)$ denotes quadratically small terms in η and $f'(x_*)$ stands for $\left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x=x_*}$ which is the derivative with respect to state variable x evaluated at x_*. Since x_* is a fixed point, $f(x_*)=0$. Hence

$$f(x_*+\eta)=\eta f'(x_*)+O(\eta^2). \tag{4.8}$$

Now if $f'(x_*)\neq 0$, the $O(\eta^2)$ terms are negligible and it is conceivable to make the approximation

$$\dot{\eta}=\eta f'(x_*). \tag{4.9}$$

This is a linear equation in the perturbation η, and is called the *linearization about a fixed point* x_*. Since we are only keeping linear terms near a fixed point, this theorem is called a *linear stability analysis*. In fact, $f'(x_*)$ is the slope at the fixed point and determines its stability. The perturbation $\eta(t)$ grows exponentially if $f'(x_*)>0$ and decays if $f'(x_*)<0$, entitling the fixed point x_* as *unstable* and *stable*, respectively . If $f'(x_*)=0$, the $O(\eta^2)$ terms are not negligible and a nonlinear phase portrait analysis is needed to determine the stability. The absolute value of $f'(x_*)$, gives the measure of how much a fixed point is stable. Inverse of this value, $1/|f'(x_*)|$, is referred as *characteristic time scale* determining the time required for $x(t)$ to vary significantly in the neighborhood of x_* [13].

For $n=2$, the dynamics of the system is defined by the following coupled equations

$$\begin{aligned}\frac{\mathrm{d}x_1}{\mathrm{d}t}&=f_1(\mathbf{x})\\ \frac{\mathrm{d}x_2}{\mathrm{d}t}&=f_2(\mathbf{x})\end{aligned}$$

where $\mathbf{x}=[x_1,\ x_2]'\geq\mathbf{0}$ is the state vector of the system at any time, and f_1 and f_2 are nonlinear regulatory functions. Let $\mathbf{x}_*=[x_{*,1},\ x_{*,2}]'$ be a fixed point and $\vec{\eta}(t)=\mathbf{x}(t)-\mathbf{x}_*$ be a small perturbation about it. Then by following the same procedure in (**4.6**), (**4.7**), (**4.8**) and (**4.9**), we obtain

$$\dot{x}_1 = (x_1 - x_{*,1}) \frac{\partial f_1}{\partial x_1}\bigg|_{\mathbf{x}=\mathbf{x}_*} + (x_2 - x_{*,2}) \frac{\partial f_1}{\partial x_2}\bigg|_{\mathbf{x}=\mathbf{x}_*},$$
$$\dot{x}_2 = (x_1 - x_{*,1}) \frac{\partial f_2}{\partial x_1}\bigg|_{\mathbf{x}=\mathbf{x}_*} + (x_2 - x_{*,2}) \frac{\partial f_2}{\partial x_2}\bigg|_{\mathbf{x}=\mathbf{x}_*}.$$

Since it is possible to write the above equations as

$$\dot{\eta}_1 = \eta_1 \frac{\partial f_1}{\partial x_1}\bigg|_{\mathbf{x}=\mathbf{x}_*} + \eta_2 \frac{\partial f_1}{\partial x_2}\bigg|_{\mathbf{x}=\mathbf{x}_*},$$
$$\dot{\eta}_2 = \eta_1 \frac{\partial f_2}{\partial x_1}\bigg|_{\mathbf{x}=\mathbf{x}_*} + \eta_2 \frac{\partial f_2}{\partial x_2}\bigg|_{\mathbf{x}=\mathbf{x}_*},$$

we can use matrix notation as follows

$$\begin{pmatrix} \dot{\eta}_1 \\ \dot{\eta}_2 \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{pmatrix}_{\mathbf{x}=\mathbf{x}_*} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}.$$

The matrix $J = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{pmatrix}_{\mathbf{x}=\mathbf{x}_*}$ is called the *Jacobian matrix* calculated at the fixed point $\mathbf{x}_*$.

Now to determine whether the perturbation $\vec{\eta}$ grows or decays in time, we must find the *eigenvalues*, λ_1 and λ_2, and corresponding *eigenvectors*, $\vec{v_1}$ and $\vec{v_2}$, of J. Then, in theory of linear differential equations, the solution is written as the superposition of terms that are in form $exp(\lambda_i)$.

If trace $(J_{11} + J_{22})$ and determinant $(J_{11}J_{22} - J_{12}J_{21})$ of J are denoted by τ and Δ, respectively, then the eigenvalues can be calculated by (**4.10**)

$$\lambda_{1,2} = \frac{1}{2}\left(\tau \mp \sqrt{\tau^2 - 4\Delta}\right). \tag{**4.10**}$$

Let $\lambda_i = a_i + jb_i$, where a_i and b_i are, respectively, the *real* $[Re(\lambda_i)]$ and *imaginary* $[Im(\lambda_i)]$ parts of the eigenvalue λ_i. Then the exponential terms are can be written as

$$exp(\lambda_i t) = exp(a_i t) exp(jb_i t), \qquad j = \sqrt{-1}. \tag{**4.11**}$$

The complex exponential can be rewritten

$$exp(jb_it) = \cos(b_it) + j\sin(b_it). \tag{4.12}$$

As it can easily be seen from (**4.12**), complex part of an eigenvalue contributes only an oscillationary component to the solution. It is the real part that matters:

If $a_i > 0$ for any i, $exp(a_it)$ grows with time, which indicates that trajectory of the small perturbation will tend to move away from the fixed point. This makes the fixed point unstable [14]. For $n = 2$, phase plane behavior of trajectories in the cases of different kinds of fixed points is schematically represented in Fig.4.12. In the figure, the stability character of fixed points is shown depending on the value of determinant (Δ, x-axis) and trace (τ, y-axis) of the Jacobian matrix.

To generalize for $n = N$, one may follow the same procedure in (**4.6**), (**4.7**), (**4.8**) and (**4.9**) and finally write the Jacobian at the fixed point as,

$$J = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_N} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_N}{\partial x_1} & \frac{\partial f_N}{\partial x_2} & \cdots & \frac{\partial f_N}{\partial x_N} \end{pmatrix}_{\mathbf{x}=\mathbf{x}_*}$$

If all the eigenvalues of the Jacobian matrix has real parts smaller than zero, the fixed point is then referred to as stable while the fixed point is called unstable when the Jacobian has at least one eigenvalue with a real part greater than zero. It is important to emphasize that the theorem is silent on the issue of what happens if some of the eigenvalues have zero real parts while the others are all negative. This can not be decided based on linear stability analysis. The nonlinear terms we left out in (**4.8**) in fact determine the stability in this case.

4.1.4 Stochastic master equations

Despite providing the great possibility of modeling gene regulation in a fine details even at a level of individual reaction steps, such as binding of a transcription factor

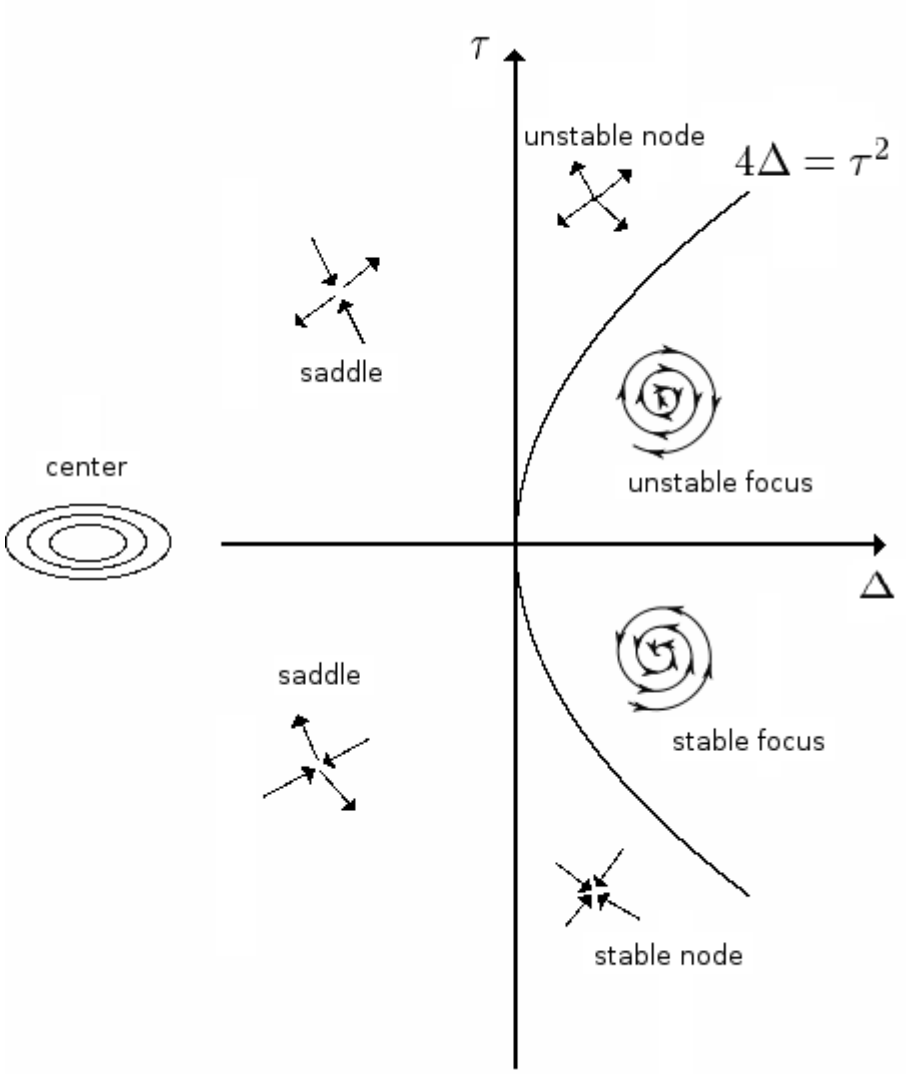

Figure 4.12: Stability of fixed points in 2-dimensional systems.

to its specific regulatory site, differential equation models assume the dynamics of the system of interest to be both *continuous* and *deterministic*. However, biochemical reactions involved in gene regulatory events are both *discrete* and *stochastic* in nature. Gene regulatory events are mainly processed inside the nucleus and so may be driven by only a few tens of molecules. Hence, small changes in the number of components become more important leading to a discrete state space for the system rather than a continuous one. Moreover, molecular interactions are not really deterministic meaning that all collisions between two molecules do not necessarily result in biochemical reactions. Gene regulatory events, therefore, have a probabilistic dynamics: that is, the same initial condition does not always end with the same steady state. Given the state vector $\mathbf{x} = (x_1, x_2, \ldots, , x_N)$ of a regulatory system of $S_1, S_2, \ldots, S_n$ different

species with x_i's that are positive integers, joint probability distribution function $p(\mathbf{x},t)$, describes the probability at a time t the cell contains x_1 molecules of S_1 species, x_2 molecules of S_2 species, etc. The time evolution of the function of $p(\mathbf{x},t)$ can be computed by

$$p(\mathbf{x},t+\Delta t) = p(\mathbf{x},t)\left(1-\sum_{j=1}^{m} a_j \Delta t\right) + \sum_{k=1}^{m} b_k \Delta t, \tag{4.13}$$

where m is the number of reactions that can occur in the system, $a_j\Delta t$ is the probability that reaction j will occur in the interval $[t,t+\Delta t]$ given that the system is in the state $\mathbf{x}$ at t, and $b_k\Delta t$, the probability that reaction k will bring the system to state $\mathbf{x}$ from another state in $[t,t+\Delta t]$ [15].

Rearranging (**4.13**),

$$p(\mathbf{x},t+\Delta t) - p(\mathbf{x},t) = -\sum_{j=1}^{m} p(\mathbf{x},t) a_j \Delta t + \sum_{k=1}^{m} b_k \Delta t$$

$$\frac{p(\mathbf{x},t+\Delta t) - p(\mathbf{x},t)}{\Delta t} = \sum_{j=1}^{m} \left[b_j - a_j p(\mathbf{x},t)\right]$$

and taking the limit $\Delta t \to 0$ where Δt allows only a single reaction, time evolution of probability function $p(\mathbf{x},t)$ is obtained from the definition of derivative as follows,

$$\frac{\mathrm{d}}{\mathrm{d}t} p(\mathbf{x},t) = \sum_{j=1}^{m} \left[b_j - a_j p(\mathbf{x},t)\right]. \tag{4.14}$$

This is the stochastic analogue of rate equation named as *master equation* (as cited in [10]). In general, it is almost impossible to find exact analytical solutions to master equations. Moreover, numerical simulations of the system is highly complicated and computationally expensive since there are $n+1$ independent variables: n discrete state variables and a continuous time variable t.

A feasible algorithm has been proposed by Gillespie in 1977 [15]. This algorithm computes stochastic time evolution of a biochemical system based on two fundamental questions: (*i*) *When will the next reaction occur?* and (*ii*) *Which type will it be among*

the possible reactions? Although Gillespie's algorithm can be easily implemented into a computer program, one needs to run the program for many times under a particular initial condition to obtain a probability distribution given by (**4.13**) since the algorithm simulates only a single trajectory for the system at each run. Gillespie's algorithm was used to simulate numerous biochemical processes in the literature and gave strikingly reliable outcomes, in particular when the system involves low numbers of molecules and strong competitive feedback loops which are often encountered in the study of cell signaling mechanisms [16]. Moreover, in order to reveal spatio-temporal dynamics of biochemical mechanisms, Gillespie's algorithm was extended by adding a reaction-diffusion term that answers a third question: *Where will the next reaction occur?* (as cited in [17]). For further details about Gillespie's algorithm the reader refer to [15, 17].

5. A REGULATORY NETWORK MODEL FOR DIFFERENTIATION OF $\alpha\beta$ THYMOCYTES BEYOND THE 'INTERMEDIATE' STAGE

In this thesis, we aim to build a mathematical model which qualitatively describes differentiation of $\alpha\beta$ thymocytes, particularly beyond the Intermediate stage, as a dynamical system. We first had an extensive literature search to determine the most important regulatory interactions between Intermediate stage and mature stage in T cell developmental pathway. The elements and the associated interactions involved in the process are presented in Table 5.1. Our model comprises a network of 8 components and 13 regulatory interactions as sketched in the directed graph shown in Fig.5.1. In the directed graph, we use the green and red arrows to represent activatory and inhibitory interactions, respectively. On the other hand, the black arrow denotes the co-activatory effect of GATA3 and ThPOK on CD4.

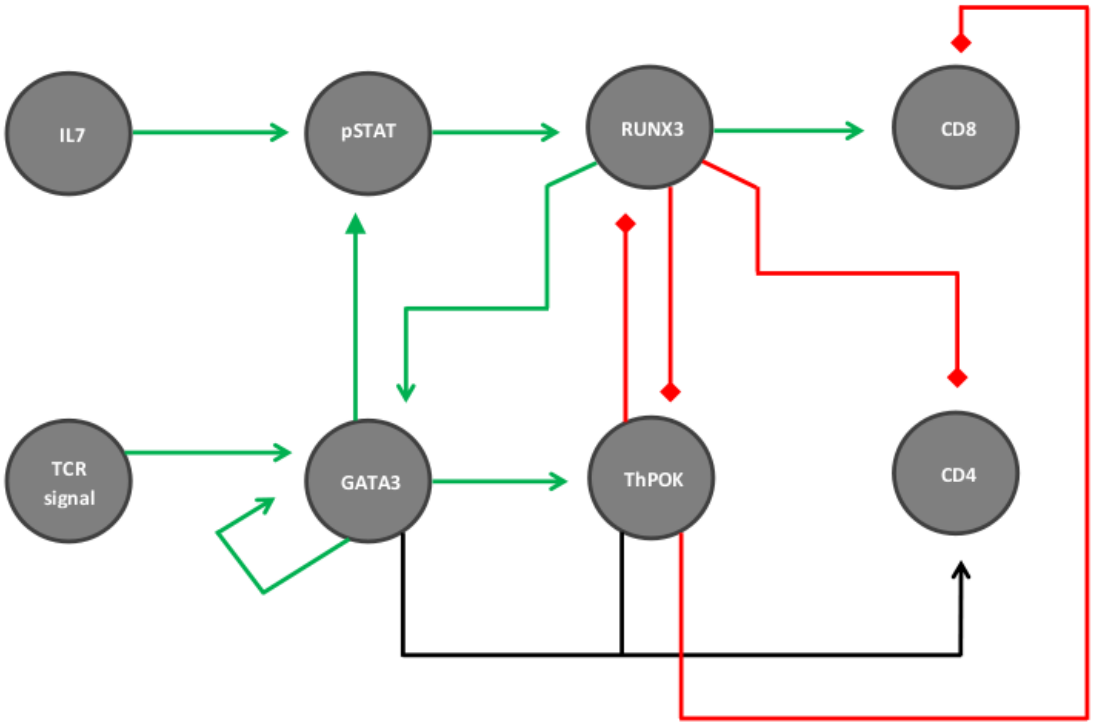

Figure 5.1: Directed graph of our regulatory network for the differentiation of $\alpha\beta$ thymocytes beyond the Intermediate ($CD4^{+}CD8^{low}$) stage.

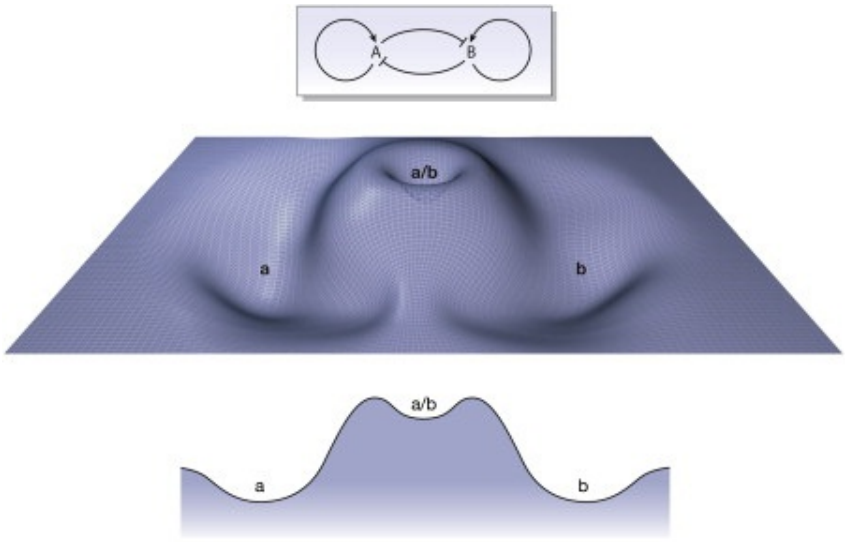

Figure 5.2: A landscape with three attractors is generated by mathematical modeling of a simple genetic regulatory network [3].

5.1 Methodology

In theoretical biology, the conventional technique in building a regulatory network model for a cell differentiation process is to define different attractors (*or* equilibrium states) in the landscape picture corresponding to different cell types. For instance, in the regulatory network illustrated at the top of Fig.5.2, two transcription factors A and B, mutually inhibit each other and autoactivate expression of themselves. Mathematical modeling of this network generates a landscape picture comprising three attractors: two stable states, a and b, in which factors A and B are exclusively expressed, and a metastable state, a/b, characterized by low coexpression of both factors A and B. In the context of cell differentiation, a precursor cell which occupies the metastable state a/b have the bipotency of moving to stable attractors either a or b, namely distinct mature cell subsets.

Despite the existence of a great amount of empirical information related the T cell differentiation in the thymus, obtaining kinetic data and real concentration values that will help to employ dynamical representations is still in its infancy. Therefore, we adopt SQDSM of Luis Mendoza and Ioannis Xenarios (2006) (Section 4.1.3.1) to convert our directed graph into a dynamical mathematical representation.

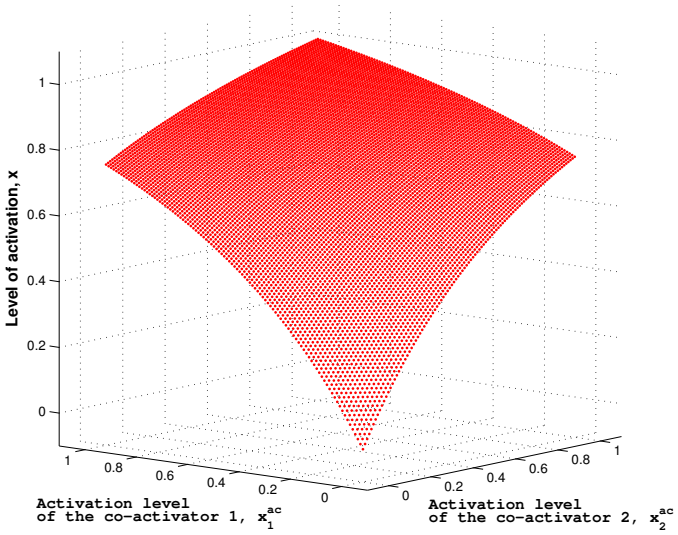

Figure 5.3: In SQDSM, the change in the activation level of an element having two co-activators with respect to the activation levels of the co-activators , x_1^{ac} and x_2^{ac} (with the strength $\eta = 1$).

In this method, the state of each element (termed *node*) in a regulatory network can be updated in time by taking into account the regulatory effects by the others and itself. But, in biology it is very ubiquitous that a regulatory event can only occur in the co-existence of two or more regulatory elements and the formalism fails to mimic such events. (As an example, the change in the activation level of an element having two co-activators as a function of the activation levels of the co-activators is illustrated in Fig.5.3.). With this in mind, we contribute to their formalism by adding co-regulatory (only second order) terms into the definition of the input function, ω, as follows

$$\omega_i = \begin{cases} \left(\dfrac{1+\sum_p \alpha_{pi}+\sum_k \eta_{ki}}{\sum_p \alpha_{pi}+\sum_k \eta_{ki}}\right)\left(\dfrac{\sum_p \alpha_{pi}x_p^{ac}+\sum_k \eta_{ki}x_{k,1}^{ac}x_{k,2}^{ac}}{1+\sum_p \alpha_{pi}x_p^{ac}+\sum_k \eta_{ki}x_{k,1}^{ac}x_{k,2}^{ac}}\right)\times \\ \left(1-\left(\dfrac{1+\sum_m \beta_{mi}+\sum_z \xi_{zi}}{\sum_m \beta_{mi}+\sum_z \xi_{zi}}\right)\left(\dfrac{\sum_m \beta_{mi}x_m^{in}+\sum_z \xi_{zi}x_{z,1}^{in}x_{z,2}^{in}}{1+\sum_m \beta_{mi}x_m^{in}+\sum_z \xi_{zi}x_{z,1}^{in}x_{z,2}^{in}}\right)\right) & \text{(a)} \\ \left(\dfrac{1+\sum_p \alpha_{pi}+\sum_k \eta_{ki}}{\sum_p \alpha_{pi}+\sum_k \eta_{ki}}\right)\left(\dfrac{\sum_p \alpha_{pi}x_p^{ac}+\sum_k \eta_{ki}x_{k,1}^{ac}x_{k,2}^{ac}}{1+\sum_p \alpha_{pi}x_p^{ac}+\sum_k \eta_{ki}x_{k,1,}^{ac}x_{k,2}^{ac}}\right) & \text{(b)} \\ \left(1-\left(\dfrac{1+\sum_m \beta_{mi}+\sum_z \xi_{zi}}{\sum_m \beta_{mi}+\sum_z \xi_{zi}}\right)\left(\dfrac{\sum_m \beta_{mi}x_m^{in}+\sum_z \xi_{zi}x_{z,1}^{in}x_{z,2}^{in}}{1+\sum_m \beta_{mi}x_m^{in}+\sum_z \xi_{zi}x_{z,1}^{in}x_{z,2}^{in}}\right)\right) & \text{(c)} \end{cases} \quad \textbf{(5.1)}$$

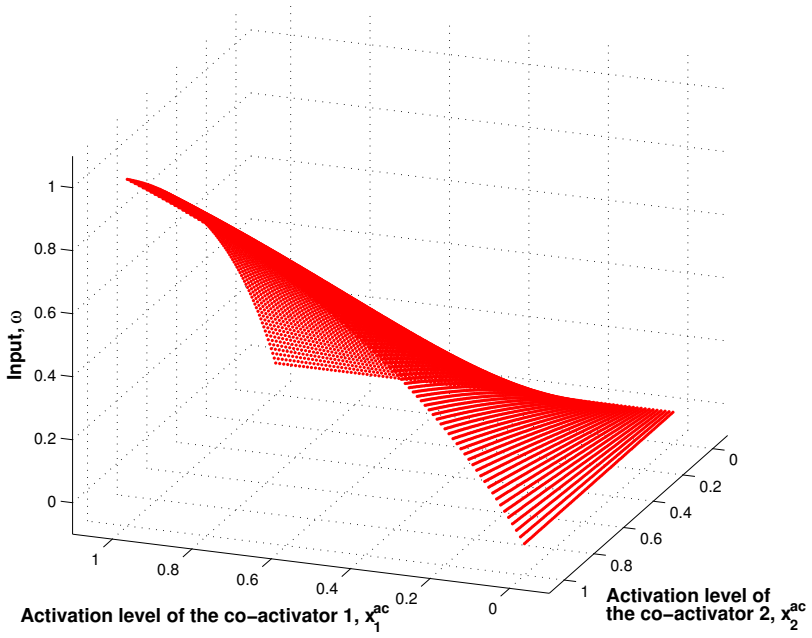

Figure 5.4: Input to any node due to effect of two co-activators, x_1^{ac} and x_2^{ac}, with the strength $\eta = 1$.

where $\{x_{k,1}^{ac} x_{k,2}^{ac}\}$ is the set of positive co-regulators acting on x_i, $\{x_{z,1}^{in} x_{z,2}^{in}\}$ is the set of negative co-regulators acting on x_i. The maximum numbers for indices k and z are defined by the numbers of second order positive and negative regulations, respectively, involved in the given network. Here, the new parameters η and ξ stand for strengths of co-regulatory (second order) events and they should be greater than zero. Here, (a) is used if x_i has both positive regulators and negative regulators, (b) is used if x_i has only positive regulators, and (c) is used if x_i has only negative regulators. If x_i has no regulatory inputs then ω_i is taken as 0. We have tested the modified version of the method to see if it still keeps the values of the variables in a normalized interval and provides trustworthy outcomes for modeling regulatory networks. The inputs and activation levels of any node with the two co-activator x_1^{ac}, x_2^{ac}, strength $\eta = 1$ and with the two inhibitors x_1^{in}, x_2^{in}, strength $\xi = 1$ are plotted in Fig.5.4, Fig.5.5, Fig.5.6 and Fig.5.7. As it is clearly seen in the figures the modified method still keeps the normalized values for the variables and allows one to simulate (second order) co-regulatory events.

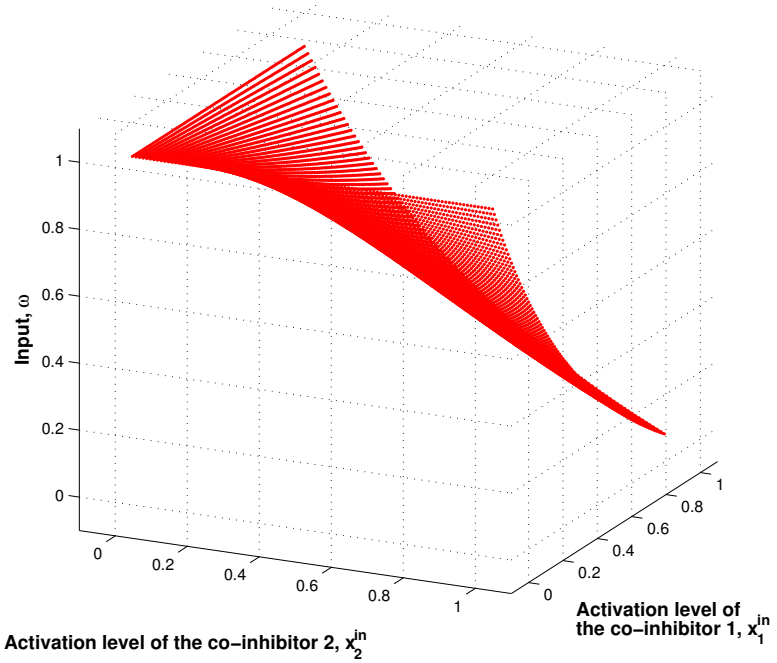

Figure 5.5: Input to any node due to effect of two co-inhibitors, x_1^{in} and x_2^{in}, with the strength $\xi = 1$.

5.2 Molecular description and construction of our regulatory network

First of all, in our model CD4 and CD8 coreceptor proteins are treated as the end-products (Fig.5.1). In our network model, there are two components (TCR signal and IL7) referring to strengths of intrathymic T cell antigen receptor (TCR) and IL7 cytokine signalings which are the essential driving factors of CD4/CD8 lineage choice for a maturing thymocyte (Section 3.2.2). GATA3, ThPOK, RUNX3, CD4 and CD8 represent expression level of respective genes and pSTAT corresponds to phosphorylation level of the STAT molecule. In the model, GATA3 and pSTAT are forming the bridges which convey the incoming stimuli by intrathymic signals to the inner regulatory mechanism. GATA3, ThPOK, RUNX3 and pSTAT act as transcription factors (TFs) since they produce proteins to fulfill particular functions for eventually regulating the activation levels of CD4/8 coreceptor expression. (Of note, we assume that the activation level of a gene has the same meaning with the amount of protein encoded by itself.).

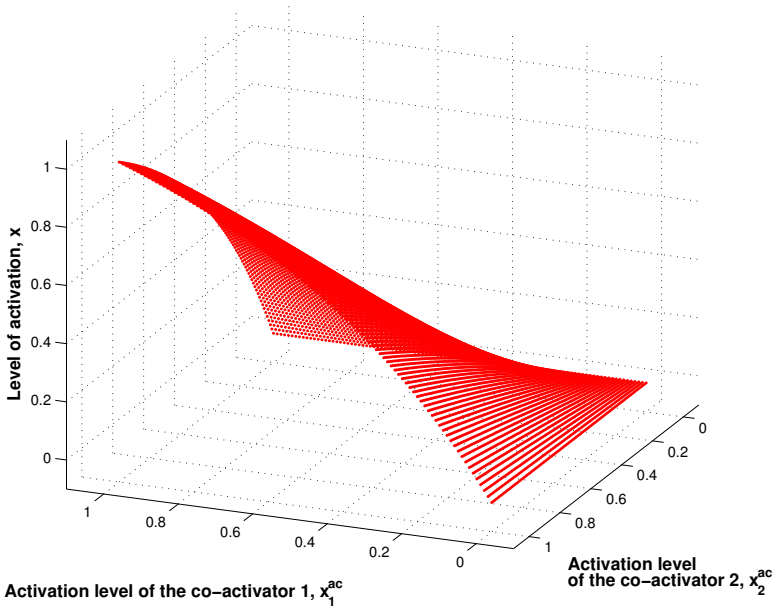

Figure 5.6: Level of activation of any node due to effect of two co-activators, x_1^{ac} and x_2^{ac}, with the strength $\eta = 1$ ($h = 1$).

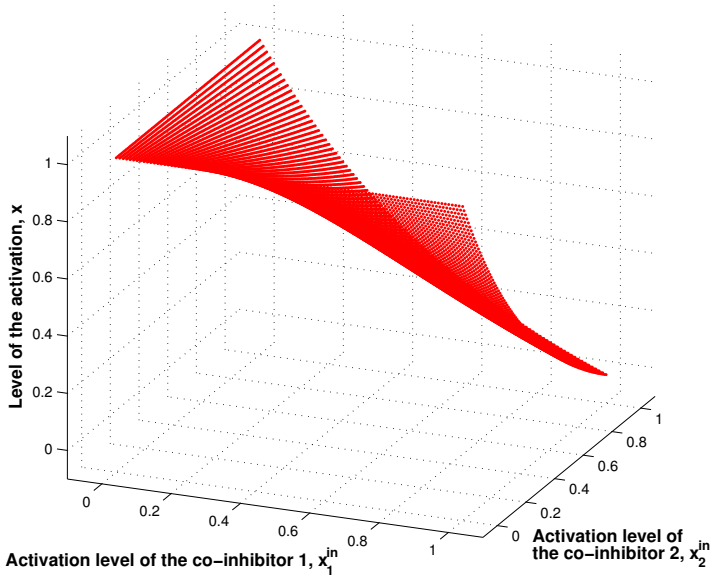

Figure 5.7: Level of activation of any node due to effect of two co-inhibitors, x_1^{in} and x_2^{in}, with the strength $\xi = 1$ ($h = 1$).

In the model, GATA3 forms a unique bridge between strength of TCR signaling and lineage decision because it is triggered/regulated by TCR signals [18], and it is the

Table 5.1: Regulatory interactions in our network.

Interaction	Reference(s)	Strength Parameter
TCR $\rightarrow$ GATA3	[18]	α_1
GATA3 $\rightarrow$ GATA3		α_2
RUNX3 $\rightarrow$ GATA3		α_3
GATA3 $\rightarrow$ ThPOK	[19]	α_4
RUNX3 $\dashv$ ThPOK	[23]	β_1
ThPOK $\dashv$ RUNX3	[20]	β_2
pSTAT $\rightarrow$ RUNX3	[25]	α_5
RUNX3 $\dashv$ CD4	[29, 30]	β_3
RUNX3 $\rightarrow$ CD8	[28]	α_6
GATA3 $\rightarrow$ pSTAT		α_7
GATA3 + ThPOK $\rightarrow$ CD4	[19]	η_1
IL7 $\rightarrow$ pSTAT	[25]	α_8
ThPOK $\dashv$ CD8	[20]	β_4

gateway to a lineage choice. It initiates expression of ThPOK [9, 19] starting from DP stage. ThPOK inhibits expression of CD8 and RUNX3 proteins [20] and activates CD4 expression co-operatively with GATA3 [19,21]. Expression of ThPOK is specific for CD4 SP (helper) lineage [22] and inhibited by RUNX3 proteins [23]. Expression of RUNX3 is first induced through IL7 signals, that cause to phosphorylation of STAT molecules allowing them to translocate inside the nucleus [24], in positively selected thymocytes [25, 26]. RUNX3 expression is specific for CD8 SP (cytotoxic) cells [27], and increased through the CD8 differentiation program while it is excluded through the CD4 SP differentiation pathway [28]. RUNX3 represses expression of CD4 [28–30] and activates CD8 expression [28]. In addition to these interactions, we further suggest two positive regulatory effects (one is from RUNX3 to GATA3 and the other is from GATA3 to pSTAT) and an autoactivation loop for GATA3 in order to make the activation patterns of the fixed points in the model consistent with those ones observed in Intermediate, CD4 SP and CD8 SP cells.

We list all these regulatory interactions in Table 5.1 and use $\rightarrow$ and $\dashv$ symbols to represent positive and negative regulatory interactions, respectively. Then using (**5.1**), we specify the input functions, ω' s, for each independent variable and present them in Table 5.2.

Table 5.2: Mathematical notation and specification of the input functions.

Component	State, Input	Input function
CD4	x_1, ω_1	$\omega_1 = \left(\frac{1+\eta_1}{\eta_1}\right)\left(\frac{\eta_1 x_3 x_4}{1+\eta_1 x_3 x_4}\right)\left(1-\left(\frac{1+\beta_3}{\beta_3}\right)\left(\frac{\beta_3 x_6}{1+\beta_3 x_6}\right)\right)$
CD8	x_2, ω_2	$\omega_2 = \left(\frac{1+\alpha_6}{\alpha_6}\right)\left(\frac{\alpha_6 x_6}{1+\alpha_6 x_6}\right)\left(1-\left(\frac{1+\beta_4}{\beta_4}\right)\left(\frac{\beta_4 x_4}{1+\beta_4 x_4}\right)\right)$
GATA3	x_3, ω_3	$\omega_3 = \left(\frac{1+\alpha_1+\alpha_2+\alpha_3}{\alpha_1+\alpha_2+\alpha_3}\right)\left(\frac{\alpha_1 x_7+\alpha_2 x_3+\alpha_3 x_6}{1+\alpha_1 x_7+\alpha_2 x_3+\alpha_3 x_6}\right)$
ThPOK	x_4, ω_4	$\omega_4 = \left(\frac{1+\alpha_4}{\alpha_4}\right)\left(\frac{\alpha_4 x_3}{1+\alpha_4 x_3}\right)\left(1-\left(\frac{1+\beta_1}{\beta_1}\right)\left(\frac{\beta_1 x_6}{1+\beta_1 x_6}\right)\right)$
pSTAT	x_5, ω_5	$\omega_5 = \left(\frac{1+\alpha_7+\alpha_8}{\alpha_7+\alpha_8}\right)\left(\frac{\alpha_7 x_3+\alpha_8 x_8}{1+\alpha_7 x_3+\alpha_8 x_8}\right)$
RUNX3	x_6, ω_6	$\omega_6 = \left(\frac{1+\alpha_5}{\alpha_5}\right)\left(\frac{\alpha_5 x_5}{1+\alpha_5 x_5}\right)\left(1-\left(\frac{1+\beta_2}{\beta_2}\right)\left(\frac{\beta_2 x_4}{1+\beta_2 x_4}\right)\right)$
TCR signal	x_7, ω_7	$\omega_7 = 0$
IL7	x_8, ω_8	$\omega_8 = 0$

Table 5.3: Zero-point states of *default* system.

	CD4	CD8	GATA3	ThPOK	pSTAT	RUNX3	TCR signal	IL7
$\mathbf{x}_\star$	0	0	0	0	0	0	0	0
$\mathbf{x}_\triangle$	1	0	0.9998	1	1	0	0	0
$\mathbf{x}_\square$	0	1	1	0	1	1	0	0
$\mathbf{x}_\dagger$	0	0	1	0.3448	1	0.3448	0	0

5.3 Results

5.3.1 Investigation of the structure of the state space

By substituting the input functions given in Table 5.2 into Equation (**4.5**), we obtain a system of 8-coupled nonlinear ODEs. We then take all parametric values (α' s, β' s and γ' s) as 1 since there is no available kinetic data on these specific interactions in the model. Also such an assumption does not favor any regulatory interactions or decay events. Furthermore, we call the system as *default* by also taking h as 50 to obtain step-like response curves, thus making our work easily comparable with possible Boolean models.

In order to explore the fixed points of *default* system and their stabilities (in biology, the fixed points represent distinct developmental stages and their level of maturity), we first find the zero-point states of the system by using *fsolve* optimization toolbox of MATLAB and present the results in Table5.3.

We then apply the linear stability analysis (as mentioned in Section 4.1.3.2) about these zero-point states. When the Jacobian matrix of the system is evaluated at $\mathbf{x}_\star$, the eigenvalues are found to be $\lambda^\star_{1,2,3,4,5,6} = -1$, $\lambda^\star_7 = -1 + 7.411 \times 10^{-8}$j and $\lambda^\star_8 = -1 - 7.411 \times 10^{-8}$j (here j stands for the imaginary unit). $\mathbf{x}_\star$ represents an attractor when the real parts of all the eigenvalues are negative. For the Jacobian matrix computed at $\mathbf{x}_\triangle$, the eigenvalues are $\lambda^\triangle_{1,2,3,4,5,6,7} = -1$ and $\lambda^\triangle_8 = -0.9960$ denoting $\mathbf{x}_\triangle$ as an attractor. At $\mathbf{x}_\square$, eigenvalues of the Jacobian matrix are almost the same as in the case of $\mathbf{x}_\star$ and thus $\mathbf{x}_\square$ is an attractor. On the other hand, the Jacobian matrix at $\mathbf{x}_\dagger$ gives the following

eigen values: $\lambda^{\dagger}_{1,2,3,4,5,6} = -1$, $\lambda^{\dagger}_{7} = -13.5014$ and $\lambda^{\dagger}_{8} = 11.5014$ which indicate an unstable fixed point.

To make the findings given above much more reliable and a comment on the sizes of the basins of attraction of these attractors, it would be helpful to computationally simulate time evolution (numerical integration) of the system of equations for all possible different initial conditions (initial values of the variables). Our dynamical representation is deterministic and asymptotic behavior of the system depends on its initial conditions. Since the variables change continuously in the interval of $[0,1]$, it is not realistic to search over all possible combinations of initial conditions. However, one can scan a big portion of the state space of the model by randomly picking up large number of initial conditions. We randomly choose 50 000 independent initial conditions and simulate the time evolutions for each case by using *ODE45* solver of MATLAB until the system converges to any steady state. Then, we cluster all the resultant steady states into three subsets ($\mathbf{x}_{\star}$, $\mathbf{x}_{\triangle}$ and $\mathbf{x}_{\square}$) and compute the mean and variance values for variables in each subset, presented in Table 5.4. (Values close to 1 are emphasized by using bold numbers.). On the other hand, any time evolution results in the $\mathbf{x}_{\dagger}$ point meaning that $\mathbf{x}_{\dagger}$ has a very tiny basin of attraction. This can also be interpreted as $\mathbf{x}_{\dagger}$ is not a biologically meaningful zero point state. Among the three subsets of attractors, two of them ($\mathbf{x}_{\triangle}$ and $\mathbf{x}_{\square}$) represent the specific activation patterns of CD4 SP (helper) and CD8 SP (killer) mature T cells. These two attractors have large basins of attraction. While 49.0% of all runs ends up with a final state of CD4 SP attractor, 47.9% results in CD8 SP attractor. The rest ($\mathbf{x}_{\star}$) with a very small basin of attraction belongs to a distinct subset which has no activated elements. This subset is a cluster of attractors from where the system moves to one of the other two attractors in response to sufficient TCR and IL7 perturbations. Then, we call it as Intermediate ($CD4^{+}CD8^{low}$) cell type attractor since this population is extensively accepted as the last intermediate common precursor to follow either CD4 or CD8 differentiation pathways during thymic development [7].

Table 5.4: Attractors of our network as a continuous dynamical system.

	$CD4^{+}CD8^{low}$ (Intermediate)		CD4 SP (T helper cell)		CD8 SP (T killer cell)	
	Mean	Variance	Mean	Variance	Mean	Variance
CD4	2.3897×10^{-9}	3.9677×10^{-18}	**1.0000**	1.3290×10^{-16}	2.8640×10^{-9}	5.6079×10^{-18}
CD8	2.4254×10^{-9}	4.4955×10^{-18}	2.8790×10^{-9}	5.6852×10^{-18}	**1.0000**	1.0647×10^{-16}
GATA3	8.6543×10^{-10}	8.3177×10^{-19}	**0.9998**	6.9078×10^{-18}	**1.0000**	1.0501×10^{-17}
ThPOK	2.8897×10^{-9}	4.1818×10^{-18}	**1.0000**	2.7547×10^{-17}	2.1108×10^{-9}	4.3198×10^{-18}
pSTAT	3.6466×10^{-9}	7.8987×10^{-18}	**1.0000**	1.1635×10^{-17}	**1.0000**	1.1163×10^{-17}
RUNX3	1.0695×10^{-9}	1.6978×10^{-18}	2.2083×10^{-9}	4.4105×10^{-18}	**1.0000**	2.9110×10^{-17}
TCR signal	7.5626×10^{-10}	7.7725×10^{-19}	2.9674×10^{-9}	5.6418×10^{-18}	2.8744×10^{-9}	5.4612×10^{-18}
IL7	1.9706×10^{-9}	3.0986×10^{-18}	2.8110×10^{-9}	5.6585×10^{-18}	2.9562×10^{-9}	5.5605×10^{-18}

5.3.2 Effects of intrathymic TCR and IL7 signalings onto CD4 vs. CD8 lineage choice

After determining *the most possible* steady states of the system, we apply many combinations of TCR and IL7 signals onto a single precursor cell type to see whether they drive the cell to one of the two mature subsets or not. Of note, we use the *default* set of parametric values again. Our simulations show that a weak and short-lived TCR signaling does not lead to CD4 SP lineage as shown in Fig.5.8 and Fig.5.9, while a short-lived but strong TCR signaling results in differentiation to single positive CD4 lineage (see Fig.5.13). In Fig.5.10, Fig.5.11, very weak TCR signals with different durations are introduced and the Intermediate thymocyte cannot differentiate to any mature subset again. On the other hand, a TCR signal has the same strength with that one in Fig.5.9 is subjected to the Intermediate cell for a longer duration and the result is now differentiation to CD4 SP helper lineage (See Fig.5.12). These results underline the importance of both strength and duration of TCR signals onto lineage decision of an Intermediate thymocyte. In Fig.5.13, we also test the stability of the CD4 SP attractor through introducing a maximum level of a short-lived IL7 signaling and see that the system doesn't move away from it: The CD4 SP attractor stays stable. Next, we want to see if IL7 signaling itself would be enough to secure CD8 lineage. We predict that the CD8 lineage can be preserved if the precursor cell is subjected to a strong IL7 signaling for a longer period of time (see Fig.5.14 and Fig.5.15). In addition, our calculation predicts that when strong and sharp (short-lived) IL7 signaling accompanied by a weak and short-lived TCR signaling the CD8 lineage to be secured (See Fig.5.16). Let us remind that, introducing a maximum level of TCR signaling doesn't change the CD8 lineage choice (See Fig.5.15) manifesting that the CD8 SP attractor is stable.

In another scenario, a stronger but short-lived IL7 signal following a weak and short-lived TCR signaling is found to be not sufficient to converge CD4 SP directed lineage to CD8 SP attractor as shown in Fig.5.17.

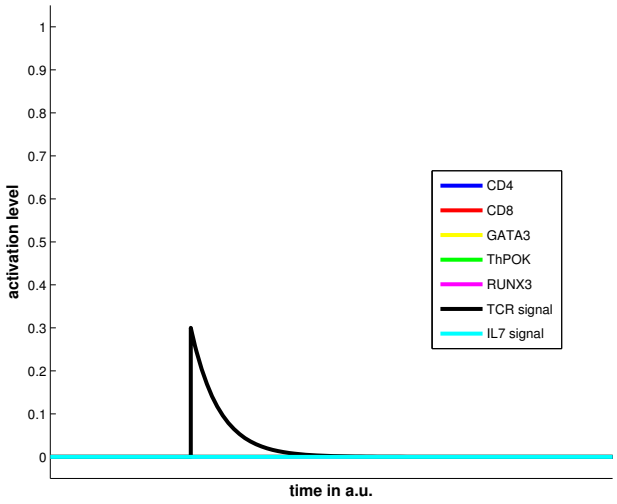

Figure 5.8: A very weak and short-lived TCR signaling cannot promote CD4 SP lineage choice of the Intermediate cell.

5.3.3 Effects of mutations *in silico*

In our network model, we specify four *key* regulatory nodes: GATA3, ThPOK, RUNX3 and pSTAT. Despite the lack of quantitative measurements on these specific regulatory factors in T cell differentiation, qualitative measurements on their salient knockout and/or overexpression are available. To cross-validate results from our network model with those from the literature, we implement corresponding experiments *in silico*[1] by manipulating their related variables in our system of equations. Knockout and overexpression of a component can be implemented by keeping its activation level value as 0 and 1, respectively, at all times during simulations. If there is no manipulation about the variables of the system then the system is called *Wild Type* (WT). By setting the activation level value of any node as 0 or 1 in the entire simulation, *i*) a specific one or more of steady states related to the WT system of equations can become unreachable, *ii*) a new one or set of either biological or non-biological steady states can emerge, and *iii*) all the steady state motifs can remain as in the WT case.

[1] 'In silico' is a computational biology jargon that means 'in computer simulation' .

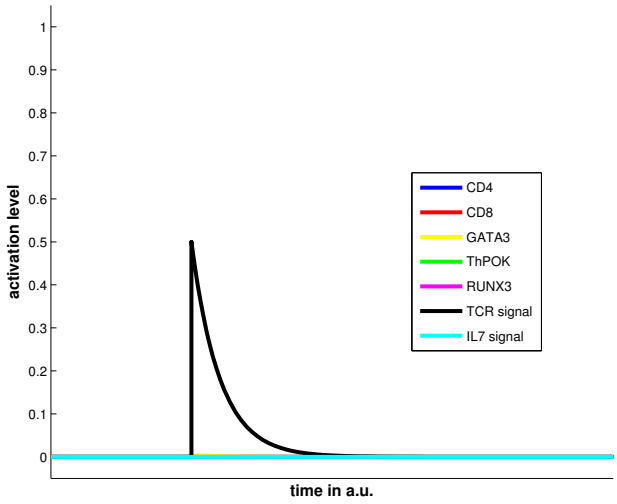

Figure 5.9: A weak and short-lived TCR signaling cannot promote CD4 SP lineage choice of the Intermediate cell.

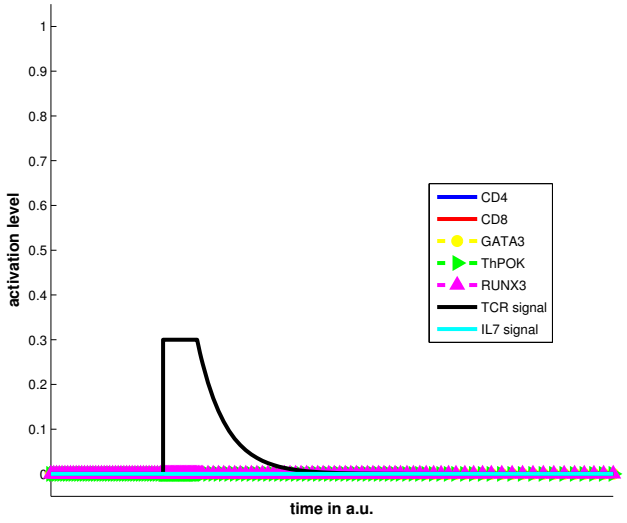

Figure 5.10: A very weak and a bit longer TCR signaling cannot determine the lineage choice of the Intermediate cell.

We first analyze steady state patterns of single node knockouts for all key regulatory nodes with *default* set of parametric values by following the same procedure defined in Section 5.3.1. We form Table 5.5 to present the corresponding results. Then, we test single node overexpressions and give the corresponding results in Table 5.6. Finally,

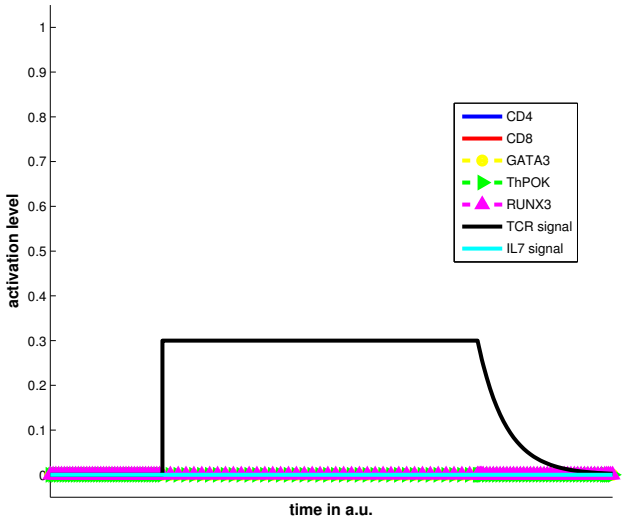

Figure 5.11: A very weak and much longer TCR signaling cannot determine the lineage choice of the Intermediate cell.

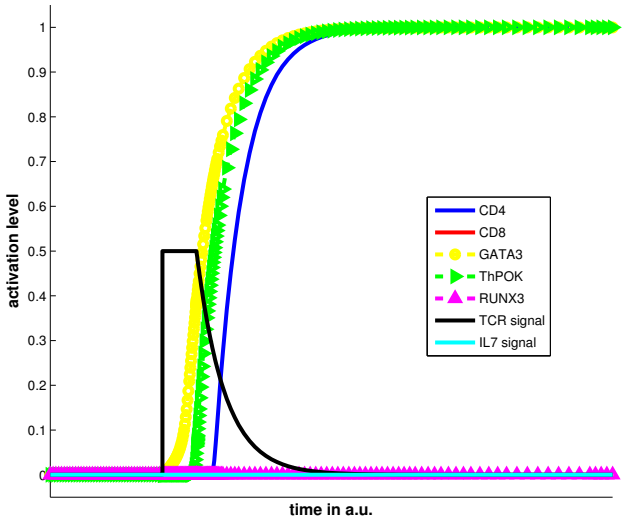

Figure 5.12: A weak but a bit longer TCR signaling leads to CD4 SP lineage choice of the Intermediate cell.

we apply two synergistic mutations on the system *in silico* by knocking GATA3 out with overexpressing ThPOK and knocking ThPOK out with overexpressing GATA3 as summarized in Table 5.7. For knockout/overexpression of ThPOK and RUNX3, there is almost a perfect agreement between our findings and those found in the literature.

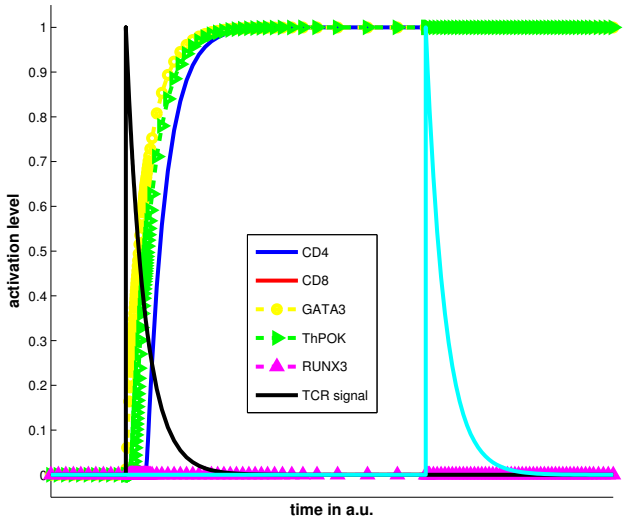

Figure 5.13: A strong and short-lived TCR signaling leads to CD4 SP lineage choice of the Intermediate cell.

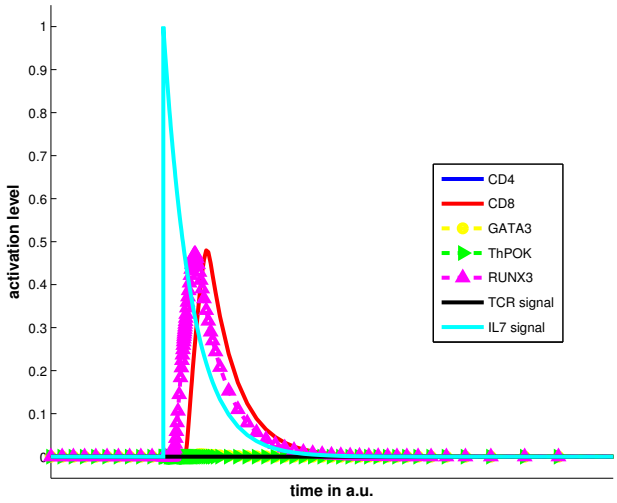

Figure 5.14: A strong but short-lived IL7 signaling is not enough to drive CD8 lineage choice of the Intermediate cell alone.

For mutations of pSTAT, we have no experimental knowledge to compare our results. And finally, our findings for knockout/overexpression of GATA3 are not consistent with experiments.

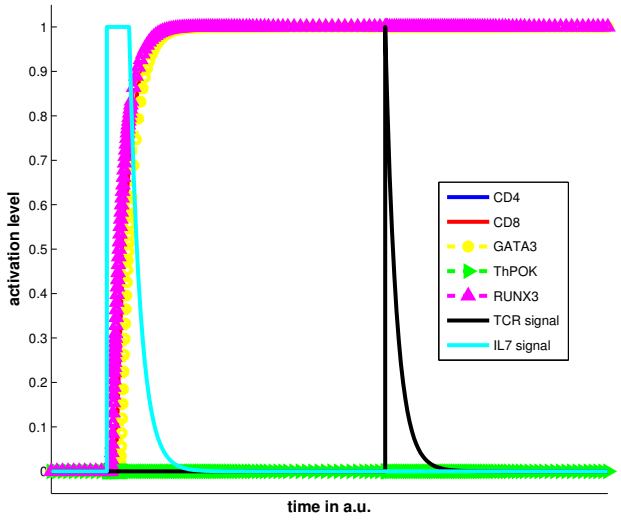

Figure 5.15: A strong and a bit longer IL7 signaling leads to CD8 SP lineage choice of the precursor cell of the Intermediate cell.

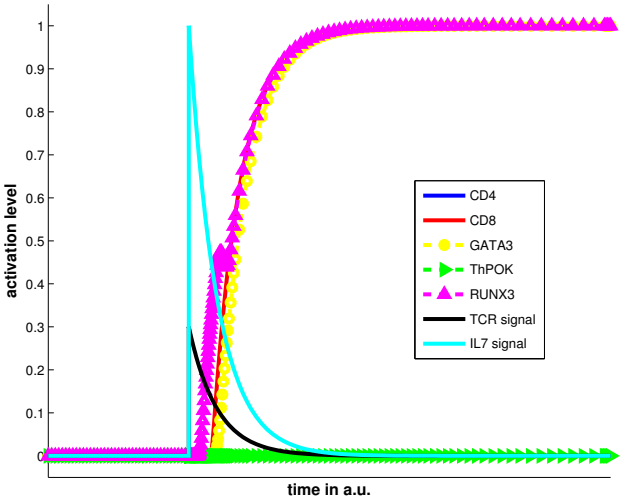

Figure 5.16: A strong but short-lived IL7 signaling with a concurrent weak (and short) TCR signaling leads to CD8 SP lineage choice of the Intermediate cell.

In 'Computational Finding' columns of Tables 5.5, 5.6 and 5.7, we determine whether three WT steady states are still reachable and new attractors emerge. In 'Implications' columns, we interpret the biological effects of each corresponding mutation on the intrathymic differentiation. In 'Evidence' columns, we give a list of publications (we are aware of), supporting our predictions.

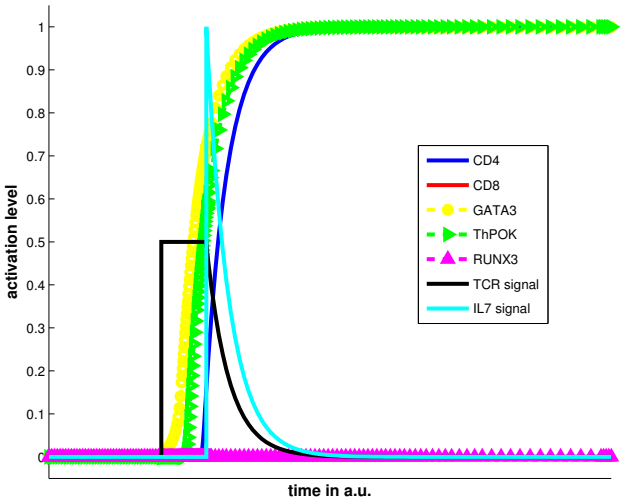

Figure 5.17: A weak but a bit longer TCR signaling leads to CD4 SP lineage choice of the Intermediate cell even if a strong IL7 signaling is received as soon as TCR signals ceases.

Table 5.5: Single node knockout experiments *in silico*.

Single Node Knockout	**Computational Finding**	**Implications**	**Evidence**
GATA3	CD4 SP and CD8 SP attractors disappear	Absence of both helper and cytotoxic lineage	
ThPOK	CD4 SP attractor disappears	Absence of helper lineage MHC II-restricted thymocytes may be redirected into cytotoxic lineage	[9, 22]
RUNX3	CD8 SP attractor disappears	Absence of cytotoxic lineage MHC I-restricted thymocytes may be redirected to helper lineage	[23, 29]
pSTAT	CD8 SP attractor disappears	Absence of cytotoxic lineage MHC I-restricted thymocytes may be redirected to helper lineage	

5.4 Conclusions and Recommendations

We give a compact but coarse-grained mathematical description for CD4/CD8 lineage choice of $\alpha\beta$ thymocytes development beyond the Intermediate ($CD4^{+}CD8^{low}$) stage in the thymus. Being an ODE model, our mathematical model also needs kinetic data,

Table 5.6: Single node overexpression experiments *in silico*.

Single Node Overexpression	Computational Finding	Implications	Evidence
GATA3	Intermediate attractor disappears		
ThPOK	Intermediate and CD8 SP attractors disappear	Absence of cytotoxic lineage MHC I-restricted thymocytes may be redirected into helper lineage	[19, 22]
RUNX3	Intermediate and CD4 SP attractors disappear	Absence of helper lineage MHC II-restricted thymocytes may be redirected into cytotoxic lineage	[31]
pSTAT	Intermediate attractor disappears		

Table 5.7: Synergistic perturbation experiments *in silico*.

Synergistic Perturbation	Computational Finding	Implications	Evidence
GATA3 Knockout & ThPOK Overexpression	All WT steady state patterns disappear and a new attractor emerges with non-zero ThPOK activation only	Helper lineage can not be rescued even if ThPOK is overexpressed in the case of GATA3 deficiency	[19]
ThPOK Knockout & GATA3 Overexpression	Intermediate and CD4 SP attractors disappear	Absence of helper lineage MHC II-restricted thymocytes may be redirected into cytotoxic lineage	

but even in the *default* case of parametric values it gives a qualitative explanation of the dynamical properties for the differentiation.

In our model, each regulatory link between any nodes represent flow of information between them in the specified direction. It means that, even if any other intermediary

element has a role on the line of a link it would not change the behavior of information flow. So, the model can be extended into a more detailed one by only locating intermediary component(s) on the lines of current links.

Our model has the ability to show the effects of intrathymic signals onto lineage choice in line with the implications of kinetic signaling model. We first mimic two fundamental assumptions of kinetic signaling model: *i*) Strong and/or long TCR signals lead to differentiation to CD4 lineage, *ii*) IL7 cytokine signaling pathway promotes CD8 gene program by activating CD8 and inhibiting CD4. Our model further suggests that short-lived IL7 cytokines cannot promote CD8 lineage choice alone. But, if a proper and simultaneous TCR signal is received then an Intermediate cell can differentiate to CD8 lineage. The latter finding may be supported by the limited amount of IL7 cytokines in the thymus [24].

Future studies should be planned in order to explain 'coreceptor reversal' [8] mechanism after positive selection which is not achieved by our current model since its scope is "Intermediate to SP" transition and it does not involve the feedback roles of CD4 and CD8 coreceptors in TCR signaling event.

Our simulations involving salient mutations give results that are in good agreement with experimental findings except for the cases of GATA3 knockout[2] and overexpression[3]. The model can be further improved by using another component to feed IL7 signaling pathway rather than GATA3 as being an essential TCR signaling downstream factor. Or it may be possible that GATA3 or any other GATA3 mediated factor weakly contributes IL7 signaling pathway, that can be implemented by a relatively small activation strength parameter for corresponding interaction in our mathematical model.

[2]Disruption of GATA3 expression arrests development of MHC class II-restricted thymocytes during 'DP to CD4 SP' transition but it does not redirect them to CD8 SP lineage. It just minimally affects the CD8 SP population [19].

[3]Overexpression of GATA3 blocks the differentiation to CD8 SP lineage [6], but it doesn't result in redirection of MHC class I-restricted thymocytes to CD4 SP lineage [18].

REFERENCES

[1] **Sompayrac, L.**, (2003). How the Immune System Works, Blackwell Publishing, Massachusetts, USA.

[2] **Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P.**, (2008). Molecular Biology of the Cell, Garland Science, New York, USA.

[3] **Enver, T. et al.**, (2009). Stem cell states, fates, and the rules of attraction, *Cell Stem Cell*, **4**, 387–397.

[4] **Janeway Jr., C.A., Travers, P., Walport, M., Shlomchik, M.J.**, (2001). Immunobiology, Garland Publishing, New York, USA.

[5] **Url-1**, http://en.wikipedia.org/wiki/Transcription_factor, date retrieved 04.30.2012.

[6] **Singer, A., Adoro, S., and Park, J.H.**, (2008). Lineage fate and intense debate: myths, models and mechanisms of CD4/CD8 lineage choice, *Nature Reviews Immunology*, **8(10)**, 788–801.

[7] **Dave, V.P. et al.**, (1998). HD mice: A novel mouse mutant with a specific defect in the generation of $CD4^+$ T cells, *Proceedings of National Academy of Sciences of the USA*, **95**, 8187–8192.

[8] **Brugnera, E. et al.**, (2000). Coreceptor Reversal in the Thymus: Signaled $CD4^+CD8^+$ Thymocytes Initially Terminate CD8 Transcription Even When Differentiating into $CD8^+$ T Cells, *Immunity*, **13**, 59–71.

[9] **Kappes, D.**, (2010). Expanding roles for ThPOK in thymic development, *Immunological Reviews*, **238**, 182–194.

[10] **De Jong, H.**, (2002). Modeling and Simulation of Genetic Regulatory Systems: A Literature Review, *Journal of Computational Biology*, **9(1)**, 67–103.

[11] **Krumsiek, J. et al.**, (2011). Hierarchical Differentiation of Myeloid Progenitors Is Encoded in the Transcription Factor Network, *PLoS ONE*, **6(8)**, e22649.

[12] **Mendoza, L. & Xenarios, I.**, (2006). A method for the generation of standardized qualitative dynamical systems of regulatory networks, *Theoretical Biology and Medical Modelling*, **3**, 13.

[13] **Strogatz, S.H. et al.**, (1994). Nonlinear Dynamics and Chaos With Applications to Physics, Biology, Chemistry and Engineering, Perseus Book Publishing, Cambridge, MA.

[14] **Roussel, M.**, (2005). Stability Analysis for ODEs.

[15] **Gillespie, D.**, (1977). Exact stochastic simulation of coupled chemical reactions, *Journal of Physical Chemistry*, **81**, 2340.

[16] **Artyomov, M.N. et al.**, (2007). Purely stochastic binary decisions in cell signaling models without underlying deterministic bistabilities, *Proceedings of National Academy of Sciences of the USA*, **104(48)**, 18958–18963.

[17] **Artyomov, M.N. et al.**, (2010). CD4 and CD8 binding to MHC molecules primarily acts to enhance Lck delivery, *Proceedings of National Academy of Sciences of the USA*.

[18] **Hernandez-Hoyos, G. et al.**, (2003). GATA3 expression is controlled by TCR signals and regulates CD4/CD8 differentiation, *Immunity*, **19**, 83–94.

[19] **Wang, L. et al.**, (2008). Distinct functions for the transcription factors GATA-3 and ThPOK during intrathymic differentiation of $CD4^+$ T cells, *Nature Immunology*, **9(10)**, 1122–1130.

[20] **Egawa, T. and Littman, D.R.**, (2008). The transcription factor ThPOK acts late in helper T cell lineage specification and suppresses Runx-mediated commitment to the cytotoxic T cell lineage, *Nature Immunology*, **9(10)**, 1131–1139.

[21] **Muroi, S. et al.**, (2008). Cascading suppression of transcriptional silencers by ThPOK seals helper T cell fate, *Nature Immunology*, **10**, 1113–1121.

[22] **He, X. et al.**, (2005). The zinc finger transcription factor Th-POK regulates CD4 versus CD8 T-cell lineage commitment, *Nature*, **433**, 826–833.

[23] **Setoguchi, R. et al.**, (2008). Repression of the transcription factor Th-POK by Runx complexes in cytotoxic T cell development, *Science*, **319**, 822–825.

[24] **Jiang, Q. et al.**, (2005). Cell biology of IL-7, a key lymphotrophin, *Cytokine & Growth Factor Reviews*, **16**, 513–533.

[25] **Park, J.H. et al.**, (2010). Signaling by intrathymic cytokines, not T cell antigen receptors, specifies CD8 lineage choice and promotes the differentiation of cytotoxic-lineage T cells, *Nature Immunology*, **11(3)**, 1257–1265.

[26] **Cui, Y. et al.**, (2004). Inactivation of Stat5 in Mouse Mammary Epithelium during Pregnancy Reveals Distinct Functions in Cell Proliferation, Survival, and Differentiation, *Molecular and Cellular Biology*, **24(18)**, 8037–8047.

[27] **Woolf, E. et al.**, (2003). Runx3 and Runx1 are required for CD8 T cell development during thymopoiesis, *Proceedings of National Academy of Sciences of the USA*, **100(13)**, 7731–7736.

[28] **Sato, T. et al.**, (2005). Dual functions of Runx proteins for reactivating CD8 and silencing CD4 at the commitment process into CD8 thymocytes, *Immunity*, **22**, 317–328.

[29] **Egawa, T. et al.**, (2007). The role of the Runx transcription factors in thymocyte differentiation and in homeostasis of naive T cells, *The Journal of Experimental Medicine*, **204(8)**, 1945–1957.

[30] **Taniuchi, I. et al.**, (2002). Differential requirements for Runx proteins in CD4 repression and epigenetic silencing during T lymphocyte development, *Cell*, **111**, 621–633.

[31] **Kohu, K. et al.**, (2005). Overexpression of the Runx3 Transcription Factor Increases the Proportion of Mature Thymocytes of the CD8 Single-Positive Lineage, *The Journal of Immunology*, **174**, 2627–2636.

CURRICULUM VITAE

Name Surname: Emrah ŞİMŞEK

Place and Date of Birth: Fatih / 25.09.1986

Adress: Soğanlı Mah. Bağcılar Cad. Gökçe Sok. No:9/7
34183 Bahçelievler/İSTANBUL

E-Mail: emrahsimsek@itu.edu.tr

B.Sc.: Fatih University, 2009

Professional Experience and Rewards:

- M.Sc., Research Assistantship at Physics Engineering Department of Istanbul Technical University, Febraury 2010 - Present
- B.Sc., Graduation degree with the highest GPA in the department, 2009

List of Publications and Patents:

PUBLICATIONS/PRESENTATIONS ON THE THESIS

- *Short Talk Presentation*: "A Qualitative Dynamical Model for T-cell Differentiation in the Thymus" (in English)

5th Mirror Conference on Statistical Physics, Condensed Matter Physics and Social Responsibility,
December 26, 2011, Sabancı University, Turkey

- *Seminar Presentation*: "A Mathematical Model for $\alpha\beta$ T-cell Differentiation in the Thymus" (in Turkish)

December 21, 2011, Experimental Medicine Research Institute (DETAE), Istanbul University, Turkey

- *Short Talk Presentation*: "T-cell Differentiation in the Thymus" (in Turkish)

18th İstanbul Statistical Physics Days,

July 1, 2011, Sabancı University, Turkey