Modular Analysis
of
Signal Transduction
Networks

Der Fakultät für Verfahrens- und Systemtechnik
der Otto-von-Guericke-Universität Magdeburg
zur Erlangung des akademisches Grades

Doktoringenieur
(Dr.-Ing.)

am 02.04.07 vorgelegte Dissertation

von Dipl. Ing. Julio Sáez Rodríguez

Hauptberichter: Prof. Dr.-Ing. Dr. h.c. mult. E.D. Gilles
Mitberichter: Prof. Dr.-Ing. J. Stelling

Tag der mündlichen Prüfung: 05.07.2007

Modular Analysis
of
Signal Transduction
Networks

Julio Saez-Rodriguez

Acknowledgments

First and foremost I would like to thank Professor Ernst Dieter Gilles for his excellent supervision. He always found the perfect balance between letting scientific freedom and providing the key idea and motivating remark whenever I got stuck. He is a mirror where any future professor should look at.

I am also indebt to Professor Jöerg Stelling, who accepted to review my thesis.

I am deeply grateful to all the members of the Systems Biology Group at the Max-Planck-Institute, specially to my closest collaborators: Holger, Conzelmann, Steffen Klamt, Carsten Conradi, Jeremy Huard, Martin Ginkel, Sebastian Mirschel and Andreas Kremling. Special thanks to Hanna Sharp and Regina Reiner for a great proofreading of the text, and to Renate Wagner and Janine Holzmann for help with organizatorial issues and a great coffee every morning. Also thanks to the students I supervised, Holger Conzelmann, Andrea Hammerle, Stefan Gayer, Immaculada Ramos, Octavian Bucur, Jona Epperlein, and Rebecca Hemenway: I learnt from them much more than what I could teached them.

I also have much to thank to colleagues in other groups at the MPI, specially Odon Angeles, Michael Mangold, Martin Haefele, and Kostya Teplynskyy. A special word deserves Dietrich Flockerzi: whenever I run to him with a mathematical question, he was able to answer and explain it in abstract, yet beautiful, terms. Thanks to the whole institute for the excellent working atmosphere, in particular to the IT staff and those with whom I shared my office, Detlev Bannasch, Hanna Sharp, Sophia Fischer, Immaculada Ramos, Octavian Bucur, and Rebecca Hemenway.

I am specially indebt to Birgit Schoeberl for helping me getting started into the field, and for her support and being 'always there' despite the deep ocean between us. I also thank the colleagues at the Institute of Immunology, directed by Prof. B. Schraven. Special thanks to Luca Simeoni, Jon Lindquist and Xiaoqian Wang. I had the honour to exchange very stimulating discussions with a number of excellent scientist, from whom I would like to highligh Boris Kholodenko, Eduardo Sontag, and Isabel Merida.

I want to express my thanks to my friends who helped me in non-scientific matters, specially Marta, Javi, Dimas and Elena.

This thesis, and everything I have done and what I am, would have never been possible without the unconditional love of my family, specially of my parents Julio and Blanca, and my brother Edu. Only the enjoyment I have working in science compensates spending so little time with them.

Finally, words can not express how thankful I am to Livi for her support, love, and patience through all these years.

Boston, 16th July 2007.

Julio Saez Rodriguez

Kurzfassung

Biologische Systeme, und insbesondere Signaltransduktionsnetzwerke, sind durch eine hohe Komplexität aber auch durch eine Modularstruktur gekennzeichnet. Das Leitmotiv dieser Arbeit ist, durch Ausnutzung der letzteren die erstere zu beherrschen. Um diesen Modularansatz anzuwenden, müssen folgende Schritte durchgeführt werden:

1. Zerlegung der Signalnetzwerke in sinnvolle Module,

2. Grundsätzliche Analyse dieser Module, und

3. Zusammenschaltung der Module (oder eines reduzierten Ersatzes davon) und Analyse des gesamten Netzwerkes.

Diese Arbeit beschäftigt sich mit allen drei Schritten, deren Implementierung in nützliche Werkzeuge und deren Anwendung an praktischen Beispielen.

Ein Ingenieurwissenschaftliches Kriterium für die Zerlegung in Module wurde entwickelt. Dieses Kriterium basiert auf dem Konzept der Rückwirkungsfreiheit. Dieses Konzept wurde durch einen automatischen Algorithmus implementiert, der in der Lage war, selbst sehr komplexe Signalnetzwerke in sinnvolle Untereinheiten zu zerlegen.

Ein domänenorientierter Ansatz liefert einen komplementären Gesichtpunkt zu der Modularität von biologischen Systemen: Er ergibt, dass die Moleküldomänen die Grundeinheiten von Signaltransduktionsnetzen sind. Deswegen wurde ein Baukasten von Domänen (Motiven) definiert, mit dem man beliebige Netzwerke aufbauen kann.

Die Motive wurden sorgfältig in Hinblick auf drei entscheidende system-theoretische Eigenschaften untersucht: Stabilität, Monotonie, und Eingang/Ausgang Verhalten.

Die Analyse zeigt, dass einige Motive Multistabilität aufweisen können, aber diese Eigenschaft gegenüber Konzentrationsfluktuationen nicht robust ist. Darüber hinaus sind fast alle Motive monoton, und alle sind durch ein monotones stationäres Verhalten gekennzeichnet. Weiterhin können diese Motive größtenteils durch eine nichtlineare Kennlinie und ein einfaches lineares System ersetzt werden.

Als Grenzfall können Motive durch logische Funktionen ersetzt werden. Dies erlaubt eine Vereinfachung der Signalnetzwerke, wodurch sehr komplexe Netzwerke untersucht werden können. Am Beispiel der Signalwege in Lymphozyten konnte gezeigt werden, dass

damit nicht nur die aktuellen biologischen Kenntnisse zusammengefasst und getestet werden können, sondern auch neue, unerwartete Erkenntnisse gewonnen werden können.

Anschließend wurden die dynamischen Eigenschaften von Zusammenschaltungen kleine Module untersucht. Unter anderem wurde eine minimale mathematische Realisierung für ein ganz bestimmtes dynamisches Verhalten erarbeitet, welches man in der Aktivierung der MAPK Kaskade in Lymphozyten beobachtet hat.

Außerdem wurde diese Implementierung in der Signalkaskade identifiziert. Letztlich wurde die Anwendung des Modularansatzes für die Analyse der Dynamik komplexerer Signaltransduktionsnetze anhand eines detaillierten Modells der EGF-induzierten MAPK Kaskade demonstriert.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Cells are fascinating objects, able to perform extremely complex tasks. One of their most remarkable properties is the ability to sense their environment, process external signals, and react accordingly. Thanks to the spectacular development of molecular biology during the last years, our view of cellular processes in general, and signal transduction in particular, has evolved into a picture of captivating, but almost discouraging, complexity. Particularly, in the case of signal transduction, the large number of elements involved, their complicated non-linear interactions, the high number of feedback loops, and the crosstalk among different pathways (see Figure 1.1), make impossible an intuitive understanding[10], specially of how this signaling machinery works as a whole.

This holistic understanding is the goal of systems biology[141]. Thereby, the goal is not only to gain insights into the fundamental principles of life, but there is also the hope to improve treatment of important diseases: essential processes like proliferation, cell development or even apoptosis (cell suicide) are controlled, at least in part, by cell signaling. Thus, defective signal processing can lead to important diseases such as cancer or diabetes, and therefore signaling pathways are important targets for disease therapy[158].

Three main fields converge in systems biology[248]: (i) Experimental molecular biology, which provides large sets of (as quantitative and reproductive as possible) data; (ii) bioinformatics, contributing with tools to process, organize and visualize these data, and (iii) mathematical foundations (coming mainly from biophysics and systems theory) to analyze, on the basis of mathematical models, cellular processes. The knowledge and amount of data available about signaling networks grows steadily, boosting the development of increasingly complex models. These models offer a highly detailed picture of signaling pathways, but the properties of these systems as a whole are difficult to understand. Therefore, the definition of functional units, i.e. entities whose function is separable from those of other units, has been proposed as a promising rationale for the analysis of large biochemical networks[103;151;197;220]. This modular approach follows a simple rationale: divide and win. As
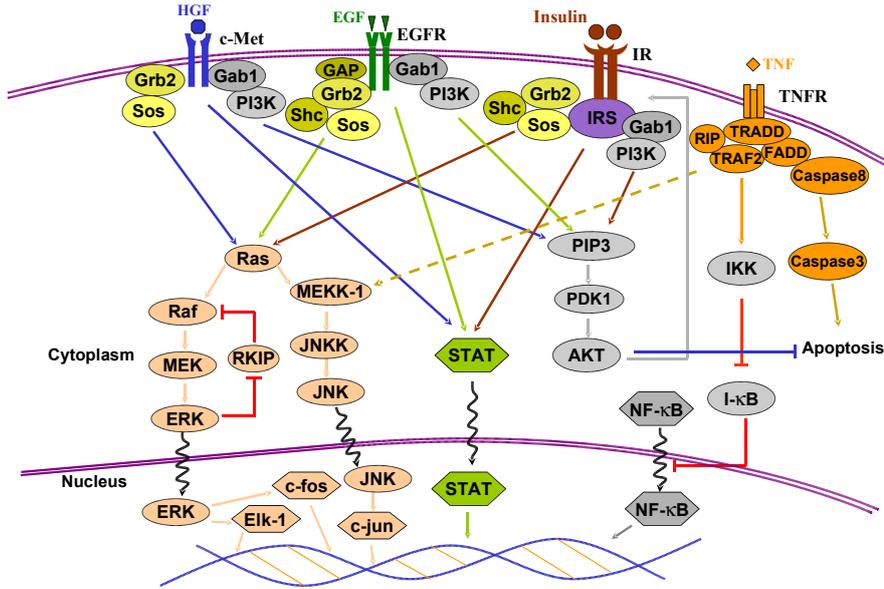
Figure 1.1: Complexity in signaling networks. The figure represents the key elements involved in the signaling processes triggered by the Epidermal Growth Factor Receptor (EGFR), Insulin Receptor, Tumor Necrosis Factor Receptor (TNFR) and Hepatocyte Growth Factor Receptor (HGFR), see Section 2.2.1). These pathways are key for the cell fate of many mammal cell. Some of the mechanisms responsible for the complexity of signaling networks can be seen here: high number of elements involved, crosstalk among the different pathways, and feedback loops (for the sake of clarity, only two are depicted).

depicted in Figure 1.2, by decomposing a system into subunits, one obtains modules which are significantly easier to handle. Once these *relatively* simple units are well understood, they can be re-assembled in order to analyze the emergent properties of the resulting systems. Furthermore, one could set up a kit of reusable elements, simplifying the setup of models, since many parts of biological networks are found in several signal transduction pathways. The present work is devoted to this modular approach, and tries to contribute to the different aspects it involves, as outlined in the following section.

1.2 Outline

As mentioned above, to apply a modular approach (see Figure 1.2), the following tasks must be performed:

- decompose the signaling network into modules,

- analyze the modules, and

- rewire the modules into the network and analyze the network.

This thesis is structured accordingly, after a biological prelude in Chapter 2, where the basic principles of signal transduction are introduced and the signaling networks used as
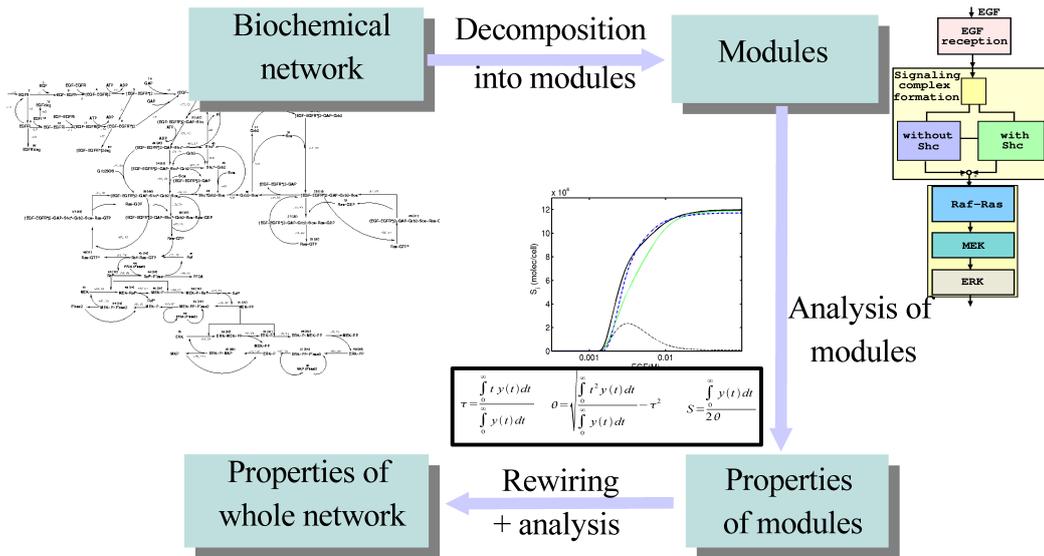
Figure 1.2: General procedure of a modular analysis approach. The approach starts with the decomposition of a large network in a suitable manner into smaller, easier to analyze subunits. Upon a thorough analysis of the resulting modules, they can be rewired together (either in their original form or in a reduced one[47]), and new insights into the network as a whole can be obtained[220].

case studies throughout this work are presented. These networks are the Epidermal Growth Factor (EGF) induced MAPK cascade and the T-cell receptor (TCR)-dependent signaling network.

Chapters 3 and 4 deal with the first point: the decomposition of signaling networks into functional units. Even though modularity is an accepted property of biological systems, how biochemical modules should be delimited remains an open question[276]. It starts with a review of the current understanding of the modularity of biochemical networks in Section 3.1. In Section 3.2, a novel system-theoretical criterion for this decomposition, namely the absence of retroactivity, is presented. Subsequently, Section 3.3 presents an algorithm for the automatic identification of modules for a given model (network), based on the concept of retroactivity, whose applicability is demonstrated with several examples in Section 3.4. The modularly structured models can be imported into ProMoT, a modeling tool which offers a natural environment for a modular set up of networks, providing thus a framework to analyze isolated modules or combination of them.

Chapter 3 deals with the decomposition of a *given* model; however, a related yet different question which is also of great importance in this context is the *inherent* modularity of signaling networks. Chapter 4 is devoted to this issue. There, it is shown how the application of a domain-oriented approach recently introduced[48] allows to define a discrete set of minimal modules (motifs) which appear recurrently. Since most networks can be set up as a combination of them, they can be seen as a construction kit for signaling networks. Therefore, a set of motifs was defined and implemented in ProMoT[95]. The applicability of this library

is exemplified in Section 4.3.1 by a model for the T-cell receptor (TCR) induced MAPK cascade.

Once modules are defined or found in a given network, they should be systematically analyzed. However, thorough systems-theory-oriented analyses of signaling networks based on their modularity are still scarce. The application of system-theoretical methods to biology, yet appealing, is not a trivial task and in fact, systems biology poses stimulating challenges to systems theory[241]. The motifs resulting from Chapter 4 are analyzed with respect to their system-theoretical properties in Chapter 5. In particular, stability, monotony, and dynamic behavior are examined. The dynamic was analyzed by comparing these motifs to simple linear systems combined with a non-linear characteristic curve (so-called Hammerstein modules).

Finally, Chapters 6 and 7 deal with the last question: how to cope with the network as a whole once it is structured in a modular manner. In Chapter 6 it is proposed to handle large signaling networks as a set of modules described as Boolean functions. It is shown that, even though this approach imposes a strong simplification of the reality, it allows to gain holistic insights into large networks. To illustrate its applicability, a large, curated model of T-cell signaling was set up. The model is, to the best of our knowledge, the largest one of its sort, comprising 94 different compounds. To set up large logical models in an efficient manner, ProMoT's abilities were extended, providing thus a new tool for a visual set up of large signaling networks within a Boolean formalism.

For a more detailed analysis, particularly addressing points dealing with the dynamics, a kinetic description is certainly required, as will be discussed in Chapter 7. Here, an important issue is the emergence of new properties due to the connection of simple models[20]. In Section 7.1 this topic will be explored via two simple models, one of a MAPK cascade[134] and a simple feedback system motivated by a non-trivial dynamic observed in the TCR-induced MAPK cascade. Finally, a remarkably complex model of the EGF induced MAPK cascade[234] will be thoroughly analyzed, using the concepts introduced in Section 5.3, exploring which new insights can be obtained.

To sum up, through this thesis contributions to the different steps involved in a modular approach will be presented.

Chapter 2

Molecular Biology
of Signal Transduction:
Basic Principles and Cases of Study

This chapter introduces the basic concepts required to understand signaling networks and their functioning, and presents succinctly the systems used through this thesis. The explanations are kept basic and concise, as they are aimed to provide the reader not familiar with signal transduction a certain background to understand this work. For deeper information the reader is referred to books devoted to signal transduction such as the book of Krauss[146].

2.1 Basic Principles of Signal Transduction

Cells, ranging from bacteria (which need to be aware of changes in their environment, e.g. of temperature, pH, or concentration of nutrients) to human cells (which must coordinate themselves with the rest of the cells in the organism), are equipped with exquisite sensing systems which allow them to receive and process a myriad of signals, and response accordingly. The complexity of the corresponding molecular machineries, in accordance with the complicated tasks they have to perform, is overwhelming.

Typically, the binding of extracellular ligands to molecular receptors at the cell membrane results in changes in the intracellular part of the receptor. The activated receptor transmits the signal to intracellular signaling intermediates. Finally, these intermediates activate transcription factors which move into the nucleus and modify the gene expression of the cell, resulting in the production of the proteins required to react to the external stimuli[60] (see Figure 2.1).

The key components in this transfer of information are proteins that can change their state by interaction with other proteins or by biochemical modifications, such as phosphorylations, catalyzed by other proteins. The proteins that catalyze phosphorylations are known as kinases and their counter partners, which remove phosphate groups from proteins, phosphatases. An-
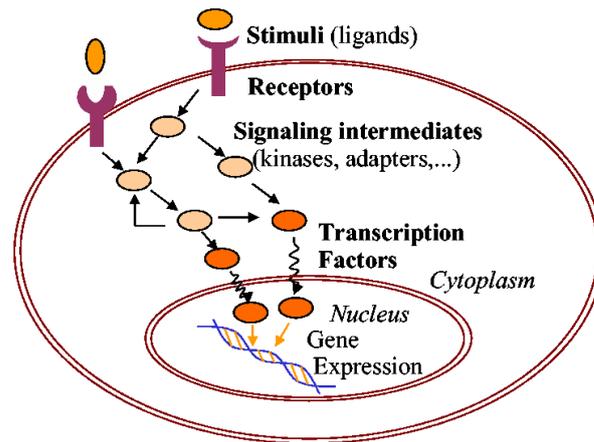
Figure 2.1: Simplified schema of a signal transduction network. Binding of ligand(s) to the receptor(s) typically lead to conformational changes in their cytoplasmatic part. This, in turn, triggers the activation of a number of intermediate components. Eventually, transcription factors are activated which control the gene expression.

other important biochemical modification is the conversion between an active, GTP-bound form to an inactive, GDP-bound form*. Molecules regulated by this mechanism are known as GTPases. Activators of the GDP →GTP conversions are the guanine nucleotide exchange factors, and promoters of the inverse reaction GTPase activating proteins (GAPs)[146]. In Section 2.2.1 the most famous GTPase, Ras, will be presented. Proteins without catalytic activity that play an important role by binding to other proteins are known as adaptors[146]. For example, an adaptor may bind simultaneously an enzyme and its substrate, allowing thus the former to act on the latter.

Sets of proteins activating subsequently each other are known as cascades. However, there are strong interactions (known as crosstalk) among the different cascades, giving rise to complex signaling networks that perform signal-processing tasks integrating information coming from different stimuli[28]. The complexity (and thereby the potential to process signals) is enhanced by feedback mechanisms. Spatial localization also increases the complexity of these networks: the cell is structured into different compartments, and a protein might be present only in a certain region of the cell (e.g. bound to the membrane). All these factors make signaling networks extremely complicated systems[10]. Therefore, even though much has been learned on the functioning of these networks over the last years, how this signaling machinery works as a whole, is still far from clear.

Proteins are built in a modular manner, as a composite of so-called domains. Each domain allows them to perform a specific task. For example, SH2 and PTB domains bind to so-called Tyrosine-based signaling motifs (TBSMs) when the TBSMs are phosphorylated, PH domains bind to phospholipids, and SH3 domains which bind to domains containing sequences of aminoacids rich in proline[146]. The interface between protein domains is built in

---

*Guanosine triphosphate (GTP) and Guanosine diphosphate (GDP) are purine nucleotides[272]

a modular fashion, with few interactions between the domains[212]. Therefore, these domains can be considered as the fundamental elements of signal transduction[199]. This property of the proteins, combined with a novel modeling approach based on it, will be exploited in Chapter 4 to define a construction kit of modules to set up models of signaling networks.

Interestingly, although eukaryotes systems are generally more complex, both prokaryotes and eukaryotes follow the same logic. Therefore, even though this work will focus on eukaryotes, the methodology and conclusions obtained here are extendible to prokaryotes.

In the following sections, two specific signaling systems from mammalian cells, namely the Epidermal Growth Factor (EGF) induced MAPK cascade and the T-cell Receptor (TCR) induced signaling network, will be presented. These two systems will serve as case studies for the different analyses used in this thesis.

2.2 Systems under Study in this Work

2.2.1 EGF induced MAPK cascade

An important family of receptors are the receptor tyrosine kinases (RTKs), which share many elements and mechanisms[231]. As the name says, RTKs possess a domain with tyrosine kinase activity in their intracellular part. When a ligand binds to a RTK it causes pairwise binding of the receptor proteins to create dimers, a process called dimerization, resulting in the activation of the receptor's kinase[231]. The kinase can then phosphorylate either the receptor itself or a substrate protein. The phosphorylated residues are binding sites for several proteins. RTKs are a large family of receptors for different ligands such as Hepatocyte Growth Factor (HGF), Epidermal Growth Factor (EGF) and Insulin, and they share to a large extent the proteins they bind to and activate (see Figure 1.1).

The Epidermal Growth Factor Receptor (EGFR) is the prototype of the EGFR family, the best-studied group of the RTKs, and perhaps the best understood cellular signaling system in mammalian cells. The EGF receptor can bind to several growth factors including EGF and TGF-α[283]. Activation of the EGFR can trigger responses that include growth and cell migration[270]. There is a tight connection between EGFR and cancer, as evidenced by the fact that EGFR is over-expressed in a wide variety of human tumors[270]. Therefore, the EGFR pathway has been intensively analyzed as a drug discovery target for cancer therapy, and some of the resulting drugs are currently in clinical development[43]. Since EGFR is both so well-known and important, many modeling efforts have been devoted to this system[20;29;135;234], recently reviewed by Wiley et al.[275].

An important process involved in the regulation of EGFR signaling is the internalization[274], a process in which the receptors are retrieved from the cell surface and moved into special compartments known as endosomes.

2.2.1.1 MAPK cascade

Among the main targets of the RTKs are the highly-conserved Mitogen-Activated Protein Kinases (MAPKs), which play a pivotal role in the transduction of signals in eukaryotes[37]. There are several families of MAPKs, and at least four expressed in mammals: ERK-1/2, JNK, p38 and ERK5[37]. MAPKs have different names, but they share the same mechanism of activation: each MAPK (see Figure 2.2), is phosphorylated at two points by another kinase - hence called MAPK kinase (MAPKK) ,- which is also activated through a double phosphorylation by another kinase - called MAPKK kinase (MAPKKK). There are also phosphatases which reverse these phosphorylation steps (see Figure 2.2).
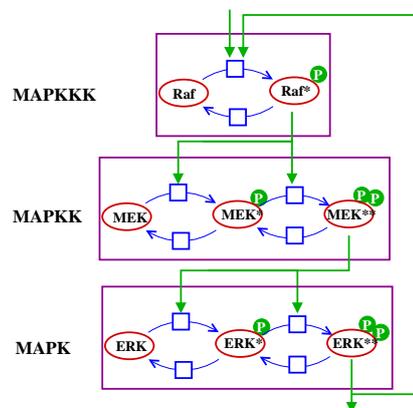
Figure 2.2: Structure of the MAPK cascade. the Raf/MEK/ERK cascade, the most prominent MAPK cascade, is used as example here. A positive or negative feedback from the last to the first module can be present.

In mammals, MAPK cascades are involved in the response to a wide range of stimuli, ranging from growth factors to stress, which result in the regulation of essential cellular processes such as differentiation, cell proliferation and survival[229]. How MAPKs are able to produce specific responses to different stimuli is an issue not fully understood yet. A particularly important MAPK cascade in the context of RTKs is the Raf/MEK/ERK-1/2 cascade. Its main activation pathway starts with binding of the adaptor Grb2 to phosphorylated RTK (or alternatively, binding of Shc to RTK and binding of Grb2 to Shc), and binding of the guanine nucleotide exchange factor (see Section 2.1) Sos to Grb2. Then Sos activates the GTPase Ras (see Figure 1.1). Subsequently, Ras binds to the MAPKKK Raf, targeting it to the membrane and thus leading to its activation. Active Raf activates MEK, which in turn activates ERK-1/2 (see Figure 2.2).

Both a model of the MAPK cascade and one for the EGF-induced MAPK will be analyzed with regard to their modularity in Section 3.4 and to their dynamic properties in Chapter 7.

2.2.2 T-cell receptor induced signaling

T-cells play a key role within the immune system: cytotoxic T-cells destroy cells infected by viruses or malignant cells, and helper T-cells coordinate the functions of other cells of the immune system[18]. Cytotoxic and helper T-cells are also known as CD8+ and CD4+ T-cells, respectively, since they express in one case the coreceptor CD8 and in the other the core-ceptor CD4. Loss or dysfunction, especially of CD4+ T-cells (as it occurs e.g. in the course of HIV infection or in immuno-deficiencies) has severe consequences for the organism and results in susceptibility to infections as well as in the development of malignancies.

The importance of T-cells for immune homeostasis is due to their ability to specifically recognize foreign, potentially dangerous, agents and, subsequently, to initiate a specific immune response that is aimed at eliminating them. T-cells detect foreign antigens by means of their T-Cell Receptor (TCR) which recognizes peptides only when presented on MHC (Major Histocompatibility Complex) molecules. The peptides that are recognized by the TCR are typically derived from foreign (e.g. bacterial, viral) proteins and are generated by proteolytic cleavage within the so-called antigen presenting cells (APCs). Subsequent to their production, the peptides are loaded onto the MHC-molecules and the assembled peptide/MHC-complex is then transported to the cell surface of the APC were it can be recognized by T-cells. The whole process of antigen uptake, proteolytic cleavage, peptide loading onto MHC, transport of the peptide/MHC complex to the surface of the APC and the recognition of the peptide/MHC-complex by the TCR is called antigen presentation and provides the molecular basis for the fine specificity of the adaptive immune response[143] (see Figure 2.3(a)).
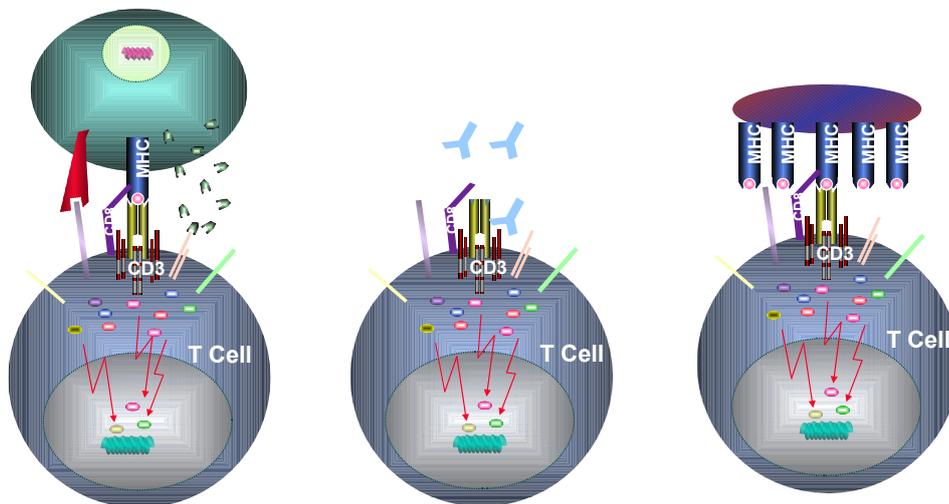
(a) Stimulation by APC   (b) Stimulation by antibodies   (c) Stimulation by streptamers

Figure 2.3: Different types of stimulation of T-cells. (a) Physiological stimulation by an antigen presenting cell, (b) stimulation by antibodies, and (c) stimulation by streptamers.

The binding of peptide/MHC to the TCR, and the additional binding of a different region of the MHC molecules to the coreceptors (CD4 or CD8) and the costimulatory molecule

CD28, initiate a plethora of signaling cascades within the T-cell. As a result, several transcription factors - most importantly, AP1, NFAT and NFκB - are activated. These transcription factors, in turn, control the cell's fate, e.g. whether it becomes activated and proliferates[118;143] or not (see Figure 2.4).

In laboratory studies, the use of antibodies to stimulate T-cells is widespread (see Figure 2.3(b)). For example, antibodies specific for the CD3 subunit of the TCR and/or for the coreceptor CD4/CD8 and the costimulatory molecule CD28 are routinely used. An advantage of the antibodies with respect to the use of APC is that one deals with a defined system (APCs are cells and as such highly variable) but at the price of using unphysiological conditions, since the antibodies have a much higher affinity for the receptors and coreceptors than the physiological ligands. A compromise solution engineered to join the advantages of both methods are the so-called streptamers (or tetramers[278]): several MHC/peptide monomers bound to a large molecule to form a complex†. (see Figure 2.4). The streptamers are particularly useful when used with transgenic mice such as the OT-1 mice. The T-cells of the OT-1 mice express all the T-Cell Receptor specific for ovalbumin (and the CD8 coreceptor, i.e., are cytotoxic T-cells). On the one hand, since the variety in the TCR allows T-cells to recognize virtually any foreign elements, OT-1 mice have a severely hindered immune system. On the other hand, they represent an ideal system for experiments, as the nature of the TCR is clearly defined. In Section 7.1.2 the effect of the use of these different stimuli on the dynamics of the signaling network will be addressed.

In contrast to RTKs, the TCR has no enzymatic activity and how it triggers signaling is still a topic of intense debate[41;55;157;263]. Different hypotheses have been formulated such as

- the heterodimerization model, which claims that TCR signaling can be initiated by the dimerization of TCR with the coreceptor[263],

- the pseudodimer model, which proposes that TCR engaged to an antigen can be crosslinked by the coreceptor to another TCR[263],

- the clustering model, where clustering of receptors allows kinases, which are bound to the receptors but inactive for esterical reasons, to cross-phosphorylate[190],

- the conformational model, where binding of a ligand promotes conformation changes in a certain unit of the TCR, triggering signaling[92], and

- the kinetic-segregation model, where the binding of ligand signaling is initiated in regions of close APC-TCR, what would exclude the 'big' phosphatase CD45, which constitutively dephosphorylates several proteins, allowing thus the kinases to act[42;55].

---

†Specifically, MHC/peptide monomers are biotinylated (bound to biotin). Biotin has a strong affinity for the large molecules avidin or streptavidin, which are used then build the complex.

No model seems to be conclusive, and the existence of coreceptor dependent and independent activation is still puzzling[263]. A combination of the models is also possible, since they are not exclusive[41]. Despite this lack of knowledge about the first - and arguably most important-step of T-cell activation, it is generally accepted that the first step involves the activation of the Src-family kinases p56lck (in the following termed Lck) and p59fyn (Fyn in the following), followed by the activation of ZAP70 (Zeta Associated Phosphoprotein of 70 kDa), triggering a number of signaling pathways[118].
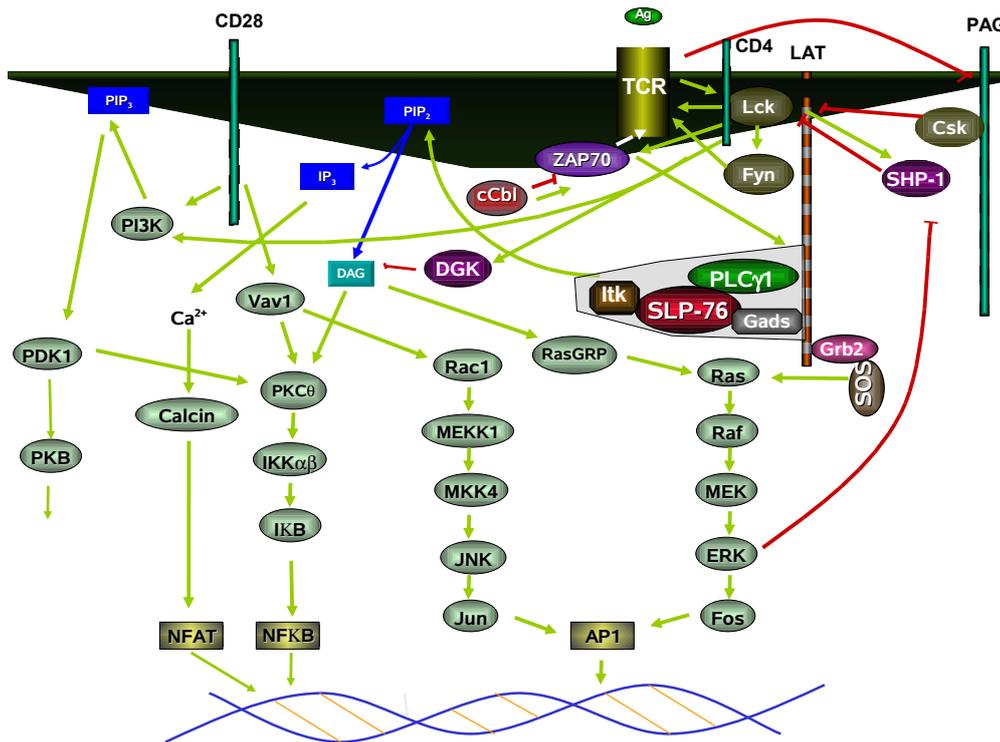
Figure 2.4: Schema of the main signaling paths triggered by T-cell receptor activation. A detailed description can be found in the appendix Section A.6

In the following, the biochemical steps involved in the TCR-induced signaling network will be briefly described (see Figure 2.4); for a detailed description the reader is referred to reviews[118;260] and the references therein. Additionally, a comprehensive logical model describing TCR-mediated signaling will be presented in Section 6.3, and a detailed documentation of all molecules and reactions involved can be found in the appendix, Section A.6.

- Upon binding of peptide/MHC to the TCR, the first main step in the TCR-mediated signaling cascade is, as stated above, the activation of Lck and Fyn, although the exact mechanism is still unclear.

  In resting T-cells, the major negative regulator of Lck, the kinase Csk (C-terminal Src kinase) is bound via a SH2-domain (see Section 2.1) to the constitutively tyrosine phosphorylated transmembrane adaptor protein PAG (Protein Associated with Gly-

cosphingolipid enriched microdomains) and consequently inhibits membrane-bound Lck by phosphorylating a C-terminal negative regulatory tyrosine residue of Lck[114].

Upon ligand binding, PAG is dephosphorylated by a so far unknown phosphatase, thereby leading to the detachment of Csk from PAG, and hence releasing Lck from the inhibitory effect of Csk. The release of Csk from PAG, together with the activity of the membrane associated tyrosine phosphatase CD45 (which dephosphorylates Lck on the same inhibitory residue that is phosphorylated by Csk), and the concomitant binding of the MHC molecule to the coreceptor CD4, lead to full activation of Lck.

After a few minutes, PAG is rephosphorylated[262], probably by Fyn, and subsequently Csk is re-recruited to PAG inhibiting Lck again.

Another important regulatory mechanism of Lck involves the phosphatase SHP-1. Lck activates SHP-1, which in turn dephosphorylates Lck at its positive regulatory site, resulting thus in a negative feedback. Additionally, ERK, downstream of Lck, can inhibit the effect of SHP-1 on Lck, creating a double negative feedback (Lck →...→ERK ⊣ SHP-1 ⊣ Lck) which has the net effect of a positive feedbackAltan-Bonnet and Germain[3]; Stefanova et al.[245].

• Activated Lck can phosphorylate Fyn (Fyn can probably also be activated in a Lck-independent, TCR-dependent manner[81]). Additionally, Lck phosphorylates the so-called ITAMs (Immunoreceptor Tyrosine-based Activation Motifs) that are present in the cytoplasmic domains of the TCR-complex (the latter if the TCR is close to Lck, i.e., if there is a concurrent activation of the TCR). Subsequently, the kinase ZAP70 binds to the phosphorylated ITAMs and, if Lck is active, becomes activated by Lck-mediated tyrosine phosphorylation. Thus, during the initial phase of signal transduction via the TCR, three tyrosine kinases become activated in a sequential manner, first Lck and Fyn and then ZAP70. Together, these three kinases propagate the TCR-mediated signal by phosphorylating a number of membrane associated and cytosolic signaling proteins.

• Active ZAP70 can phosphorylate LAT (Linker for Activation of T-cells), a second transmembrane adapter protein, at four different tyrosine residues. Subsequently, cytoplasmic signaling molecules containing SH2-domains, including the scaffolding proteins Grb2, Gads, and the lipid kinase PLCγ1 (Phospholipase gamma 1), can bind to phosphorylated LAT. Additionally, Grb2 binds to the nucleotide exchange factor Sos, and Gads to the adapter protein SLP-76. The latter, upon phosphorylation by ZAP70, can bind to the kinase Itk. Binding to SLP76 and additional phosphorylation by ZAP70 activates Itk. Finally, Itk phosphorylates and thereby fully activates PLCγ1.

• Activated PLCγ1 hydrolyzes phosphatidyl-inositol-4,5 biphosphate (PIP2), thereby

generating the second messenger molecules diacyloglycerol (DAG) and inositol trisphosphate (IP3)[118;260].

- IP3 mediates calcium flux. Calcium (together with calmodulin) activates the serine phosphatase calcineurin, which dephosphorylates the cytosolic form of the transcription factor NFAT (Nuclear Factor of Activated T-cells). The calcineurin-mediated removal of phosphate groups allows NFAT to translocate to the nucleus and to regulate gene expression.

- The second messenger DAG activates PKCθ and activates the nucleotide exchange factor RasGRP1. The amount of DAG is tightly regulated by the DAG kinases (DGKs), which degrade DAG into phosphatic acid[261].

- RasGRP1 and Sos (the latter if it is close to the membrane, that is, if it is bound to LAT by means of Grb2), can activate Ras, which in turn activates the Raf/MEK/ERK MAPK cascade.

- ERK, activated by the Ras/Raf/MEK cascade, activates Fos which, together with Jun, forms the fundamental transcription factor AP1.

- PKCθ is involved in the activation of the essential transcription factor NFκB (via phosphorylation and subsequent degradation of the NFκB inhibitor, IκB, by the PKCθ-activated IκB-kinase, IKK).

- Lck, in addition to the cascade described above, triggers the PI3K/PKB pathway that regulates many aspects of cellular activation and differentiation, particularly survival.

- The costimulatory molecule CD28 plays an important role in T-cell signaling. Its mechanism, however, is still unclear. Nevertheless, it has two main targets: PI3K and Vav1[175]. Vav1, in turn, acts on many elements such as PKCθ and Rac1[284]. This pathway will be discussed in detail in Section 6.3.

- The E3 ubiquitin ligase c-Cbl is important for shutting off TCR-mediated signaling processes by ubiquitination of key proteins, which are subsequently targeted for degradation[61]. One important target of c-Cbl is ZAP70; upon tyrosine phosphorylation of ZAP70, c-Cbl binds to ZAP70, leading to ZAP70's ubiquitination and degradation as well as to the downregulation of the TCR.

This signaling network as a whole will be analyzed using a boolean formalism in Section 6.3 and a portion of it, the TCR-induced MAPK cascade, will be dynamically modeled in Section 4.3.1 and studied with respect to its dynamics in Section 7.1.2.

Chapter 3

On the Modularity
of Signal Transduction Networks

The previous chapter illustrates the high complexity of signal transduction. In Chapter 1, it was introduced that a sound rationale to untangle it is a modular approach, where systems are decomposed into subunits.

This chapter is devoted to an analysis of the modularity of signal transduction networks from a system-theoretical perspective. Although the modularity of biological processes is generally accepted, a distinctive criterion for defining modules is still lacking. Different proposals, such as evolutionary conservation, robustness, and genetic co-expression have been suggested[276]. Section 3.1 reviews briefly different efforts towards unraveling the modularity of biochemical networks.

In Section 3.2, a novel criterion for the definition of modules, namely the absence of retroactivity in the connections between the modules, is proposed. This approach, inspired by systems theory, provides a theoretical framework to analyze signaling networks in a modular manner. Most approaches use a description of the system under study as a protein interaction network. The approach presented here, however, relies on the description of a signaling system as a biochemical network. Thus, it uses more refined information and acts on the kinetic formalism most mathematical models are set up with. Therefore, it is particularly convenient to decompose models of signaling networks to facilitate their analysis and provides a framework to set up modular models. The different situations that can lead to a retroactivity-free connection are first examined by means of the network theory[93] in Section 3.2.3. Subsequently, in Section 3.3 an algorithm to automatically detect modules connected in a retroactive-free manner will be outlined, and finally its applicability will be illustrated with several examples in Section 3.4.

3.1 An Overview on the Current View of Modularity

That biology is modular is a largely accepted notion. Specifically, the concept of functional units is widely prevailing[103;148;151]: functional units are entities whose interaction with their environment is significantly smaller than their internal interaction, and can be thus seen as semiautonomous modules. However, a general, unique definition of module is lacking[276]. For example, some propose modules to be a group of molecules chemically isolated from their environment or clustered according to graph-theory methods. Others, a set of elements active at a certain time scale or place, or connected in a statistically relevant fashion. Alternative criteria may involve to be evolutionary conserved or robust. Accordingly to this large list of definitions, there are an extensive number of works attempting to unravel the modularity of biochemical networks from all kind of perspectives.

3.1.1 Statistically relevant motifs

Since the concept of modularity can be applied to different levels of detail in a hierarchical manner[222], these subunits may comprise anything ranging between a single domain of a particular molecule to a whole organism. Usually, the simplest units are referred to as motifs, while larger components are named modules[276]. Motifs comprise normally no more than 2 or 3 proteins, appear repeatedly, and are characterized by their structure (topology) or dynamic properties (e.g. a switch, an amplifier, a filter, etc)[103;276]. Statistical analyses have uncovered that there are motifs which appear significantly often in signaling and regulatory networks which can be connected to specific functions[177;178;239].

In a modular system, the potential damage of perturbations are confined to a separable subunit. Therefore, It has been proposed that modularity combined with other design principles such as redundancy contributes to robustness. Importantly, this source of robustness is not incompatible with evolvability (i.e., the ability to evolve in order to adapt to new conditions), which can take place either by rewiring of the modules or by modifications inside the modules which do not affect directly other functional units[251]. In fact, modularly varying goals (different targets, set up of a certain combination of sub-targets, which change repeatedly), characteristic of biological evolution, when applied to evolving networks, lead to the generation of modular network structure and network motifs[131].

3.1.2 Graph-based analyses

A large number of works have addressed the properties of large biological networks, among them the modularity, using technics of graph theory. It could be shown that metabolic networks[210] as well as protein networks in yeast[125;216] and human[217] have a modular, hierarchical structure[14] using different clustering methods.

A remarkable effort towards the decomposition of networks into modules has been devel-

oped by Newman and colleagues[183;185], whose approach has also been applied by others to metabolic[100] and protein-protein networks[40]. Particularly appealing is their idea of defining the modularity as a mathematical formula, which is subsequently optimized[100;183]. In principle, one would intuitively quantify the modularity as a function of the number of connections *among* the modules with respect to the connections *within* the modules. This, however, would lead to the trivial solution that the modularity is maximum if all elements are in one single module. Therefore, one requires to somewhat 'normalize' the number of interconnections. Accordingly, their definition is based on the idea that the modularity is high if there are fewer edges than expected between modules, i.e., if there are fewer than in a randomly generated network. Therefore, the modularity Q is defined by Newman as the sum, over all pairs of vertices, of the differences between the network of interest and a random network regarding the number of connections inside the modules[183]. For a network containing n vertices and m edges decomposed into 2 modules the modularity reads

$$Q = \frac{1}{4m} \sum_{ij}^{n} (A_{ij} - \frac{k_i k_j}{2m})(s_i s_j + 1) \qquad (3.1)$$

where $s_i = 1$ if vertex i belongs to module 1 and $s_i = -1$ if it belongs to module 2, so that $\frac{1}{2}(s_i s_j + 1)$ is 1 if i and j are in the same module and 0 otherwise. The adjacency matrix A contains the information about the edges between pairs of vertices, i.e., $A_{ij} = 0$ if there is no connection between vertices i and j, 1 if there is one, etc. The number of edges between i and j expected in a random network is $\frac{k_i k_j}{2m}$, where k_i and k_j are the number of edges connected to i and j, respectively, and m is the total number of edges in the network ($m = \frac{1}{2}\sum_i k_i$). The leading factor $\frac{1}{4m}$ is just conventional for compatibility purposes with other definitions of modularity[100;183].

The works described here analyze large biological networks as graphs, and aim to unravel general structural principles. Even though we will take use of some of their ideas (in particular, the approach of Newman) in Section 3.3, the goal of this work is not to tackle such networks but rather, to decompose a given (model of a) signaling network into units in order to facilitate its analysis, providing thus a rationale for modular models. To that extent, it is connected to the approach of Ederer et al.[63], where also a kinetic model is analyzed by clustering methods. There, however, typical trajectories of the concentrations are used to cluster compounds into modules, while here the focus is on the network structure. There are also related ideas from the analysis of metabolic networks, also relying on a description of biochemical networks as in Equation 3.6, which we shall revise in the next section.

3.1.3 Insights from Metabolic Networks Analysis

Elementary flux modes[107] (EFMs) and extreme pathways[230], two very close concepts[142], are key instruments in the analysis of metabolic networks[142]. An EFM is a minimal sub-

network that can operate (i.e., present a flow) in steady state[142]. While their applicability for enlighting signal processing is possible[195;196] but limited (since they capture the mass-transfer rather than the information transfer[143], and are thus not applicable in important cases, see Chapter 6), they provide an interesting basis to define functional units: the ubiquitous cycles of activation/deactivation are elementary modes that can be seen as modules with regenerative capacity, connected via either shared compounds or enzymatic influence[17]. However, in cases where there is no flux (see Figure 3.6 and Section 3.6(a)), EFMs fail to identify signal process units. As will be discussed later, application of the concept of retroactivity captures not only the modules one would capture applying EFM but also where there is no flux, and is thus a more suitable approach.

Important efforts towards a modular understanding of biochemical networks have been performed in the field of Metabolic Control Analysis (MCA), which has proofed to be successful in the analysis of metabolic networks, and has been extended to signal transduction networks[129;136]. Importantly, MCA provides a framework for modular analysis. The approach was firstly restricted to modules which do not share mass flows[129], but later extended to modules which can share mass flows and have to fulfill conditions similar to the absence of retroactivity that is proposed and formalized in the next section.

3.2 The Absence of Retroactivity as a Criterion to Demarcate Modules

3.2.1 System-theoretical decoupling vs. absence of retroactivity

Since engineering sciences are used to work in a modular manner, it is tempting to approach the definition of biological modules from a technical perspective. From a *system-theoretical* point of view an interesting criterion might be the definition of elements where both the input and the output are unidirectional. This is actually the form in which most technical systems are devised, facilitating their analysis and design: for example, a thermometer is constructed in such a way that it receives information (in the form of thermic energy) about the temperature of a certain object, but it does not affect significantly the energy (and thus temperature) of the object itself.

Consider a signaling network as a general non-linear dynamical system described by a set of ordinary differential equations (ODEs) of the form

$$\dot{\vec{c}} = \frac{d\vec{c}}{dt} = f(\vec{c}, \vec{u}, \vec{p}), \qquad (3.2)$$

where $\frac{d\vec{c}}{dt} \in I\!R^n$ is the vector of the 'balances' of the concentrations c_i, $\vec{u}$ the vector of inputs and $\vec{p}$ the vector of parameters. A vector of outputs $\vec{y} = g(\vec{c})$ may also be defined. The goal

would be to decompose $\vec{c}$ into two sub-systems $\vec{c_1}$ and $\vec{c_2}$ so that

$$
\begin{aligned}
\dot{\vec{c}}_1 &= f_1(\vec{c}_1, \vec{u}, \vec{p}) \\
\dot{\vec{c}}_2 &= f_2(\vec{c}_1, \vec{c}_2, \vec{u}, \vec{p}).
\end{aligned}
\tag{3.3}
$$

If modules are connected in the same form as $\vec{c_1}$ and $\vec{c_2}$, they fulfill the requisite of independence postulated for functional units[103]: the behavior of $\vec{c_1}$ is only influenced by the input $\vec{u}$ and is independent of what is downstream of it. Importantly, decoupled units can be relatively straightforwardly analyzed by means of systems theory's tools. Actually, the decomposition into decoupled systems of the form of Equation 3.3 is a well studied problem in the field of systems theory[243].

Unfortunately, a hallmark of biochemical - in particular signaling - networks is the high degree of coupling. Hence, a clean decomposition in the form of Equation 3.3 is in most cases not possible. We therefore introduce a subtly different, more relaxed definition, which we shall call the absence of retroactivity, illustrated in Figure 3.1: two modules $\vec{c_1}$ and $\vec{c_2}$ are connected without retroactivity if there is no pair of elements (compounds), one in each module, which influence each other, i.e.

$$
\nexists (i,j) : \dot{c}_{1i} = f_{1i}(c_{2j}, \dots) \lor \dot{c}_{2j} = f_{2j}(c_{1j}, \dots) \text{ with } c_{1i} \in \vec{c_1}, c_{2j} \in \vec{c_2}.
\tag{3.4}
$$

The key difference with the previous concept of decoupling is that here, instead of the *global* decoupling between the modules as a whole imposed in Equation 3.3, we just require a *local* decoupling between all elements of one module and all elements of the other module. Note that both criteria work on the systems considered as autonomous, since neither the inputs $\vec{u}$ nor the outputs $\vec{y}$ are taken into account for the delimitation of the modules.
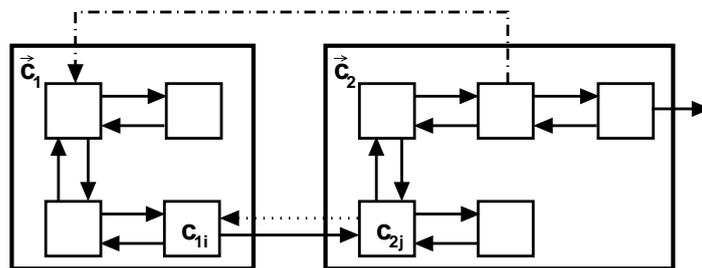
Figure 3.1: Schematic representation of the concept of retroactivity. If the state c_{1i} of the module $\vec{c_1}$ influences the submodule state c_{2j} of the module $\vec{c_2}$ (solid line), but the state c_{2j} does not directly influence c_{1i} (dotted line), the connection between $\vec{c_1}$ and $\vec{c_2}$ is free of retroactivity. A *unidirectional* feedback from another element $\vec{c_2}$ to $\vec{c_1}$ (dashed-dotted line) does not change the input/output behavior of module $\vec{c_2}$, but restricts the range of possible values for the input c_{1i}[220]. Note that hear each elemental block corresponds to one state associated to a differential equation.

Even the relaxed concept of absence of retroactivity (Equation 3.4) may not be completely fulfilled by biological systems. Therefore, the algorithm to decompose signaling networks which we shall see in Section 3.3 will rely on methods optimizing the modularity (i.e., finding the set of modules so that the number of retroactive connections among modules is minimized), rather than pursue a clean separation.

The particularities of biochemical systems should be taken into account. Specifically, a biochemical system is often described as a set of ordinary differential equations of the form

$$
\begin{aligned}
\frac{dc_1}{dt} &= \dot{c}_1 = N_{11}v_1 + N_{12}v_2 + ... + N_{1k}v_k + ... + N_{1m}v_m, \\
&\vdots \\
\frac{dc_i}{dt} &= \dot{c}_i = N_{i1}v_1 + N_{i2}v_2 + ... + N_{ik}v_k + ... + N_{im}v_m, \\
&\vdots \\
\frac{dc_n}{dt} &= \dot{c}_n = N_{n1}v_1 + N_{n2}v_2 + ... + N_{nk}v_k + ... + N_{nm}v_m,
\end{aligned}
\tag{3.5}
$$

or in matrix form

$$
\frac{d\vec{c}}{dt} = \dot{\vec{c}} = N\vec{v},
\tag{3.6}
$$

which is a special form of Equation 3.2. Here, $\vec{v}(\vec{c}, \vec{u}, \vec{p}) \in \mathbb{R}^m$ is the vector of the m reactions, and $N \in \mathbb{R}^{n \times m}$ the stoichiometric matrix (see e.g.[107]). This structure of the differential equations will be helpful in Section 3.3 to cleanly characterize, from a biochemical point of view, the coupling among modules.

The structure of Equation 3.6 emerges naturally from a description of signaling networks following the network theory, which is also a convenient framework for a modular approach. We shall hence introduce in the next section the network theory and, thereafter, discuss in Section 3.2.3 the different cases which lead to the absence of retroactivity.

3.2.2 Network theory and biological systems

A suitable frame for developing modular models is provided by the network theory introduced by Gilles[93]. Systems are described as a combination of two types of elementary units: *components*, which have storages of physical quantities and *coupling elements*, which describe the interactions between the components. These elements can be aggregated into a single elementary unit (corresponding to the modules introduced in the previous section*) on a higher level, which can be again described by means of components and coupling elements, leading to a hierarchical structure[167]. Components and coupling elements are connected by two types of signal vectors: potential vectors, which are outputs of components and inputs of

---

*Note that in Section 3.2.1 (e.g. in Figure 3.1) a unit was assigned to each component, and now a unit is assigned to both the components and coupling elements.

coupling elements, and current vectors, which are outputs of coupling elements and inputs of components. For example, in a chemical network the compounds would be the components, the reactions the coupling elements, potential vectors would carry information about the concentrations from the compounds to the reactions and current vectors would bring information about the rates back to the compounds (see Figure 3.3(a)), leading to a system of differential equations of the form of Equation 3.6. The application of the network theory to biochemical systems leads to a modular modeling concept introduced elsewhere[148].

One argument supporting the application of the network theory to cellular pathways is the proposed hierarchy of biological systems[148;156]. Actually, this hierarchical structure can be represented similarly for biological systems and chemical processes (see Figure 3.2). If we consider a human body, we can divide it into different systems which fulfill different tasks (e.g. the digestive system, the locomotive system, etc.), connected mainly (but not only) by blood vessels. Each of these systems can be described as a sum of organs connected also mainly by blood vessels. Organs are made up of several tissues, set up of cells. The coupling between cells takes place by means of exchange of different substances as well as by direct contact between the cells. The machinery of a cell can be decomposed into functional units which perform different functions. These biochemical pathways are connected by common compounds. Finally, each of these modules can be decomposed into molecules which interact by means of molecular interactions or reactions (see Figure 3.2). This work focuses on the last two levels of detail.
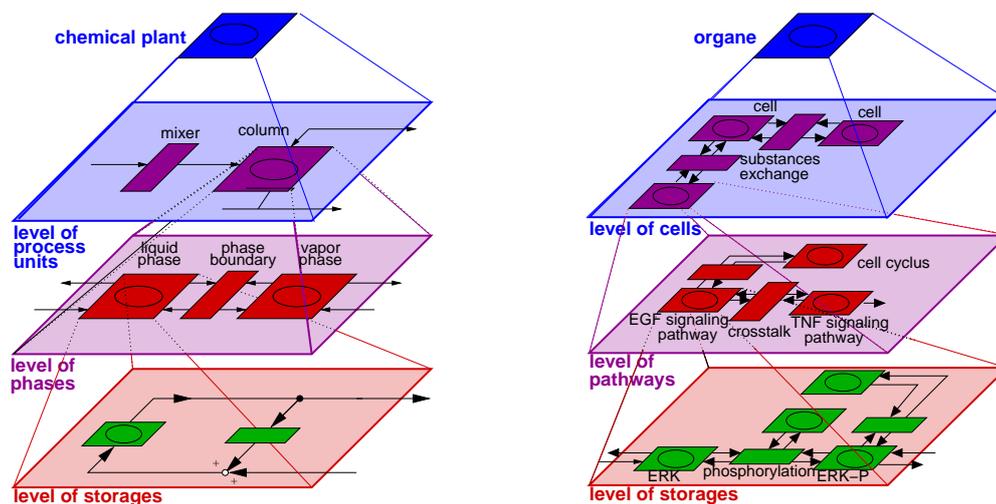
Figure 3.2: Hierarchical structure of biological systems. Analogously to chemical processes[167], biochemical systems are structured in a hierarchical manner. The network theory is a convenient framework for both kind of systems.

3.2.3 The absence of retroactivity in biochemical systems

Consider the simple general schema depicted in Figure 3.3(a), which represents one reaction (coupling element) r and three compounds (components), A, B and C, involved in the reaction r, according to the network theory. If one of the potential or current vectors can be neglected, the system shows a junction free of retroactive effects. But, under which conditions can a current vector (rate) or a potential vector (concentration) be neglected? In the following we discuss some typical simple cases.

Neglect of a potential. A potential can be neglected if the concentration of one of the compounds, say C, does not affect the reaction rate, which corresponds to neglect vector 1 in Figure 3.3(a). An example is an irreversible reaction, where the product does not affect the reaction rate. Hence, an irreversible reaction of A and B to give C would be represented as in Figure 3.3(b). There are some common irreversible reactions in biochemistry.
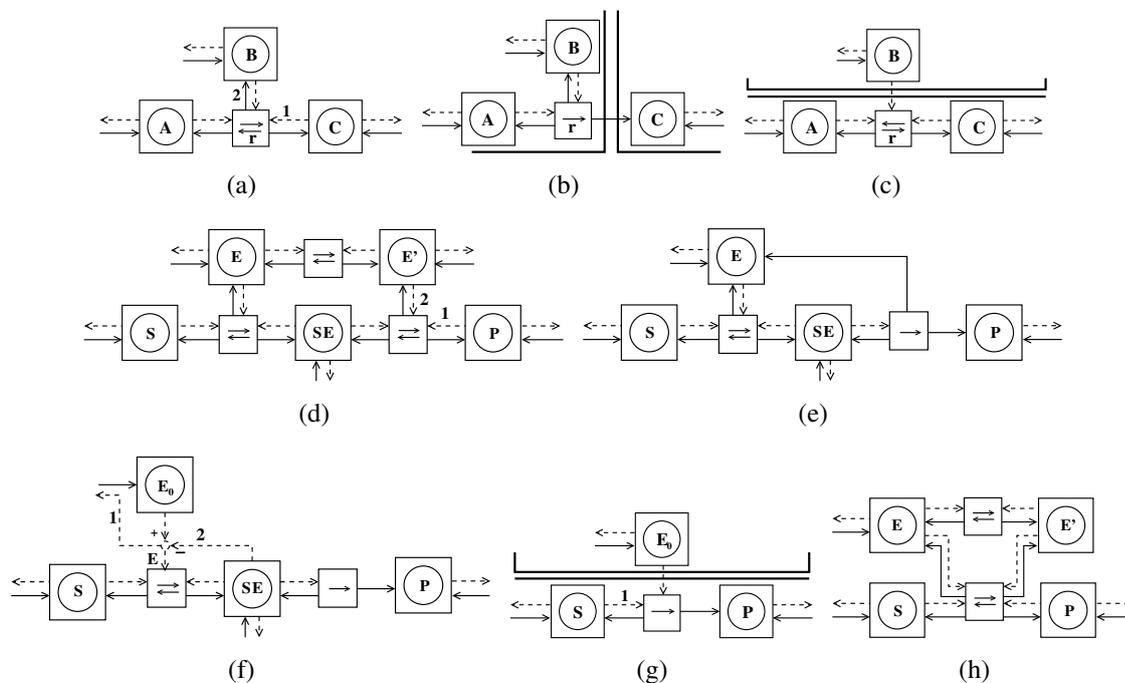
Figure 3.3: Representation of different reactions schemes according to the network theory. Dashed lines represent potential (concentration) vectors, solid thin lines current (rates) vectors, and solid thick lines the borders of the modules. (a) general case; (b) neglect of a potential; (c) neglect of a current; (d) system defined by Equations 3.7 and 3.8; (e) system defined by Equation 3.10; (f) same system as in (e) but with a change of variable $E_0 = E + SE$; (g) system defined by Equation 3.11; (h) system defined by Equations 3.15 and 3.16.

Neglect of a current. A retroactive-free connection by neglect of a current is possible if a compound influences a reaction rate, but the reaction rate does *not* influence this component (i.e., if the vector 2 in Figure 3.3(a) can be neglected), leading to the system depicted in Figure 3.3(c). One possibility would be if a compound is consumed or produced in a reaction, but the amount involved in the reaction is negligible compared to the total amount. For

example, if one of the substrates is in excess, say $A \ll B$, then the amount of B consumed in the reaction will be negligible compared to the total amount of B, leading to a unidirectional connection.

If we consider the general case where a compound S is transformed into P, by reaction with another compound E, being E regenerated in an additional step, as defined by the equations

$$E' \rightleftharpoons E \tag{3.7}$$

and

$$S + E \rightleftharpoons SE \rightleftharpoons P + E', \tag{3.8}$$

we arrive at the schema depicted in Figure 3.3(d). The system is highly interconnected, without unidirectional connections. If the second step of the second reaction (Equation 3.8) is considered irreversible we obtain

$$S + E \rightleftharpoons SE \rightharpoonup P + E' \tag{3.9}$$

instead of Equation 3.8. The representation of the new system is obtained by deleting the vectors 1 and 2 in Figure 3.3(d). In this system, there is a unidirectional connection defined by the irreversible step, but the connection between $E \leftrightarrow E'$ and $S \leftrightarrow P$ has still retroactivity (see Figure 3.3(d)). If, additionally, $E = E'$, the system

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} SE \overset{k_2}{\rightharpoonup} P + E \tag{3.10}$$

is obtained, which is shown in Figure 3.3(e) and represents the irreversible conversion of S into P catalyzed by an enzyme E. The reactions are normally described according to the mass action law. Defining a new variable $E_0 = E + SE$ we obtain an alternative representation (Figure 3.3(f)). Analyzing this schema we can see that a connection free of retroactivity from the enzyme to the reaction can be achieved if:

(i) The reactions that influence E_0 but are not represented in Figure 3.3(f) are not influenced by E, which is equivalent to neglect the vector 1 in Figure 3.3(f). This is actually the case introduced above of absence of retroactivity by an irreversible reaction.

(ii) The dynamics of the compound SE can be neglected (i.e., if $dc_{SE}/dt \approx 0$, which means that the vector 2 in Figure 3.3(f) is negligible). This approximation is known as the quasi-steady-state assumption, and leads to the reduced system (see for example[236])

$$S \rightarrow P, \tag{3.11}$$

following the reaction rate r the classical Michaelis Menten equation[235]

$$v = \frac{V_{max} \cdot S}{K_m + S} = \frac{k_2 \cdot E_0 \cdot S}{K_m + S}, \tag{3.12}$$

where $K_m = (k_{-1} + k_2)/k_1$. We obtain thus a connection free of retroactivity by absence of a current vector, as represented in Figure 3.3(g). If, additionally, the enzyme is saturated by the substrate ($K_m \ll S$), then the reaction rate r becomes

$$r = k_2 E_0 \tag{3.13}$$

and the system can be represented as in Figure 3.3(g) deleting the vector 1. We obtain hence an additional connection free of retroactivity between the reaction r and the substrate S.

The assumption $dc_{SE}/dt \approx 0$ is correct for the system defined in Equation 3.10 if $\varepsilon \ll 1$, where $\varepsilon = E_0/(K_m + S_0)$, being E_0 and S_0 the total concentration of E and S, respectively[236]. This condition is fulfilled if $E_0 \ll S_0$ and if $E_0 \ll K_m$. $E_0 \ll S_0$ (much less enzyme than substrate, a usual situation in many *in vitro* experiments) is the usual assumption for the application of Michaelis Menten equation.

The condition $E_0 \ll K_m$ can be rewritten as $E_0 k_1 \ll k_2 + k_{-1}$. Since k_1 is the kinetic constant for the formation of the complex SE, and k_{-1} and k_2 the kinetic constants for the dissociation of the complex SE (see Equation 3.10), this condition can be interpreted as the decomposition of SE being much faster than the formation of SE. This situation is analogous to many electrical measuring devices, e.g. a thermocouple. In a thermocouple, a difference of temperature generates a voltage V, which in turn produces a current I through a conductor. A very high value is given to the resistance R and therefore the current is very low ($V = IR$). This current provides a measurement of the voltage that does not affect the source of the signal. In the case of an enzymatic reaction where $E_0 \ll K_m$, the reaction rate (or the amount of product) is a "measurement" of the concentration of the enzyme, but, since $E_0 \ll K_m$, there is a high resistance against the consumption of the enzyme, which is thus not affected by its "measuring device".

The Michaelis Menten expression (Equation 3.12) is widely used for enzymatic reactions without considering whether the assumptions described above are fulfilled or not.

If the Equation 3.7 of the general case can be neglected, but the second term of the Equation 3.8 can not be considered irreversible, we obtain a system defined by the equation

$$S + E \rightleftharpoons SE \rightleftharpoons P + E \tag{3.14}$$

which, under the quasi-steady-state assumption, can be transformed into a system with a unidirectional connection by neglect of a current as depicted in Figure 3.3(c).

On the other hand, the neglect of the complex *SE* in the general case (Equations 3.7 and 3.8) leads to the system

$$E' \rightleftharpoons E \tag{3.15}$$

$$S + E \rightleftharpoons P + E'. \tag{3.16}$$

As can be seen in Figure 3.3(h) this system is still highly interconnected.

In the following section, we shall formalize these ideas and devise an algorithm to realize them.

3.3 Automatic Identification of Modules Considering the Absence of Retroactivity

In Section 3.2.3 it was introduced how, if one of the potential or current vectors can be neglected, the system shows a junction free of retroactive effects. Additionally, typically biochemical cases were analyzed which lead to the neglect of a current vector (rate) or a potential vector (concentration). In this section, we formalize the concept of absence of retroactivity and show how this can be implemented in an algorithm to detect modules considering the concept of absence of retroactivity.

3.3.1 Formalizing the concept of absence of retroactivity into mathematical equations

As described in the previous Section 3.2, there are two cases which can lead to a connection free of retroactivity: neglect of a potential and neglect of a current. Therefore, the first step towards a mathematical formalization of the concept of retroactivity is to determine which reactions influence which storages and vice versa.

3.3.1.1 Neglect of a current

Recall from Equation 3.6 that the balance of the concentration of a storage c_i is a function of the reaction rates,

$$\frac{dc_i}{dt} = \dot{c}_i = N_{i1}v_1 + N_{i2}v_2 + \ldots + N_{ik}v_k + \ldots + N_{im}v_m.$$

Therefore, a reaction v_u does not influence *significantly* a storage c_i (see Figure 3.4) if its contribution to the balance is negligible, that is to say, if

$$N_{iu}v_u \ll \sum_{k=1}^{m} N_{ik}v_k \Rightarrow \frac{N_{iu}v_u}{\sum_{k=1}^{m} N_{ik}v_k} = \widehat{g}_{iu}^{c} \ll 1. \tag{3.17}$$

This definition might lead to numerical problems since in steady state $\frac{dc_i}{dt} = \sum_{k=1}^{m} N_{ik}v_k = 0$. Using the absolute values would circumvent this problem; we use thus

$$\frac{|N_{iu}v_u|}{\sum_{k=1}^{m} |N_{ik}v_k|} = g_{iu}^c \ll 1, \tag{3.18}$$

which is a more strict condition than Equation 3.17. Therefore, the time-dependent function g_{iu}^c defines the effect of the reaction v_u on a storage c_i.
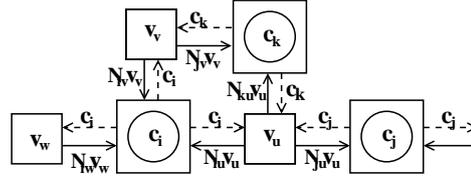
Figure 3.4: Representation of the conditions leading to a retroactive-free connection. Dashed lines represent potential (concentration) vectors, and solid thin lines current (rates) vectors.

3.3.1.2 Neglect of a potential

Since in the general case the reaction rates $v_u(\vec{c}, \vec{p}, \vec{u})$ are *not* a linear function of the concentrations, the effect of a storage c_j on a reaction v_u can be estimated by the derivative

$$\varepsilon_{uj}^* = \frac{\partial v_u}{\partial c_j} \tag{3.19}$$

This derivative is known as (unscaled) elasticity in the field of Metabolic Control Analysis[107], and we adopt this nomenclature here. For example, in the case of a reaction v_1

$$c_1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} c_2 \tag{3.20}$$

following simple mass action law kinetics, $\varepsilon_{11}^* = k_1$ and $\varepsilon_{12}^* = -k_2$.

However, the derivative alone does not capture the full effect of a component on a reaction. For example, if $c_2 = 0$, then the actual effect of c_2 on v_1 is zero, but $\varepsilon_{12}^* = k_2$. Therefore, it would be more convenient to use the term $\varepsilon_{uj}^* c_j$. Accordingly, a storage c_j does not affect significantly a reaction v_u (see Figure 3.4) if its effect is negligible when compared to that of the rest of storages, i.e.

$$|\varepsilon_{uj}^*|c_j \ll \sum_{k=1}^{n} |\varepsilon_{uk}^*|c_k \Rightarrow \frac{|\varepsilon_{uj}^*|c_j}{\sum_{k=1}^{m} |\varepsilon_{uk}^*|c_k} = g_{uj}^p \ll 1. \tag{3.21}$$

The time-dependent function g_{uj}^p defines thus the effect of a storage c_j on a reaction v_u. Note that, as in the case of the neglect of a current g^c, we use absolute values. If we divide

numerator and denominator by v_u we obtain

$$g^p_{uj}(t) = \frac{\frac{|\varepsilon^*_{uj}|c_j}{|v_u|}}{\sum_{k=1}^m \frac{|\varepsilon^*_{uk}|c_j}{|v_u|}} = \frac{|\varepsilon_{uj}|}{\sum_{k=1}^m |\varepsilon_{uk}|}, \tag{3.22}$$

where $\varepsilon_{uj}(t)$ is the scaled elasticity, $\varepsilon_{uj} = \frac{\partial v_u}{\partial c_j}\frac{c_j}{v_u} = \frac{\partial v_u/v_u}{\partial c_j/c_j}$ [107].

To sum up, the information about the retroactivity is stored in the time-dependent matrices $g^p : \mathbb{R}^+ \to \mathbb{R}^{n \times m}$ and $g^c : \mathbb{R}^+ \to \mathbb{R}^{m \times n}$.

3.3.1.3 Some considerations on the functions g^c and g^p and the structural retroactivity

The values of g^c_{ij} or g^c_{uj} are strictly zero if there is a *structural* absence of retroactivity. For example, a strictly irreversible reaction will lead to $N_{iu} = 0 \Rightarrow g^c_{iu}(t) = 0$ between the reaction u and the product i, and an enzymatic reaction modeled with Michaelis Menten kinetics will result in $\varepsilon_{uj} = 0 \Rightarrow g^p_{uj} = 0$ between the enzyme j and the reaction u.

However, in many other cases, g^c or g^p might not be strictly zero but have very low values. For those cases, a criterion determining what is low enough is required. Simple criteria could be that the maximal or the average value might be lower than a certain threshold θ, that is to say

$$\max_t (g^c_{uj}(t)) < \theta \tag{3.23}$$

or

$$\frac{1}{\Delta t} \int_0^{\Delta t} g^c_{uj}(t) dt < \theta \tag{3.24}$$

at a characteristic trajectory. Alternatively, one could perform more exhaustive analyses using e.g. Monte Carlo methods to explore the parameter space. Note that to determine the rigorous, stricter, structural absence of retroactivity only the structure of the network is needed, while for the approximate, kinetic-dependent absence of retroactivity a parameterized model and a particular 'experimental (simulation) condition' has to be defined.

3.3.1.4 Definition of the retroactivity matrix

Once a certain criterion has been applied, the matrices $R^p \in \{0,1\}^{n \times m}$ and $R^c \in \{0,1\}^{m \times n}$ can be obtained which define which storages affect which reactions and vice versa. Now, at the connection between a storage c_i and a reaction v_j following cases are possible:

$R^p_{ij} = 1,\ R^c_{ji} = 1$	retroactive connection
$R^p_{ij} = 1,\ R^c_{ji} = 0$	unidirectional connection (neglect of current)
$R^p_{ij} = 0,\ R^c_{ji} = 1$	unidirectional connection (neglect of potential)
$R^p_{ij} = 0,\ R^c_{ji} = 0$	absence of connection

(3.25)

Importantly, if one considers only the *structural* retroactivity, since $N_{ij} = 0 \Rightarrow R_{ij}^c = 0$ and $\varepsilon_{ij} = 0 \Rightarrow R^p = 0$, R^c and R^p correspond to the matrices expressing the occupancy of N and ε (what we shall call the indicator matrices N^I and ε^I) , respectively:

$$R_{ij}^c = N_{ij}^I = \begin{cases} 0 & \text{if } N_{ij} = 0 \\ 1 & \text{else} \end{cases} \tag{3.26}$$

and, analogously,

$$R_{ij}^p = \varepsilon_{ij}^I = \begin{cases} 0 & \text{if } \varepsilon_{ij} = 0 \\ 1 & \text{else} \end{cases} \tag{3.27}$$

From now on we shall concentrate on the structural case, but the same arguments can be applied to the general case.

Now, defining a matrix $R \in \{0,1,2,3\}^{n \times m}$, which we shall call the retroactivity matrix , so that $R_{ij} = R_{ij}^p + 2R_{ji}^c$, we would obtain the information about the retroactivity in a compact manner:

$$\begin{array}{|c|c|} \hline R_{ij} = 0 & \text{absence of connection} \\ \hline R_{ij} = 1 & \text{unidirectional connection (potential)} \\ \hline R_{ij} = 2 & \text{unidirectional connection (current)} \\ \hline R_{ij} = 3 & \text{retroactive connection} \\ \hline \end{array} \tag{3.28}$$

3.3.2 Implementation into an algorithm

The next step should be to try to demarcate the modules in such a way that the number of retroactive connections (where $R_{ij} = 3$) among modules is minimized (ideally zero), and maximized inside the modules. Considering the matrix $R^r \in \{0,1\}^{n \times m}$ where

$$R_{ij}^r = \begin{cases} 1 & \text{if } R_{ij} = 3 \\ 0 & \text{else,} \end{cases} \tag{3.29}$$

clustering technics can be applied, as those introduced in Section 3.1.2. The methods of Newman and colleagues, relying on the maximization of a mathematical value of the modularity (see Section 3.1.2), seem particularly suited for this task, since we are trying to minimize the number of retroactive connections (recall Section 3.2.1).

However, these approaches consider mostly (undirected) interaction graphs of networks, but here we are dealing with a bipartite graph (since there are 2 kinds of nodes: storages and reactions). Unfortunately, clustering algorithms for bipartite graphs are much less developed. A detour to circumvent this problem would be to somewhat convert the information about the relationships between the storages and reactions coded in R into a quadratic matrix defining the connections among one type of elements (storages). Thereby we would move from a

bipartite to a monopartite graph.

A natural monopartite graph would be one reflecting the reciprocal influence among the compounds. A compound c_i does not influence directly a compound c_j if there is no connection from c_i to c_j through any reaction. The influence of c_i on c_j via the reaction v_v is determined by $R^c_{jv} \cdot R^p_{vi}$: the influence of c_i on v_v (R^p_{vi}) multiplied by the influence of v_v on c_j (R^c_{jv}). Thus, extending this argument to all reactions, one gets that the influence of c_i on c_j reads

$$R^c_{j1} \cdot R^p_{1i} + R^c_{j2} \cdot R^p_{2i} + ... + R^c_{jm} \cdot R^p_{mi} = \sum_{k=1}^{m} R^c_{jk} \cdot R^p_{ki}, \tag{3.30}$$

and thus, if this expression is equal to zero, there is no influence of c_i on c_j. And for the structural case, from Equations 3.26 and 3.27,

$$\sum_{k=1}^{m} N^I_{jk} \cdot \varepsilon^I_{ki} = 0 \tag{3.31}$$

There is a close relationship between this expression and the Jacobian $J \in \mathbb{R}^{n \times n}$ with

$$J_{ij} = \frac{\partial f_i}{\partial c_j} \tag{3.32}$$

of a system of differential equations in the form of Equation 3.2. The sign of J_{ij} informs whether c_j has a direct positive or negative influence on c_i, and a matrix considering the sign of these entries can be seen as the adjacency matrix of the underlying interaction graph[143]. The relation to Equation 3.31 is simply obtained by deriving f_i with respect to c_j. From Equation 3.6 it results

$$\frac{\partial f_i}{\partial c_j} = N_{i1}\frac{\partial v_1}{\partial c_j} + N_{i2}\frac{\partial v_2}{\partial c_j} + ... + N_{ik}\frac{\partial v_k}{\partial c_j} + ... + N_{im}\frac{\partial v_m}{\partial c_j}, \tag{3.33}$$

that is

$$J_{ij} = N_{i1}\varepsilon^*_{1j} + N_{i2}\varepsilon^*_{2j} + ... + N_{ik}\varepsilon^*_{ij} + ... + N_{im}\varepsilon^*_{mj}, \tag{3.34}$$

or in compact manner

$$J = N\varepsilon^*. \tag{3.35}$$

Therefore, for the structural case that we are considering, the indicator matrix of the Jacobian $J^I \in \{0,1\}^{n \times n}$,

$$J^I_{ij} = \begin{cases} 1 & \text{if } J_{ij} = \sum_{k=1}^{m} N_{ik} \cdot \varepsilon_{kj} \neq 0 \\ 0 & \text{else} \end{cases} \tag{3.36}$$

would provide a starting point for these algorithms.

However, one still needs to include the criterion of the retroactivity. We are trying to define modules so that the number of retroactive connections are maximized inside the modules and

minimized among the modules. Therefore, for the decomposition purpose, the presence of a unidirectional connection is equivalent to no connection at all. The connection between two compounds c_i and c_j is retroactive if $J_{ij}^I = J_{ji}^I = 1$. If $J_{ij}^I = J_{ji}^I = 0$, there is no connection between the elements, and if $J_{ij}^I = 1$ and $J_{ji}^I = 0$ there is an unidirectional connection from c_i to c_j (involving either the neglect of a current or of a potential). The retroactivity can thus be captured via a symmetric matrix $J^{IR} \in \{0, 1\}^{n \times n}$, so that

$$J_{ij}^{IR} = J_{ji}^{IR} = \begin{cases} 1 & \text{if } J_{ij}^I = J_{ji}^I = 1 \\ 0 & \text{else} \end{cases} \tag{3.37}$$

This has also an advantage for applying Newman's algorithms, since they are devised for symmetric matrices. We will use thus the Newman definition for Modularity (Equation 3.1) applied to J^{IR} and we will try to optimize it using different methods, which are succinctly described in the next section.

3.3.2.1 Optimization algorithms

Let us recall that we try to optimize the modularity, defined as the number of edges within modules with respect to the number of edges within modules expected for a random network. This concept can be defined mathematically as in Equation 3.1 [183], which reads for a network decomposed in 2 modules

$$Q = \frac{1}{4m} \sum_{ij}^{n} (A_{ij} - \frac{k_i k_j}{2m})(s_i s_j + 1) \tag{3.38}$$

and in matrix form

$$Q = \frac{1}{4m} \vec{s^T} \cdot B \cdot \vec{s}, \tag{3.39}$$

with $s_i = 1$ if i belongs to module 1 and $s_i = -1$ if it belongs to module 2. A number of algorithms have been used to optimize the modularity of Equation 3.1. The results of the different algorithms were compared for the examples discussed below (see Section 3.4), and for each case the best results were taken. Importantly, the algorithms as implemented here, and as should be according to the ideas of Newman, do not require *a priori* information about the number of modules, they rather propose a number and how the nodes are distributed along the modules. The methods will be presented here succinctly; for a detailed description of them, the reader is referred to the original papers cited and to [87].

Leading Eigenvector method   operates as follows: first, it separates the network into two modules. Subsequently, each module is further decomposed into two modules, and so on. The procedure stops when, upon a certain decomposition, the modularity does not increase [183]. In

each iteration, to determine which elements belong to which module, the eigenvalues † λ of the matrix B (see Equation 3.39) are evaluated. The rationale reads as follows: Let us write $\vec{s}$ as a linear combination of the normalized eigenvectors u_i so that $s = \sum_{i=1}^{n} a_i \vec{u}_i$ with $a_i = \vec{u}_i^T \vec{s}$. Then, from Equation 3.39 it follows

$$Q = \frac{1}{4m} \sum_{i=1}^{n} (\vec{u}_i^T \cdot \vec{s}) \lambda_i. \tag{3.40}$$

Since the eigenvector u_{max} associated to the largest eigenvalue (λ_{max}) has the heaviest weight on Equation 3.40, we define s as parallel to u_{max} as possible, to maximize the product $\vec{u}_{imax}^T \cdot \vec{s}$. Since s_i can only be +1 or -1, we define $s_i = +1$ (i.e., element i belongs to module 1) if the corresponding element of u_{max} is positive and $s_i = -1$ (i belongs to module 2) if it is negative [183]. This procedure is applied recursively then to each of the modules obtained: for each submodule a new objective function Q is calculated and maximized.

Multiple Eigenvector method can be considered an extension of the previous method (leading eigenvector). Here, one tries to use the information about several eigenvalues. Importantly, this method allows to determine all modules in one single step [184].

Extremal optimization is a method [13;26] that considers a population of individuals, each of them with a certain fitness. A new fitness is assigned to the individual with the smallest fitness, and subsequently a certain global fitness of the population is computed. Since a change in an individual affects the whole population, the global fitness of the population may increase (and then the change is updated) or decrease. In case of decrease, one would in principle not accept the change. However, to avoid getting trapped in local minima, those negative changes are accepted with low frequency [13].

The method can be naturally applied to find modules in networks [62]: one considers each node as an individual and its fitness, roughly speaking, as the number of edges connecting it with other members of the modules related with its total number of edges it is bound to. The global fitness of the population, which is the objective function, would correspond to the modularity as defined in Equation 3.38.

Simulated annealing [140] is a method inspired by thermodynamics used to solve optimization problems in a wide range of fields (e.g. parameter estimation). It combines a relatively high efficiency with the ability to avoid local minima. It can be straightforwardly applied to maximize the modularity, by simply defining Equation 3.38 [100] as cost function.

---

†for a certain matrix B, if there is a vector $\vec{v} \neq 0$ such that $B \cdot \vec{v} = \lambda \cdot B$, then λ is an eigenvalue, and $\vec{v}$ an eigenvector of B

Table 3.1: Types of connections between two species as a function of J^{IR} and N^{CI}. All possible dependencies between 2 storages i and j can be unambiguously determined by the values of J^{IR}_{ij}, J^{IR}_{ji}, N^{CI}_{ij} and N^{CI}_{ji}.

J^I_{ij}	0	0	1 0	1 0	1	1
J^I_{ji}	0	0	0 1	0 1	1	1
N^{CI}_{ij}	0	1	0 0	1 1	0	1
N^{CI}_{ji}	0	1	0 0	1 1	0	1
(Bio) Chemical Interpretation	No connection	Products of common irrev. reaction	Control by potential (enzyme)	Control by current (irrev. reaction)	Reciprocal control by potential (enzyme)	Coupled storages (rev. reaction)

3.3.2.2 Characterization of the connections among the modules

After applying a modularity analysis, one would like to know what kind of connections bind the different modules. This information is, however, not present in the matrix J^I alone. One can obtain this information from the matrices $R^p(\varepsilon^*)$ and $R^c(N)$. Alternatively, one can deduce the nature of the connections from J^I and N. The latter has the advantage that it is not required to compute ε^* if it is not available (J^I is available from the previous steps and N is easily obtainable). Additionally, one remains in the monopartite (concentrations) description and does not need to go back to the bipartite (concentrations+reactions) formalism. The matrix $N^C = N(-N)^T$ is a symmetric matrix so that $N^C_{ij} = N^C_{ji} \neq 0$ if there is a mass flux between the species i and j, and is 0 otherwise[69]. Considering conjunctly its indicator matrices $N^{CI} \in \{0,1\}^{n \times n}$ and J^I the different cases can be retrieved (see Table 3.1). This will allow us to characterize the connection among the modules, as will be shown in Section 3.4.

3.3.2.3 Integration of the algorithm into ProMoT

The identification of the modules underlying a given model provides useful insights into its structure. More useful would be, however, to allow one then to subsequently analyze the system in a modular manner. For example, one would like to consider only one module in isolation, or the combination of several of them, eventually testing different variants considering different connexions between modules, a reduced version of one of the modules, etc. These analyses are the topic of Section 7. Rewriting by hand the model to consider all these possibilities is a cumbersome and error prone task. Therefore, it would be convenient to have at one's disposal a framework where these tasks can be performed in an automatic manner. The modeling tool ProMoT (Process Modeling Tool)[95] provides a natural environment

for such a modular modeling: ProMoT is based on an object-oriented modeling concept, facilitating the reuse and combination of modules. For this purpose, different libraries of models and modules can be implemented, which can be easily combined via a user interface. Thus, ProMoT allows to intuitively implement models in a modular and hierarchical manner. Therefore, an extension for ProMoT was developed, which allows to import models modularly decomposed according to the procedure described above. Currently, the modularity analysis is performed in Matlab[170] and the results are imported into ProMoT. A full integration of those methods into ProMoT is to be performed in the close future.

In the following, the applicability of the criteria to real cases will be exemplified by several examples of mathematical models of signal transduction pathways, ranging from the paradigmatic MAP kinase cascade to the complicated EGF signaling pathway in humans (introduced in Chapter 2).

3.4 Examples

3.4.1 MAP kinase cascade

Considering its structure, the MAPK Cascade can be decomposed in three submodules corresponding to the 3 kinases, as depicted in Figure 2.2. The connections between the three modules belong to the type discussed above (Equation 3.10). The assumption (i) introduced in page 23 does not hold, because the concentration information about E (e.g. MEK-PP) is needed in order to compute the dephosphorylation steps. However, the assumption (ii) might hold, depending on the values of the kinetic parameters and kinases concentrations. Some mathematical models (e.g.[29;134]) that include the MAPK cascade have been set up assuming (ii), i.e., the quasi-steady-state assumption - which implies the application of Michaelis Menten kinetics,-while others have not (e.g.[234]). As a proof of principle for the approach outlined in Section 3.3, we shall analyze the MAPK Cascade model of Kholodenko[134]. This model is set up simply, with all of the reactions described by Michaelis Menten kinetics (which provides a connection between the modules free of retroactivity), see Section 2.2.1.1[‡]. The model includes a negative feedback from ERK to Raf (see Figure 2.2). As expected, the algorithm finds the 3 modules (Figure 3.5), and characterizes their connections accurately.

[‡]In Section 3.4.4 we shall explore the modularity of a MAPK cascade modeled with mass-action law, embedded in a larger model.

(a) Biochemical representation of the MAPK cascade

(b) Result of the automatic decomposition and subsequent characterization
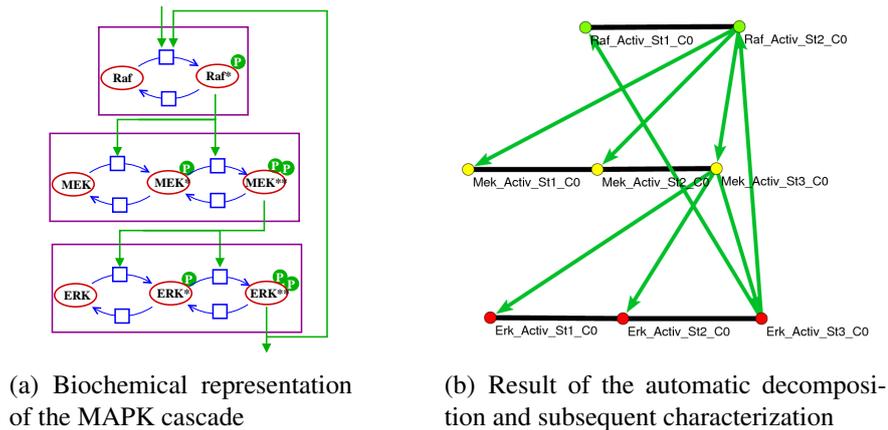
Figure 3.5: Automatic modularization of the MAPK cascade model of Kholodenko[134]. Each color of the edges represents a module: Raf (green), MEK (red) and ERK (yellow). Green lines represent control by a potential, and black bidirectional (potential+current) coupling. Figure Figure 3.5(b) was created with Pajek[15].

3.4.2 Simple receptor system

The modularity of the MAPK cascade can not only be captured applying the concept of retroactivity, but also the elementary flux modes (EFMs) (see Section 3.1.3). However, the latter fails when dealing with subsystems with no mass flow, as for example the binding of a ligand to a receptor. If there is a production and degradation of the receptor, an EFM would be present in that module but if not, there is no EFM and thus this approach can not be used. A similar problem would arise if one considers the binding of two proteins as a necessary step in a signaling cascade. A simple example, describing an extension of the MAPK Cascade of the previous section (Figure 3.6) to include a receptor, illustrates this situation.

3.4.3 EGF signaling network

We shall now consider the comprehensive map of EGF signaling (see Section 2.2.1) of Oda et al.[187], which is probably the largest model of a signaling network within a kinetic (stoichiometric) framework. The enzymatic reactions are described as the conversion of a substrate into a product, being the enzymes considered as modifiers of the reaction. This can be interpreted as an abstraction of the Michaelis-Menten kinetics: the enzyme is not sequestered by the reaction but no specific law is given. This would hinder to make simulations with this model but, fortunately, is enough for our purposes here. In spite of the remarkable size and complexity of the network (comprising over 200 reactions and 300 species), the algorithm is able to decompose it into 42 modules. The results were correct in the sense that coupled elements were assigned to the same module, but imperfect since, in many cases, completely decoupled elements were grouped together (e.g. module 7 lumped together MKK1, MKK4, and CREB, Figure 3.4.3). Since the separation into 2 modules of decoupled elements should

(a) Biochemical representation of a simple receptor system

(b) Elementary flux modes in the model

(c) Automatic decomposition and characterization
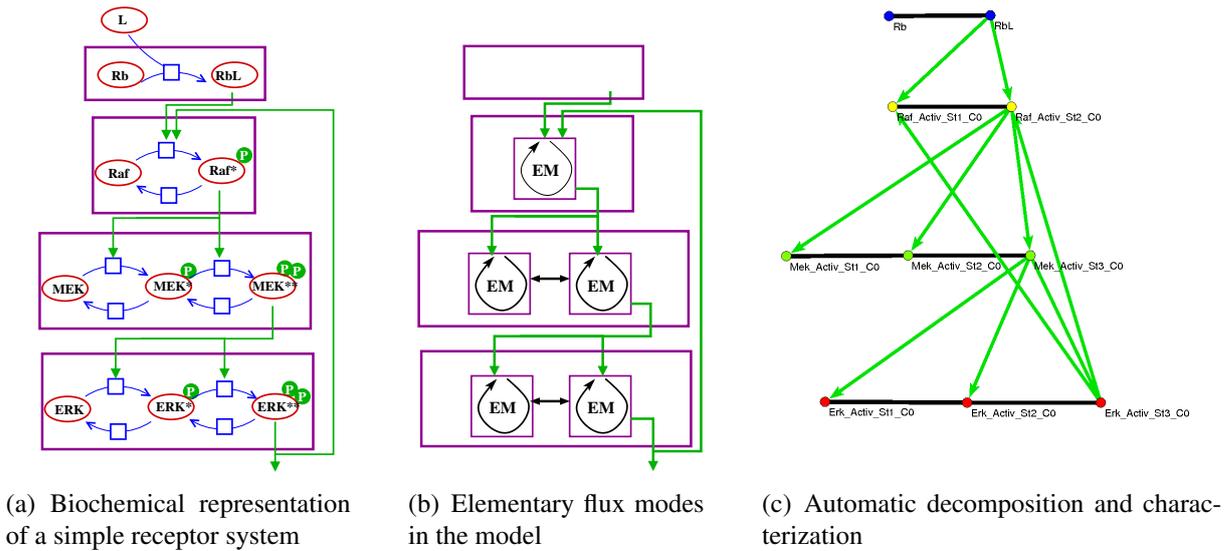
Figure 3.6: Retroactivity vs. Elementary Flux Modes as a criterion to identify modules. While both are able to capture the modules for the cycles of activation/deactivation, the EFMs (Figure 3.6(b)) fail to identify the unit describing the binding of ligand to receptor, since there is no mass flow in steady state.

increase the modularity, this is most probably due to the fact that the optimization algorithms are not able to find the global optimum, and they remain stuck in a local optimum.

3.4.4 EGF-induced MAPK cascade

The EGF induced MAPK Cascade was introduced in Section 2.2.1. Here, the model of the EGF induced MAPK Cascade of Schoeberl et al.[234] will be analyzed, which describes the activation of the ERK MAPK Cascade by EGF. The model includes (a) the reception of EGF, (b) the formation of signaling complexes by interaction of several signaling proteins (namely Sos, Grb2 and Shc), (c) the activation of a signaling intermediate called Ras and (d) the activation of the Raf/MEK/ERK MAPK Cascade (see Figures 3.8 and 3.9). Furthermore, the model also includes the internalization processes: the EGF receptor, alone or bound to other proteins, can be internalized. Once internalized, the EGFR is still active and can bind to the same compounds as the receptor on the surface[234]. Hence, the internalization duplicates all the steps described above and increases the complexity of the system (see Figure 3.9(a)). This duplication, together with the different complexes that the 13 components (from the EGF ligand to ERK) can form, gives rise to 94 states included in the mathematical model.

It constitutes thus a submodel of the network of Oda et al.[187] analyzed in the previous section. However, we consider this case at the end as it poses the hardest problem to the algorithm. The difficulty arises from the fact that all reactions are modeled with mass action law, and from the high grade of entanglement of the model (see Figure 3.8).

Let us try to demarcate the modules first 'by hand'. If we focus on the non-internalized
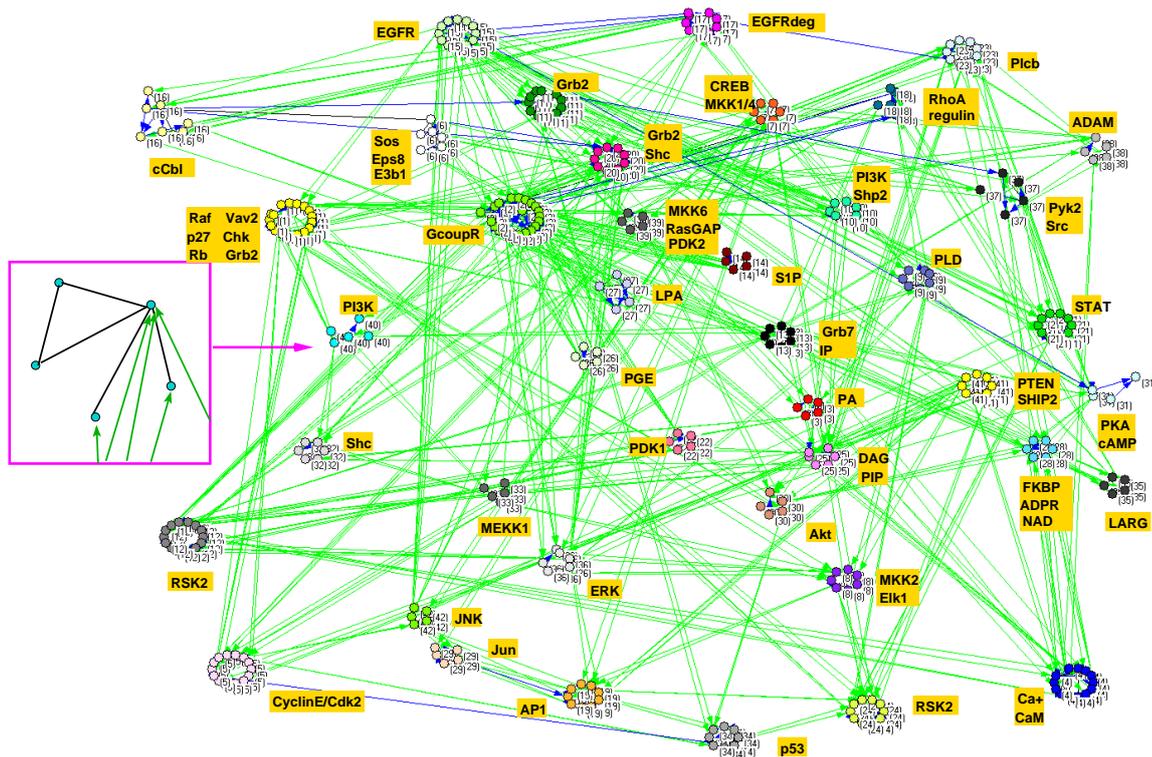
Figure 3.7: Automatic decomposition into modules of the EGF signaling network of Oda et al. [187].
For the sake of simplicity, the names of the compounds have been replace by numbers, and a label in a golden box indicates the key molecules comprised in each of the modules. The color of the edges labels the module they belong to. Green lines represent control by a potential, blue ones control by a current, and black ones bidirectional (potential+current) coupling. The figure was created with Pajek [15].

pathway of the model (see Figure 3.9(a)) , the system can be decomposed as depicted in Figure 3.8. The first module includes the EGF reception process. In principle, it is not clear which particular step should be the border of the module, e.g. the receptor phosphorylation ($v3$ in the model of Schoeberl et al.[234], see Figure 3.9(b)) or the binding of GAP. A careful analysis of simulation results reveals that, although the receptor phosphorylation is *formally* considered a reversible step in the model, it is a suitable point to separate units, since the effect of the backward term is almost negligible, and the connection is hence almost free of retroactivity (data not shown). Note that when we apply the algorithm in this section we will use only structural information. However, in Section 3.4.4.1 we will perform a precise analysis using kinetic parameters and simulation data.

The next unit is the signaling complex formation. Here we can *conceptually* distinguish two submodules, corresponding to the complex formation with and without the adaptor molecule Shc (see Figure 3.9(b)). These two submodules are however strongly coupled since they share several signaling molecules. The output signal of both units (compounds $c35$ and $c25$ in the model) is integrated into an enzymatic signal leading to a connection of the type discussed above (Equation 3.14).
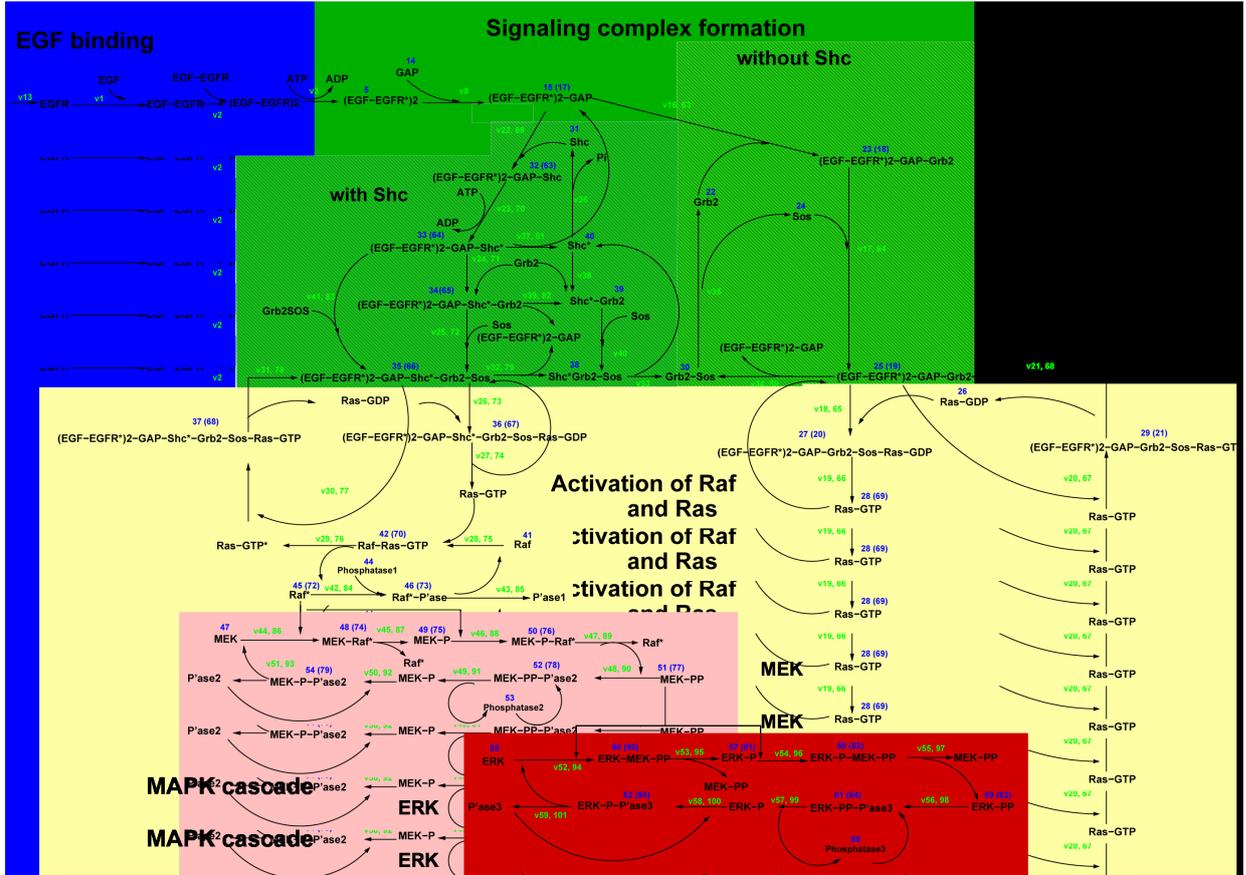
Figure 3.8: Biochemical representation of the EGF-induced MAPK Cascade of Schoeberl et al.[234]. The colored boxes represent the modules as one would define them applying the concept of retroactivity 'by hand', i.e., by inspecting the network visually[222].

The next elements are the activation of Ras, Raf, MEK and ERK. The activations of Ras and Raf are strongly coupled, since Ras activates Raf through a reaction of the type of Equation 3.9. Therefore, it is more reasonable to consider Ras and Raf as a unique module, obtaining three modules (Raf-Ras, MEK and ERK, see Figure 3.9(b)) with enzymatic outputs ($c45$, $c51$ an $c59$, respectively). The connection between Raf-Ras and MEK modules, as well as the connection between MEK and ERK modules, belong to the type defined in Equation 3.10 which, even if the quasi-steady-state assumption does not hold, are reasonable points for defining modules' borders, at least from an 'intuitive' perspective.

To check the ability of the algorithm to cope with this entangled structure (see Figure 3.10(a)), and at the same time examine the accuracy of the 'intuitive' decomposition described above, the whole model as published by Schoeberl et al.[234] is analyzed. Remarkably, the algorithm proposes the same number of modules we have discussed above and, additionally, the distribution of the nodes in the modules is in both cases very similar (see Figure 3.9(b)). By applying a more advanced optimization algorithm, a slightly higher modularity could be achieved based on the decomposition of the Raf-Ras module into the internalized and not internalized part[87].

(a)   (b)
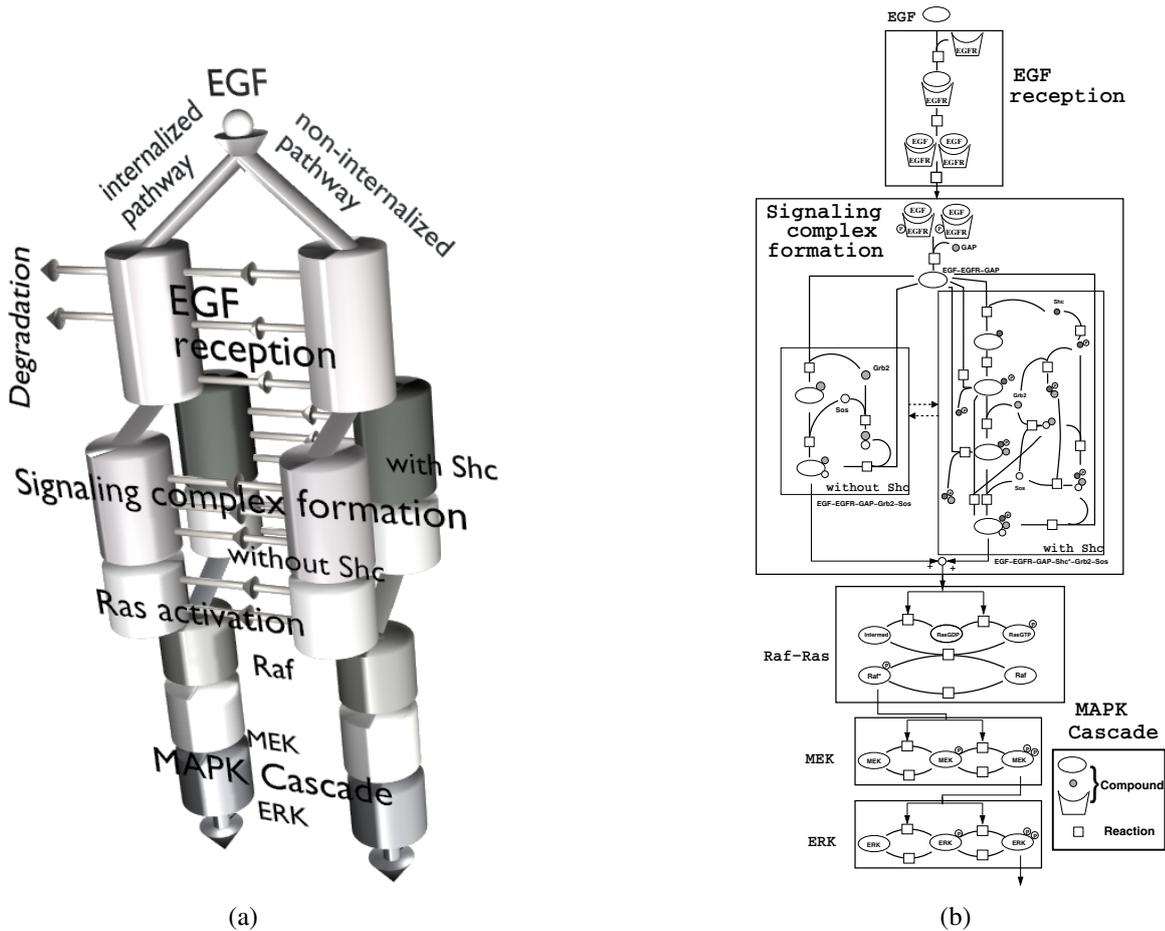
Figure 3.9: Representation of the EGF signaling model of Schoeberl et al. [234]. (a) 3D representation of the whole model; (b) Modular representation of the non-internalized part of the model. The module Raf-Ras corresponds to a modified version of the module MAPKKK of the MAPK Cascade, the module MEK to the module MAPKK and the module ERK to the module MAPK (see Figure 2.2).

3.4.4.1 Quantitative analysis of the EGF-induced MAPK cascade

The examples discussed above rely on the structural absence of retroactivity. This has been the focus of our analysis since we wanted to handle as many networks as possible: for most of them detailed and complete information about kinetic parameters and concentrations is not available, and thus a quantitative analysis of the retroactivity as outlined in Section 3.3.1 is not possible. The model of Schoeberl et al. [234] is a good example to test this approach on a realistic case, as it is a fully characterized model.

Here, we shall use typical simulation conditions of the model to *quantify* the retroactivity matrix. Subsequently, we used a criterion to define which retroactivities are important defined as follows: for each time point, we compare the value of all elements J_{ij} (which represents the influence of c_i on c_j) and $J_{max,i} := \max_j J_{ij}$ the maximum of J_{i1}, J_{i2}, etc., representing the most important influence on c_j. We define a certain J_{ij} to be relevant (and thus $J_{ij}^I = 1$), if, at least for one time point, J_{ij} is larger than a certain fraction ε of the maximum $J_{max,i}$
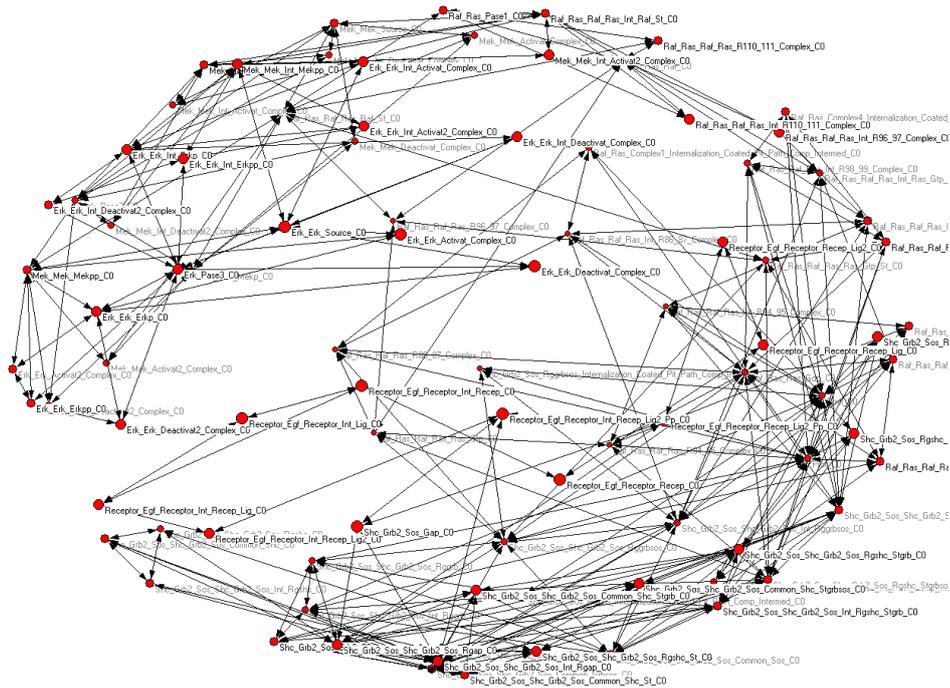
(a) J^I matrix
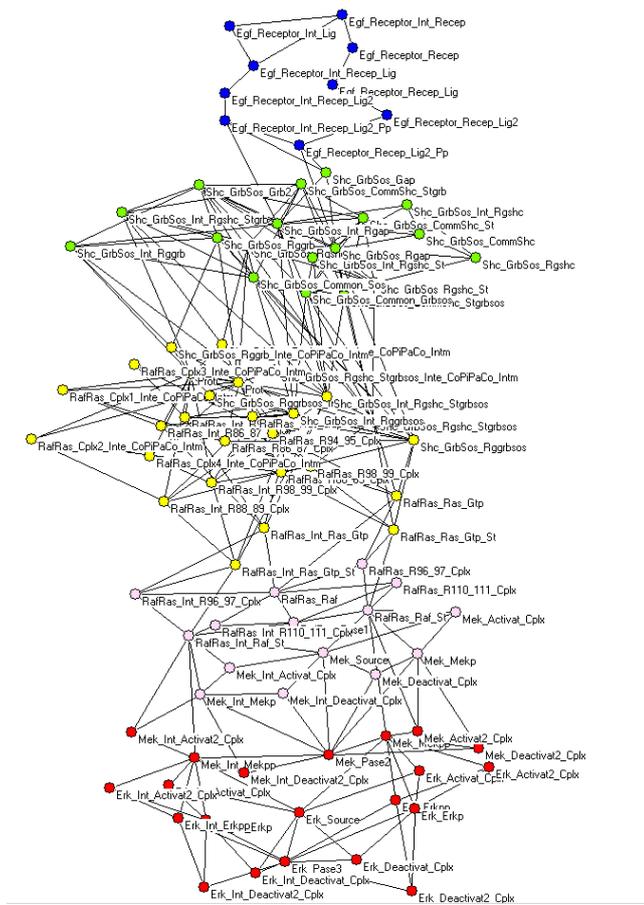
(b) Result of the automatic decomposition

Figure 3.10: Automatic modularization of the model of the EGF induced MAPK cascade model of Schoeberl et al.[234]. In Figure 3.10(b), the color of the edges labels the module they belong to. Blue lines represent control by a current and black bidirectional (potential+current) coupling. Note that the colors of the modules correspond to those of Figure 3.8. Both figures were created with Pajek[15].

(Equation 3.23),

$$J_{ij}^{I} = \begin{cases} 0 & \text{if } \frac{J_{ij}(t)}{J_{max,i}(t)} < \varepsilon \ \forall t \\ 1 & \text{else.} \end{cases} \tag{3.41}$$

For $\varepsilon = 0$ one would get the structural retroactivity, and thus the same results as in the previous section. Increasing the values of ε, elements in J_{ij}^{I} become 0 which were before 1, and thus the coupling of the network decreases. Therefore, the number of modules increases. For instance, for $\varepsilon = 0.001$, one finds 9 modules, and 11 for $\varepsilon = 0.1$ (data not shown). The resulting modules, however, were not intuitively 'good' modules. Therefore, it seems that the application of a 'fine' quantitative analysis, at least in this case, is not an advantage.

3.5 Conclusions

This chapter considers the modularity of signaling networks from a systems-theoretical perspective. A new criterion to demarcate modules, namely the absence of retroactivity, is introduced and, relying upon methods from graph theory, implemented into an algorithm. The method allows to work at two levels of detail: either at the level of the pure stoichiometric structure of the model (with the advantage that much less information is required), or at a more detailed level, where kinetic information can provide subtle additional information. The procedure was tested with a number of case studies, mainly focusing on the structural approach. Particularly important are the results with the EGF signaling map of Oda et al.[187] and the EGF-induced MAPK of Schoeberl et al.[234]. These examples have been carefully chosen as they represent, as far as signaling networks are concerned, the largest model currently available the former, and (probably) the most entangled the latter. The fact that the algorithm successfully decomposes these two models suggests that our approach is able to deal with any state-of-the-art model.

Several points may be subject of future development. For example, we have focused on a 'single-level' modularity: find a number of modules, all of them at the same hierarchical level. It would be interesting to analyze recursively the modularity of them, obtaining thus a multi-hierarchical modularity. Besides, the optimization algorithms proved to be remarkably efficient, but, particularly in the largest case (the EGF signaling map of Oda et al.[187]), not perfect. There may be additional methods which could be tested, as well as combinations of them. Particularly promising in the case of highly decoupled networks (as the EGF signaling map of Oda et al.[187]) seems to be the combination of graph theory methods to identify completely decoupled subgraphs in the graph (which are computationally not expensive and could be thus used as a pre-processing step) with the methods described here.

The integration of the decomposition method presented here into ProMoT provides a unique platform for a modular analysis: one can take a model - e.g. in SBML[119] (Systems Biology Markup Language) from a database such as BioModels[153],- 'blindly' decompose

it into sensible subunits, and easily create models of isolated subunits and combinations of them for analysis. In this context, it would be of interest to have in ProMoT a visualization of the models exploiting this modularity.

Chapter 4

Definition and Setup of a Construction Kit for Signaling Networks

As explained in Section 2.1, a hallmark of the proteins involved in signal transduction processes is their ability to bind different molecules via different domains. Recently, a new approach has been developed tailored to this fact, which thus considers the protein domains as the actual units of signal transduction. Applying this new method, signal transduction network models posses a modular structure where, if the domains do not interact allosterically, each module corresponds to a molecular domain[47;221]. Since the possible events taking place at a single domain are discrete, the number of possible modules is limited. Therefore, one could develop a construction kit of motifs, so that any signal transduction network can be set up as an aggregation of those elements.

Once such a kit is defined, it would be reasonable to analyze its elements systematically; upon thorough analysis of the units, the properties of combinations of them should be easier to understand. Chapter 5 is devoted to this goal. In this chapter, the method leading to the construction kit will be presented in Section 4.1, the elements of this library will be introduced in Section 4.2, and its implementation in ProMoT will be outlined in Section 4.3.

4.1 A Domain Oriented Approach leads to a Natural Decomposition of Signaling Networks into Modules

Consider a signaling network. If one tries to describe all possible combinations of proteins, the number of feasible states explodes to thousands or even millions[27;47;111;221] (see Figure 4.1(a)). This combinatorial complexity has typically been circumvented by assuming that only certain compounds are relevant (see Figure 4.1(b)). However, it is not possible to determine *a priori* which of all the possible molecular combinations of proteins (which we shall call micro-states) are the important ones[68], and therefore, a rigorous description has to include all possible states[23].

(a)   (b)

(c)   (d)
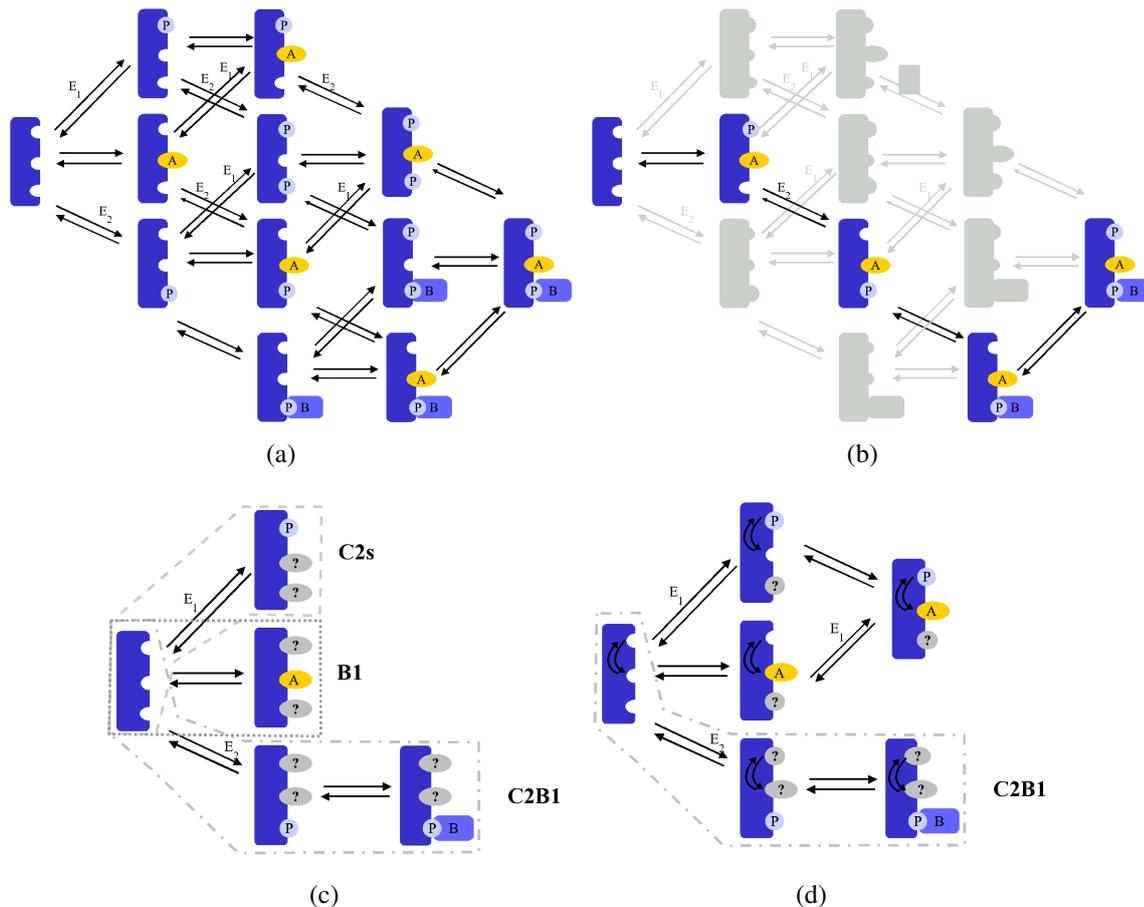
Figure 4.1: Combinatorial complexity in signal transduction networks and their rigorous, modular description using a domain-oriented approach, illustrated by a protein with 3 domains. The first domain can be phosphorylated by an enzyme E_1. At the second domain, another protein A can bind and at the third, upon phosphorylation by an enzyme E_2, a protein B can bind. The possible states (micro-states) are represented in Figure 4.1(a). Traditionally, the modeler assumes an order of reactions, e.g. as in Figure 4.1(b) which, however, is not always justified. Note that in this case, a separation into modules is not possible. If the domains do not influence each other, they can be rigorously considered as independent entities (Figure 4.1(c)), corresponding to the simple motifs B1, C2s and C2B1 depicted in Figures 4.4(a), 4.3(a), and 4.4(b), respectively. If two are somehow interconnected, they must be considered together, while the third can be considered alone (Figure 4.1(d)).

Recently, a new approach based on the macro-states (the states of the different domains considered independently, see Figure 4.1(c)) instead of the micro-states has been proposed[27] which has been extended and formalized by Conzelmann et al.[48]. The latter methodology provides a description exactly equivalent to the microscopic one, but the resulting models are of a lower dimension and -in contrast of the 'intuitive' reduction of e.g. Figure 4.1(b),- are structured in a modular manner, corresponding each module to a protein domain (and thus, to one macrostate). The key idea is that a state space transformation allows a reversible move between the macroscopic and microscopic description. This approach, which has been exposed in a detailed manner elsewhere[48], will be introduced here succinctly for the sake of

completeness. The method operates according to the following steps[48]:

1. Define a complete mechanistic description of the microstates $\vec{c}$ of the system (corresponding to Figure 4.1(a)), creating a set of ODEs in the form of Equation 3.6. The macro-states of interest $\vec{y}$ are a function of the micro-states, $\vec{y} = h(\vec{c})$,

2. Adjust the kinetic parameters according to domain interactions. For example, if a scaffold molecule M can bind to A and B, and both binding sites are independent, the kinetic parameters for the binding of M to A will be the same for all micro-states where M can bind A:

$$M + A \underset{k_{r1}}{\overset{k1}{\rightleftharpoons}} MA \tag{4.1}$$

$$MB + A \underset{k_{r1}}{\overset{k1}{\rightleftharpoons}} MAB. \tag{4.2}$$

Importantly, this simplification is independent of the parameter values; one only needs to know whether the binding sites influence each other.

3. Perform a linear transformation $\vec{z} = T\vec{c}$, where T is a quadratic, non-singular matrix. The resulting states z_i can be classified into levels or tiers, representing each tier a level of detail: the 0th tier corresponds to the total concentration of the protein, the 1st tier to the macro-states (the state of the individual domains), the 2nd tier to the state of all pairs of domains (i.e., the concentration of proteins with concurrently occupied domains 1 and 2, 1 and 3, 2 and 3, etc.), the 3rd tier to the triples of domains, and so on. Importantly, this transformation is general, i.e., independent of domain interactions and kinetic parameters[48]. Furthermore, the new states of tiers 0 and 1 represent the state of the domains (e.g. degree of phosphorylation, of binding, etc.), a quantity biologists are used to work with, and that can be measured more easily than the concentration of particular species.

4. In many relevant cases the transformed model equations for $\vec{z}$ can be decomposed into two sets $\vec{z}_1$ and $\vec{z}_2$ so that

$$\dot{\vec{z}} = \begin{bmatrix} \dot{\vec{z}}_1 \\ \dot{\vec{z}}_2 \end{bmatrix} = \begin{bmatrix} g_1(\vec{z}_1, \vec{u}) \\ g_2(\vec{z}_1, \vec{z}_2, \vec{u}) \end{bmatrix}, \tag{4.3}$$

$$\vec{y} = h(T^{-1}\vec{z}_1). \tag{4.4}$$

In systems theory $\vec{z}_2$ is said to be non-observable[122]* because $\vec{z}_2$ can not be determined from the output function $\vec{y}$. The macro-states (which are a subset of $\vec{z}_1$) are not influ-

---

*Formally, a system is said to be (completely) observable if, starting at any time point, the state vector can be determined from the output function defined over a finite time interval[122].

enced by the state vector $\vec{z}_2$ and, therefore, a reduced model only has to account for the ODEs $\dot{\vec{z}}_1 = g_1(\vec{z}_1, \vec{u})$ [48].

A key advantage is that the domains can be modeled as independent motifs when the domains do not influence each other (Figure 4.1(c)). However, if the state of one domain influences the properties of another, they must be considered together (Figure 4.1(d)). A typical example is when binding of a protein to a certain domain changes the conformation of another domain. Note, however, that if e.g. the phosphorylation of a domain converts it into an active catalytic center which in turn can affect another domain in a trans-molecular manner, but the latter domain does not affect the former, then the domains can be modeled independently (see Figure 4.2).
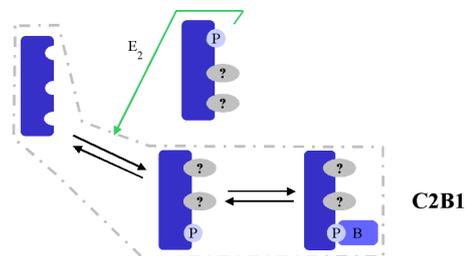
Figure 4.2: Trans-molecular interactions and their description in the domain-oriented approach.
If e.g. in the system defined in Figure 4.1, the phosphorylation of the first domain converts it into an active catalytic center which can catalyze the phosphorylation of another domain in a trans-molecular manner, but the latter domain does not affect the former, then the domains can be modeled independently.

Accordingly, to set up large signaling networks one could wire together a number of motifs belonging to a discrete set wherever domains can be considered independent, and only where an inter-domain influence is present, model them in a combined manner.

In summary, a domain-oriented approach [48] provides a mathematically rigorous rationale for the biologically intuitive fact that protein domains are the units of signaling networks. In the next section such a kit of motifs will be presented.

4.2 Definition of a Construction Kit of Motifs to Set Up Models of Signaling Networks

A kit was defined which comprises common recurring biochemical processes taking places at protein domains (see Figures 4.2, 4.5, and 4.6). A central motif is a cycle of activation/deactivation [137]. It typically involves a phosphorylation, which leads to the activation of the protein [146], and is counteracted by a dephosphorylation (see Figure 4.3(a)). Another common realization of this motif is via a guanine nucleotide exchange factor (GEF) that promotes the exchange of GDP by GTP and is opposed by a GTPase-activating protein (GAP),

see Section 2.1. Sometimes, an activated motif can phosphorylate (activate) the same domain in another molecule, leading to an auto-catalytic effect (Figure 4.3(b)). On the other hand, the activated motif may be able to deactivate the same motif in another molecule, giving rise to a negative feedback (Figure 4.3(c)[†]). Since these are cycles involving 2 states, we shall call C2 the generic case, and add an additional letter (e.g. C2p) to specifically mention a variant of this motif.
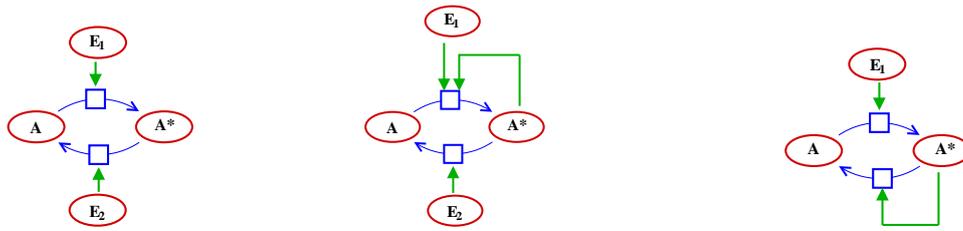
(a) **C2s**: simple cycle (b) **C2p**: cycle with positive feedback (c) **C2n**: cycle with negative feedback

Figure 4.3: Motifs involving simple cycles of activation/deactivation.

In addition to activation/deactivation cycles, another essential motif of signal transduction is the binding of two domains (see Figure 4.4(a)), leading to the formation of a complex, as it is for example the case of the SH3 domains which bind to domains containing sequences of aminoacids rich in proline (see Section 2.1)[‡]. This process can also take place combined with an activation/deactivation cycle (see Figure 4.4(b)), as is for example the case of SH2 or PTB domains, which bind to certain phosphorylated motifs[146] (see Section 2.1). Here, it may happen that two or more molecules compete to bind to the domain (see Figure 4.4(c)).
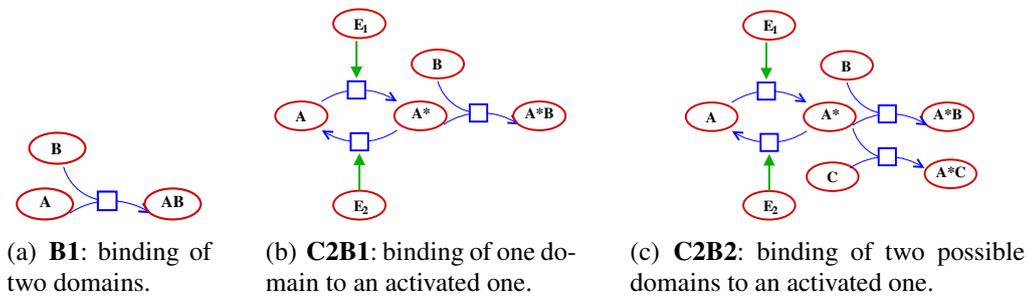
(a) **B1**: binding of two domains. (b) **C2B1**: binding of one domain to an activated one. (c) **C2B2**: binding of two possible domains to an activated one.

Figure 4.4: Motifs involving binding of domains.

It is often the case that the activation of a protein requires the phosphorylation of 2 binding sites. In principle, one would describe such a case as a combination of 2 domains.

However, in many cases it is known that the phosphorylation is not random and a sequential phosphorylation of both domains takes place, as in the case of the Mitogen Protein Activated Kinase (MAPK)[37]. Therefore, this case can be (and is virtually always) described

---

[†]Note that if an additional enzyme can also deactivate the motif, the system is symmetric, and thus regarding the purpose of this work equivalent, to the case C2p depicted in Figure 4.3(b).

[‡]In a sense, this could be interpreted as an activation/deactivation cycle, but it is reasonable to separate both cases since here (i) no chemical change in the domain(s) takes place (the binding is a pure physical process) and (ii) in the cases of the activation/deactivation cycles only one domain is considered, being the other element (e.g. the phosphate group) neglected.

as a lumped motif where a double activation cycle takes place. Since these motifs involve 3 states connected by cycles of activation/deactivation, we shall call these modules C3.

Commonly, both the phosphorylation and dephosphorylation follow a distributive mechanism[45;168], where the enzyme releases the substrate after the first phosphorylation (Figure 4.5(a)). Here, often there is one enzyme responsible for the activation of both domains, and one for the deactivation ((Figure 4.5(a)). However, in some cases, there is an specific enzyme involved in the activation and/or deactivation of each domain (see Figures 4.5(b) and 4.5(c)[§]). Alternatively, one of both enzymes may follow a processive mechanism, where it does not release the substrate between both activation processes (Figure 4.5(e)[¶])[45;168]. Furthermore, autocatalytic or autoinhibitory effects can also appear. We consider here one specific case, where the intermediate and fully activated form are both active and also produce a positive feedback via activation of the second reaction (Figure 4.5(d)). This is the case for example of Src kinases (see Chapter 2.2.2), where the deactivated form (left) is phosphorylated at a negatively regulating site and the fully activated is phosphorylated at a positive regulating site, being the intermediate form not phosphorylated in any of them[193].
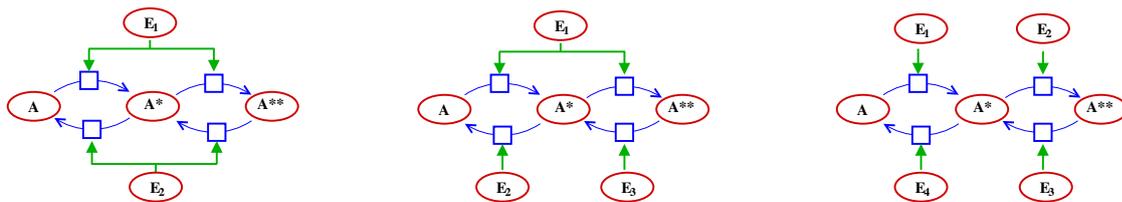
(a) **C3dd**: both reactions following a distributive mechanism.

(b) **C3di**: one reaction follows a distributive mechanism and the other one is catalyzed by 2 independent enzymes.

(c) **C3ii**: all reactions are catalyzed by independent enzymes.
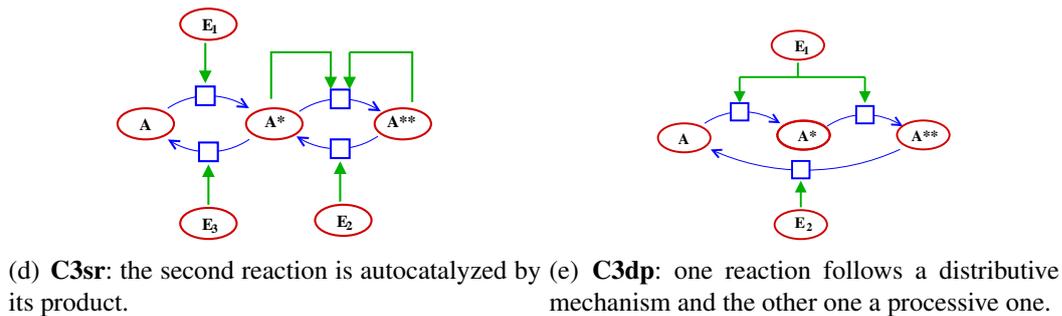
(d) **C3sr**: the second reaction is autocatalyzed by its product.

(e) **C3dp**: one reaction follows a distributive mechanism and the other one a processive one.

Figure 4.5: Motifs involving double cycles of activation/deactivation.

It is also common that cycles of activation/deactivation are coupled. An example widespread in bacteria is the two-component system, where one protein (the sensor) is phosphorylated upon perception of a stimulus and then the phosphate group is transferred to another protein (the receiver), resulting in its activation[252]. We consider here two variants of this motif, one

---

[§]Note that the case of two enzymes involved in the activation and one in the deactivation is symmetric, and thus regarding the analysis of e.g. in Section 5.1 equivalent, to the case C3di depicted in Figure 4.5(b).

[¶]As in the case of C3ii (Figure 4.5(c)), here there are two cases (either activating or inhibiting enzyme are) which are regarding posterior analysis, e.g. in Section 5.1, equivalent.

(a) **C2C2**: two single cycles coupled.

(b) **C3C2**: one single and one double cycle coupled.
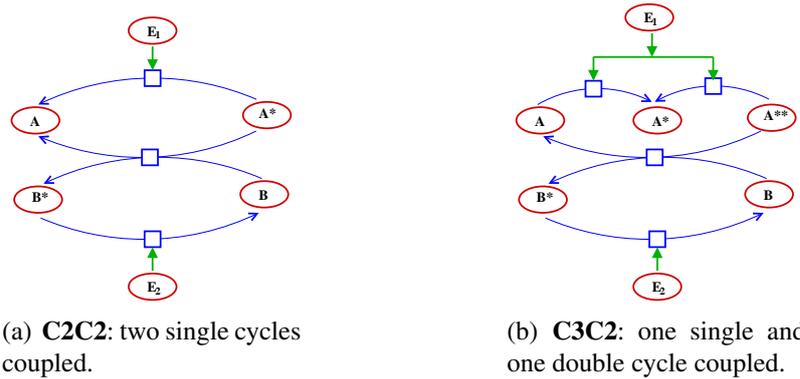
Figure 4.6: Motifs involving coupled cycles of activation/deactivation.

coupling two simple cycles (Figure 4.6(a)), and one consider a double and a simple cycle (Figure 4.6(b)).

These motifs will be thoroughly analyzed in Chapter 5. Previously, an implementation in ProMoT[95] of a construction kit describing these motifs will be presented in the next section.

4.3 Implementation of a Construction Kit of Signal Transduction Motifs in ProMoT

As discussed in Section 4.1, domains are the 'bricks' of the signaling process. However, except a few of them[23;152;163], most modeling tools -including ProMoT - only allow to set up signaling networks following the 'metabolic' paradigm, where one element is assessed to each state rather than to each domain. It would be useful to combine the binding-site-oriented perspective with the modular approach, which is already present in ProMoT (see Section 3.3.2.3). However, this is not a trivial task, since both approaches rely on very different methodologies.

Towards this goal, a library was developed in ProMoT. First, a new family of elements, describing each a domain, were implemented, using a different visual definition (see Figure 4.7). Two key variables were defined, one for the total concentration of the domain (constant if there is no degradation or formation of the molecule where the domain is embedded) and another for the fraction of the domain in the different possible states (see Section A.4 for technical details). Special care was taken to generate modules compatible with the already existing 'metabolic' ones.

Subsequently, the motifs presented in Section 4.2 were set up combining the modules for the domains and the existing reactions for different types of kinetics: mass action law, Michaelis Menten, etc. The modules were also implemented using the classical, 'metabolic' framework, for comparison purposes. Importantly, in many cases both kind of modules are exchangeable (see Figure 4.8). Using these modules, one can model any network as long as the domains can be considered independent (see Figure 4.1). It turned out that, in the

(a) C3dd, compare to Figure 4.5(a)

(b) C3C2, compare to Figure 4.6(b)
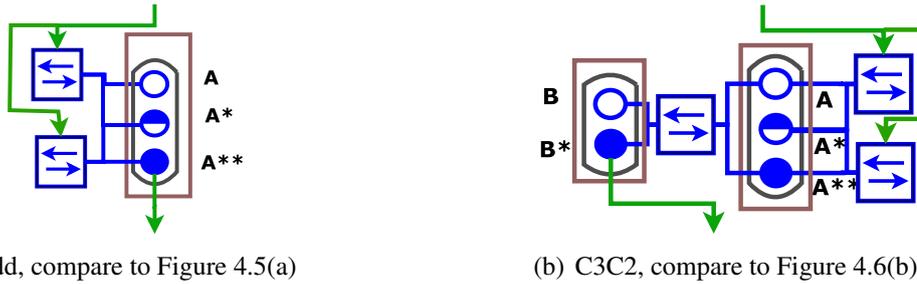
Figure 4.7: Visual definition of motifs using a domain-oriented formalism. This formalism was used in ProMoT to define the modules (see Section 4.3.1 for an example).

cases studied here (e.g. TCR-induced MAPK cascade, see below, and EGF induced MAPK cascade, not shown), there is an interaction of motifs, but always of one nature: a domain A influences a domain B in a unidirectional fashion. In that case, domain B must be modeled in a different manner[48;220], and thus a special module was defined (see Section A.4).

Once the library is implemented, using ProMoT's modular, object-oriented philosophy (see Section 3.3.2.3) and this library of modules, large signaling networks can be modeled by simply drag-and-drop and subsequent wiring of the corresponding modules. Furthermore, ProMoT models can not only be exported using the Systems Biology Modeling Language (SBML)[119], but also directly to Matlab[170] and DIVA, where the analysis described in Chapter 5 will be performed.

To describe a general system in a rigorous manner, one should create all possible species (as done e.g. in BioNetGen[23]) and, subsequently, perform a model reduction as outlined in Section 4.1[48]. The integration of both steps into ProMoT would provide a more powerful environment to set up models considering both a modular and domain-oriented perspective.

4.3.1 Proof of principle: TCR-induced MAPK cascade

To demonstrate its ability to describe large signaling networks, a dynamic model of the TCR-induced MAPK cascade (see Section 2.2.2) was set up (Figure 4.8). Note how this approach allows to combine a modular structure and the domain-oriented approach. Besides, since, as stated above, this new library in compatible with the metabolic one, it is possible to combine both approaches. Here for example the MAPK module (a C3dd motif, see Figure 4.5(a)) was described using the classical metabolic approach, see Figure 4.8, green box.
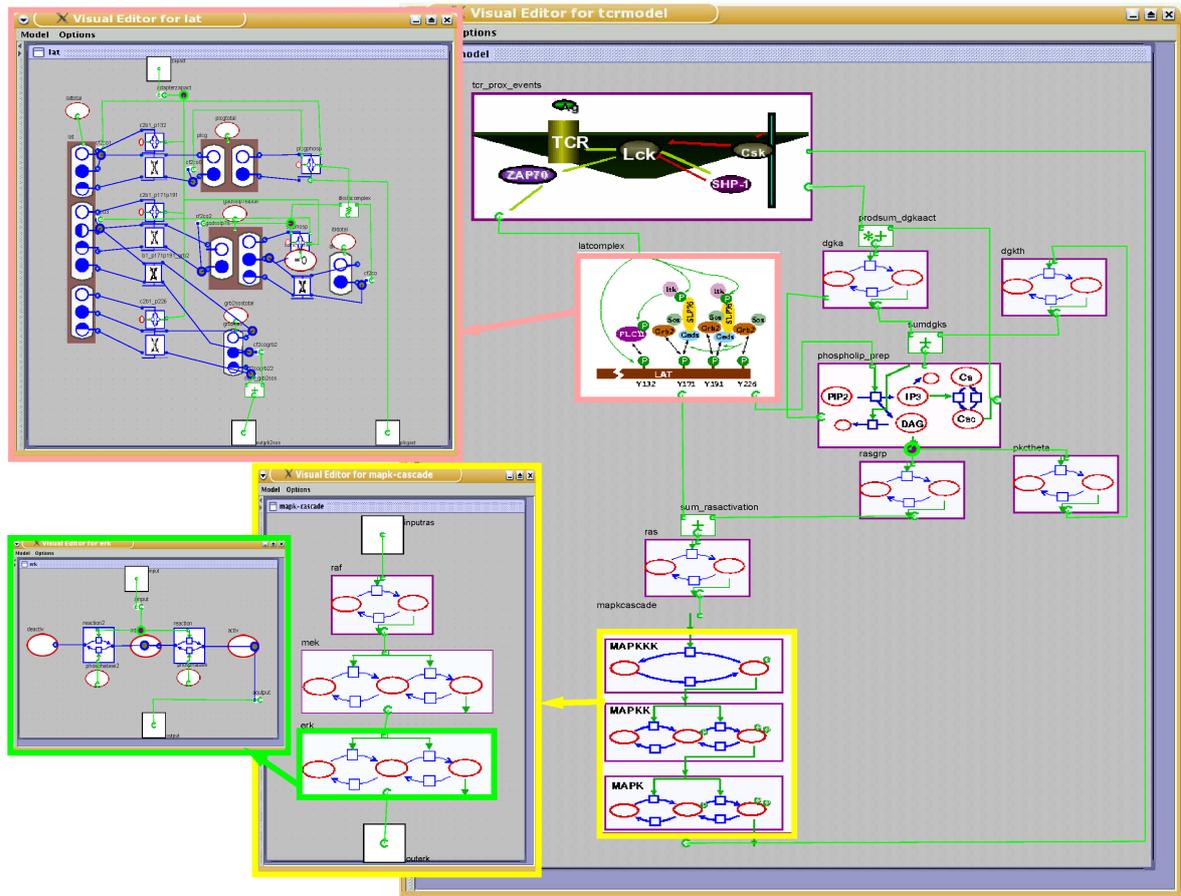
Figure 4.8: Screenshot of the modular, domain-oriented dynamic model of the TCR-induced MAPK cascade implemented in ProMoT. The model describes the main processes involved in the activation of the MAPK cascade via the T-cell receptor (see Section 2.2.2), comprising 92 compounds and 204 parameters. The formalism for the domain-oriented modules is described in Sections 4.3 and A.4.

4.4 Conclusions

In this chapter, the modularity of signaling networks was examined from a molecular perspective. This approach can be interpreted as a molecular 'bottom-up' analysis of modularity, in contrast to the system-theoretical 'top-down' of Chapter 3.

From a molecular point of view, protein domains are the units of signaling networks. A new modeling strategy developed by Conzelmann et al.[48] demonstrates that a rigorous and simplified description of signaling processes is possible, where each module corresponds to one motif, being the later connected in a unidirectional manner.

Since the number of possible processes taking place at a motif is discrete, the number of motifs to be considered is also limited, and these motifs can be combined to set up a variety of signaling networks. In the light of these results, a set of recurring themes was presented in Section 4.2, comprising the typical events occurring in motifs such as binding to another motif and cycles of activation/deactivation.

Furthermore, this construction kit was implemented in ProMoT. The resulting library can be used to set up large signaling networks, and is fully compatible with the basic library. Admittedly, it is not as general as the methodology of Conzelmann et al.[48], and represents only a first attempt to fuse the modular and domain-oriented concepts using ProMoT's current architecture. Therefore, a full integration of the methodology of Conzelmann et al.[48] would be highly desirable.

Once the set of motifs is defined, one would ask himself how they behave, how similar they are, and whether they can show a certain complex non-linear behavior. The next chapter will explore these issues.

Chapter 5

System-theoretical Analysis
of Signal Transduction Motifs

In Chapter 4 a construction kit for signal transduction networks was introduced, and it was shown how this can be used to set up large models of signaling processes. Once such a kit is defined, it is reasonable to analyze its elements systematically; upon thorough analysis of the units, the properties of combinations of them should be easier to understand. Three key properties will be analyzed in this chapter: stability, monotony, and input/output behavior.

Stability would arguably be one of the first properties to examine, not only from a system-theoretical but also from a biological point of view, since it is related to important biological processes such as irreversible decision-making events controlling cell fate and differentiation processes[65;277]. In Section 5.1, the motifs will be thoroughly analysed with respect to their multistability, using Feinberg's Chemical Reaction Network Theory[72], bifurcation methods (combined with a novel method to generate a large number of randomly chosen parameter sets leading to bistability[46]), and an analysis of the corresponding Jacobian matrix.

Monotone systems are well-behaved in a mathematical sense[242] and it is thus very interesting to find out whether our construction kit fulfills this property. This question will be tackled in Section 5.2.

Finally, the input/output (I/O) dynamic behavior will be characterized in Section 5.3. The steady state value, signal amplitude, signaling time and signal duration[106;220] will be used as characteristic parameters. Furthermore, it will explored to what extent these biochemical modules can be substituted by a linear system or a so-called Hammerstein module, composed of a non-linear characteristic curve and a simple linear dynamical system.

For the analysis, we will model the motifs as a set of ordinary differential equations (ODEs) of the form of Equation 3.6 ($\dot{\vec{c}} = N\vec{v}$) . While for the multistability analysis (Section 5.1) we will describe the enzymatic reactions with the more rigorous mass-action-law kinetics, for the analysis of the monotony (Section 5.2) and of the dynamic properties (Section 5.3), these expressions will be simplified into Michaelis-Menten kinetics (see Section 3.2.3, page 22). Admittedly, the assumptions underlying Michaelis-Menten kinetics may

not hold in all the cases studied and may introduce undesired side-effects[25], but, on the other hand, they simplify significantly the system making thus possible the analysis. Therefore, it is arguably a good compromise between detailed description and easiness of analysis*.

5.1 Multistability Analysis of Signaling Motifs

5.1.1 Theoretical principles

A number of methods will be used to tackle the issue of the possible multistability. The methods are ordered so that first those which require less effort are applied, and, to obtain more insight, more complex methods are applied thereafter (see Figure 5.1 for a description of their interconnections).
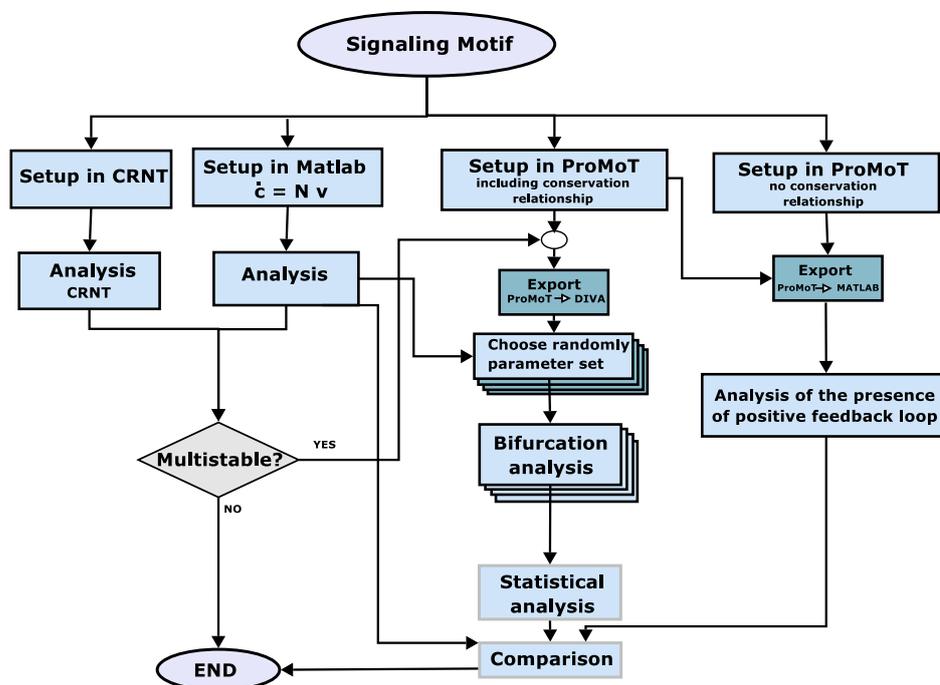
Figure 5.1: Methodology for the analysis of the multistationarity of signaling motifs. Once the modules are defined, the presence of positive feedbacks in the Jacobian of all motifs is determined. Parallelly, they are set up in CRNT Toolbox and ProMoT. Analysis with CRNT Toolbox indicates whether a motif can show multistationarity. In case a motif shows multistationarity, a large number of randomly generated parameters showing multistability is generated, and a detailed bifurcation analysis is performed for all of them in DIVA. The results are subsequently analyzed statistically. The results of the different approaches are finally compared. See the main text and Methods section for details.

5.1.1.1 Identification of feedback loops in the Jacobian matrix

First, the Jacobian matrix (see Section 3.3.2) of the corresponding models will be inspected to detect positive feedbacks. The reason is that it is well-established that there is a connection

---

*Note that one could simplify it further to first order kinetics and, on the other hand, one could argue that mass action law is not realistic enough and thus argue for a stochastic description.

between multistationarity (the ability to display two or more distinct steady states under identical conditions) and positive feedback: the presence of a positive feedback in the incidence graph associated to the Jacobian matrix (Equation 3.32) is a necessary (but not sufficient) condition for multistationarity[143;244]. The algorithm for the detection of feedbacks in the Jacobian matrix is embedded in *CellNetAnalyzer*[144], see Section 6.2. The differential equation models where set up with ProMoT[95] (see Section 3.3.2.3).

5.1.1.2 Chemical Reaction Network Theory

Subsequently, Chemical Reaction Network Theory (CRNT) is applied to determine which motifs can show multistationarity at all. CRNT is an elegant theory developed by Feinberg which allows to make assertions about the ability of a certain (bio)chemical reaction network to show multistationarity, based purely on its structure[45;72;73]. The key concept is the so-called deficiency δ (a non-negative integer), which can be derived from the structure of the network. If $\delta = 0$, the system of ODEs describing the corresponding network with mass-action-law kinetics can not admit multiple steady states for any parameter values. If $\delta = 1$, it can be determined whether a network can admit multiple steady states or not. If $\delta > 1$, however, only under certain conditions can the possibility of existence of multistationarity be gauged. For an introduction to Chemical Reaction Network from the perspective of systems biology the reader is referred to Conradi et al.[45], and for a deeper understanding of the theory itself to the work of Feinberg himself[72;73]. Fortunately, one does not need to understand the details of CRNT, since Feinberg and colleagues have developed the CRNT Toolbox[74], which provides a simple and fast way to apply CRNT to a particular case (of limited size, however). The analyses presented here have been performed with the CRNT Toolbox.

5.1.1.3 Bifurcation analysis

A method recently developed by Conradi et al.[46] allows to characterize analytically the region in the parameter space where a biochemical reaction network described with mass-action-law kinetics allows more than one positive steady state. This powerful method was used here to generate a large number of randomly-chosen parameter sets leading to multistationarity.

The differential equation models where set up with ProMoT[95] (see Section 3.3.2.3). The bifurcation analyses were performed using the continuation methods of DIVA[95].

5.1.2 Results

5.1.2.1 Analysis of the Jacobian Matrix

Surprisingly, there are positive feedbacks in the Jacobian matrix of not only those motifs exhibiting multistationarity, but of all of them. Therefore, we can in principle not exclude any motif from having the ability for showing multistability.

The analysis was performed on models assigning one differential equation to each biochemical entity. However, the modules presented several conservation relationships (since the total concentrations of the protein A and all enzymes E_i are constant), and therefore several differential equations can be replaced by algebraic equations[227].

We also considered this reduced description, since it could happen that these over-dimensioned models introduced 'artificial' positive feedbacks, masking the result of our analysis. However, also under consideration of the conservation relationships, all modules showed positive feedbacks in the incidence graph associated to the Jacobian.

In summary, even though in some cases they may not be obvious from the reaction network, all motifs possess feedbacks: a simple reversible reaction generates a positive feedback between the substrate and the product.

5.1.2.2 CRNT Analysis

The motifs described above (see Figures 4.2, 4.5, and 4.6) were systematically analyzed using CRNT. The results are displayed in Table 5.1. Except for the cases where a distributive mechanism is present (C3dd and C3di), only modules involving an explicit positive feedback are able to show multistationarity: motifs C2p (Figure 4.3(b)) and C3sr (Figure 4.5(d)). By explicit we mean 'visible' in the biochemical network, produced by an autocatalytic reaction. It should be noted that a larger set of variants of the motif C2p has been recently presented by Kholodenko[137].

The results for C3dd and C3dp were firstly presented by Markevich et al.[168] (using bifurcation methods), and confirmed with CRNT[45]. The requirement that a kinase mediates both steps of either the activation or deactivation was recently shown by analytical methods[191]. Here, we extend these results in the sense that we show that the distributive activation is the only mechanism (of a set of alternatives including simple and double cycles, coupled cycles, and binding of proteins) without an explicit autocatalytic feedback that has the ability to produce multistationarity. Furthermore, our analyses rely on a mass-action law expression for the enzymatic reactions, instead of the implicit quasi-steady state assumption of the Michaelis-Menten description used in the works above[168;191]. This is an important generalization as it is known that such assumptions can introduce important differences in the behavior of the systems [25;176].

Furthermore, even though the simple presence of a processive mechanism for either activation or deactivation excludes multistationarity (C3dp, see Figure 4.5(e)), multistationarity

Table 5.1: Summary of the analysis of the multistationarity of the motifs depicted in Figures 4.2, 4.4, and 4.5. First, the existence of a positive feedback in the Jacobian (P.F.) was determined. The deficiency δ and the possibility of multistationarity was computed with the CRNT toolbox (see methods). For the cases where multistationarity is possible (marked with MSS in the third column), a large ($\simeq 1000$) set of parameters leading to multistability was computed, its area of bistability characterized with the parameters Δ_{A_T}, $\Delta_{E_{1T}}$, and $\Delta_{E_{2T}}$ (see Equations 5.1, 5.2 and 5.3, respectively) and the resulting values were statistically analyzed. The statistical distributions are depicted in Figures A.1, A.3, and A.2. Here, the mean values $\overline{\Delta}_{A_T}$, $\overline{\Delta}_{E_{1T}}$, $\overline{\Delta}_{E_{2T}}$ are presented. Additionally, the estimated fraction of cells which would lay out of the bistability region [1-F(-$\overline{\Delta}_{A_T}$,$\overline{\Delta}_{A_T}$)] (see main text and methods) is shown.

Motif	P.F.	δ	CRNT	$\overline{\Delta}_{A_T}$	1-F(-$\overline{\Delta}_{A_T}$,$\overline{\Delta}_{A_T}$)	$\overline{\Delta}_{E_{1T}}$	$\overline{\Delta}_{E_{2T}}$
C2s	$\exists$	1	-	-	-	-	-
C2p	$\exists$	2	MSS	0.2708	0.0367-0.652	0.9037	0.3661
C2n	$\exists$	1	-	-	-	-	-
B1	$\exists$	1	-	-	-	-	-
C2B1	$\exists$	1	-	-	-	-	-
C2B2	$\exists$	1	-	-	-	-	-
C3dd	$\exists$	2	MSS	0.2562	0.3831-0.6694	0.0808	0.0806
C3di	$\exists$	2	MSS	0.0786	0.793-0.8958	0.039	0.0875
C3ii	$\exists$	2	-	-	-	-	-
C3sr	$\exists$	3	MSS	0.2773	0.3553-0.6440	–	0.3760
C3dp	$\exists$	1	-	-	-	-	-
C2C2	$\exists$	1	-	-	-	-	-
C3C2	$\exists$	1	-	-	-	-	-

is possible if both steps are performed by independent enzymes (C3di, Figure 4.5(b)).

A comparison of these results from this section with those of the previous one suggests that, even though it is a requirement for multistability, the presence of positive feedback is not a suitable criterion for identifying multistable systems, since it seems to be ubiquitous in signal transduction motifs.

5.1.2.3 Bifurcation Analysis

In order to go into the nature of the multistationarity of these modules in depth, bifurcation analyses were performed using the continuation methods of DIVA[95] (see Figure 5.1). Note that CRNT can only provide information about the *multistationarity*, i.e. the presence of more than one possible steady state for certain parameter values. If we are interested in the *multistability* (presence of more than one *stable* steady state), it is required to perform bifurcation studies.

To characterize the modules independently of particular parameter values, a large set (around 1000) of randomly chosen parameters leading to multistationarity was generated (see Section 5.1.1.3), and subsequently evaluated statistically. The analysis of the properties
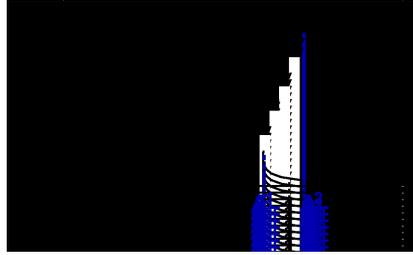
Figure 5.2: Characteristic bifurcation analysis for the module C3di (Figure 4.5(b)) with respect to A_T**.** The values of A_T^1 and A_T^2, which are used to compute Δ_{A_T} (see Equation 5.1) are depicted.

of signaling and gene networks using randomly generated parameters has been successfully applied in different contexts[24;64;139;145]. The bifurcation analyses revealed that every motif showing multistationarity is bistable (i.e. with two stable steady states) for all the parameter sets studied and that the bifurcation is characterized by a limit point. Figure 5.2 shows a characteristic example.

In every case the total concentration of protein A_T was used as a bifurcation parameter. This parameter was chosen for the following biological reason: since concentrations typically fluctuate, we argue that, in order to be a useful bistable switch, these motifs must keep bistability over a wide range of concentrations. Robustness against this sort of noise is postulated to be a hallmark of biological systems[251]. Therefore, we computed a 'robustness' against concentration fluctuations as the relative range of concentration of A_T where bistability is present as

$$\Delta_{A_T} = \frac{|A_T^1 - A_T^2|}{\frac{A_T^1 + A_T^2}{2}}, \tag{5.1}$$

where A_T^1 and A_T^2 are defined as in Figure 5.2. The mean values for Δ_{A_T} are summarized in Table 5.1. Interestingly, all modules show an average value of around 25 % for the range of protein concentration where multistability is possible, except C3di, which shows a much lower value ($\approx$ 8 %). In the case of C3di (Figure 4.5(b)), a distributive double-step activation mechanism is responsible for multistationarity. In the case of C3dd (Figure 4.5(a)), *two* double-step activation mechanisms contributing to multistationarity, while in the other cases (C2d and C3sr, see FigsC2pC3sr), an explicit autocatalytic activation is responsible. Therefore, these results suggest that autocatalytic activation is a stronger precursor of multistationarity than a distributive activation, since one of the earlier leads to similar ranges of Δ_{A_T} as two of the latter.

Recent experimental results in living human cells revealed that the variability in protein concentrations between cells shows a standard deviation of about 15 to 30 %[240]. Similar ranges have been also reported in yeast[182;208], and bacteria[66]. According to this data, we estimated the percentage of cells that would lose their switch-like behavior due to these concentration fluctuations. This is based on the idea that if a certain motif is designed to be
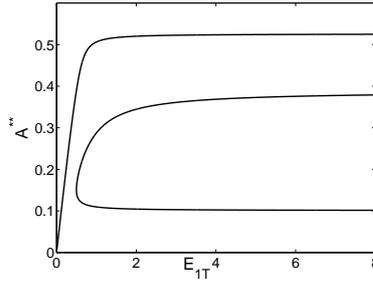
Figure 5.3: Typical continuation curve for the module C3sr (see Figure 4.5(d)) with respect to E_1. After an initial critical amount, any increase in concentration of E_1 has no influence on the steady state. the system is shifted to the right cycle and thus behaves as a single cycle without input and a autocatalytic reaction. Therefore, we do not observe the second bifurcation point.

a switch, and optimized against noise, its average concentration should be approximately in the center of the bistability region, $\mu = \frac{A_T^1 + A_T^2}{2}$. Therefore, for each motif, the proportion of cells with a concentration of A outside the range leading to bistability (assuming a normal distribution, see methods and Figure 5.4) was calculated. The resulting percentages can be thus interpreted as an estimation of the robustness against real fluctuations. Due to the focus here on mammal signaling processes, the ranges 15 to 30 % of Sigal et al.[240] were used, but the results would be similar for yeast or bacteria, since the ranges are similar. The results summarized in Table 5.1 indicate that for an implementation of a switch with an average robustness ($\overline{\Delta}_{A_T}$) a significant fraction of the cells would not be able to operate in the bistable modus, in any of the 4 motifs able to be bistable.
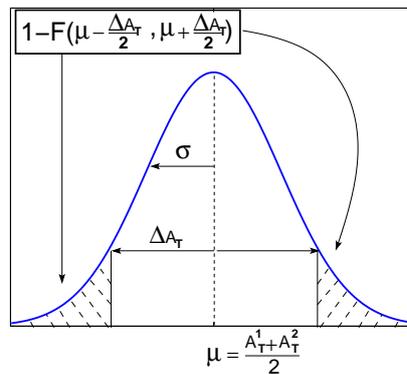
Figure 5.4: Schematic representation of the procedure to estimate the amount of cells out of the bistability regime. Assuming a normal distribution with $\mu = \frac{A_T^1 + A_T^2}{2}$, for a certain σ the amount of cells can be computed as $1 - F(\mu - \frac{\overline{\Delta}_{A_T}}{2}, \mu + \frac{\overline{\Delta}_{A_T}}{2})$.

However, if one considers the most robust (highest Δ_{A_T}) implementations of each of the motifs, (see Figure A.1) one sees that there are cases with a remarkably high Δ_{A_T}, specially for C3dd ($\cong$0.95). Thus, such a hypothetical realization, if feasible from a physical point of view, could be a quite robust one, since only for $\approx 2 - 11\%$ of the cells would the switch lie

out of the bistability region.

In summary, the use of single-level bistable switches does not seem to be a robust design as there would be a significant part of the cells where it would not be operative, even though particular implementations may be.

The same analysis was performed with respect to the enzymes E_1 (activating) and E_2 (deactivating), defining analogously to Equation 5.1

$$\Delta_{E_{1T}} = \frac{|E_{1T}^1 - E_{1T}^2|}{(E_{1T}^1 + E_{1T}^2)/2}, \tag{5.2}$$

and

$$\Delta_{E_{2T}} = \frac{|E_{2T}^1 - E_{2T}^2|}{(E_{2T}^1 + E_{2T}^2)/2}, \tag{5.3}$$

respectively. Here, however, the interpretation is different: rather than robustness against fluctuations in the concentration of the corresponding protein, $\Delta_{E_{1T}}$ and $\Delta_{E_{2T}}$ provide information about the range of operativity of the switch: it indicates for which input values (E_1 and E_2) the module shows bistability. In this context, also the range of multistationarity is much narrower for the modules where the multistationarity relies on the distributive mechanism, see Table 5.1. Note that, for C3sr, the parameter ΔE_1 can not be computed: for an input higher than a certain critical value, the system is 'shifted' to the right cycle and it converts into a single cycle without input and an autocatalytic reaction. The system is thus autonomous and does not depend on the input E_1 (see Figure 5.3).

The results are summarized in Table 5.1. Except for the cases where a distributive mechanism is present (C3dd and C3di), only modules showing an explicit feedback produced by an autocatalytic reaction are able to show multistationarity. The results for C3dd and C3dp were firstly presented by Markevich et al.[168] (using bifurcation methods), and confirmed with CRNT by Conradi et al.[45]. Here, we extend these results in the sense that we show that the distributive activation is the only mechanism without a explicit autocatalytic feedback that 'hides' an ability to produce multistability (C2p, C3sr). Furthermore, even though the presence of a processive mechanism for either activation or deactivation excludes multistationarity (C3dp), multistationarity is possible if both steps are performed by independent enzymes (C3di).

5.2 On the Monotony of Signaling Motifs

In the previous section the multistability of the signaling motifs was investigated. This results characterize one (yet essential) property, but certainly other properties should be analyzed. Since signaling is a highly dynamic process, the dynamical properties are of large interest. We shall start exploring the monotony since, if a system is monotone, many counterintuitive aspects of the dynamics can be excluded.

5.2.1 Theoretical foundations

5.2.1.1 A brief introduction to monotone systems and their relevance in biology

A system that is monotone is well-behaved in a mathematical sense[242]. Consider the response of a system to a certain stimulus. If one modifies the stimulus (or other conditions), and the system starts to response, say, with a higher value than before, one would expect that, for a 'simple' system, the response would remain higher for the whole trajectory. Also, one would expect that, if an even higher value for the input is applied, the response would be even higher. Similar behavior would be expected for the initial values, as well. This intuitive property is what characterizes a monotone system. In mathematical terms, this intuitive idea is formalized as follows: a system of the form of Equation 3.2 is said to be input/output (I/O) monotone if

$$\xi_1 \leq \xi_2 \text{ and } u_1 \leq u_2 \Rightarrow x(t, \xi_1, u_1) \leq x(t, \xi_2, u_2) \ \forall t > 0, \tag{5.4}$$

where ξ represents the initial value. A more strict and even better-behaved system is a strongly I/O monotone system, which fulfills

$$\xi_1 < \xi_2 \text{ and } u_1 < u_2 \Rightarrow x(t, \xi_1, u_1) < x(t, \xi_2, u_2) \ \forall t > 0. \tag{5.5}$$

Monotone systems have a number of interesting properties. For example, under weak additional assumptions, there is at least one steady state x_u for every stepwise input u. Furthermore, if there is only one steady state for each input, this is a global attractor, i.e., all solutions converge to x_u for $t \to \infty$[242]. Importantly, under this condition, the steady-state characteristic curve $x_{ss}(u)$ is well defined. Additionally, considering Equation 5.4 for $t \to \infty$, it results

$$(\xi_1 \leq \xi_2 \text{ and } u_1 \leq u_2 \Rightarrow \lim_{t \to \infty} x(t, \xi_1, u_1) \leq \lim_{t \to \infty} x(t, \xi_2, u_2)) \Leftrightarrow x_{ss}(\xi_1, u_1) \leq x_{ss}(\xi_2, u_2), \tag{5.6}$$

that is, the characteristic curve of a monotone system is monotone.

It would be therefore very useful to know to what extent biochemical systems, and in particular the elementary units defined in Chapter 4, are monotone. If that were the case, one could rigorously exclude 'surprises' at the single level motif and thus comfortably analyze the properties of aggregations of monotone systems. This is particularly convenient since certain combinations of monotone systems, such as chains, are also monotone[7].

A simple condition which guarantees strong monotony can be obtained by analyzing the incidence graph of the system, as outlined by Angeli et al.[7]. For a system with n states, one input u and one output y in the classical form of Equation 3.6 ($\dot{\vec{c}} = f(\vec{x}, u, \vec{p})$, with $y = f(\vec{x})$), the incidence graph has $n + 2$ nodes (one for each state plus one for the input u and one for the output y). A labeled edge is drawn from c_i to c_j if c_i influences the balance of c_j, that is to say, if the element J_{ij} of the Jacobian (see Equation 3.32) is not zero. Additionally, if the input u influences a state x_i, a labeled edge is drawn from u to c_i, and if a state c_k influences the output y, an edge is drawn from c_k to y.

If the signs of all edges are definite, and there is no negative feedback in the incidence graph, the system is said to be strongly input/output monotone[7]. This method will be applied in the following to the motifs described in Chapter 4.

5.2.2 Analysis of the monotony of signaling motifs

We shall illustrate by an example, namely C3sr (see Figure 4.5(d)), the application of the method described in the previous section. As discussed above, the enzymatic reactions will be described using Michaelis-Menten kinetics. Considering the conservation relationship

$$A + A^* + A^{**} = A_0, \tag{5.7}$$

the equations describing the motif C3sr read

$$\dot{A} = \frac{dA}{dt} = -\frac{E_1 \cdot k_1 \cdot A}{K_{m1} + A} + \frac{E_2 \cdot k_{-1} \cdot (A_0 - A - A^{**})}{K_{m-1} + (A_0 - A - A^{**})} \tag{5.8}$$

$$\dot{A}^{**} = \frac{dA^{**}}{dt} = -\frac{E_3 \cdot k_{-2} \cdot A^{**}}{K_{m-2} + A^{**}} + \frac{(A^{**} + (A_0 - A - A^{**})) \cdot k_2 \cdot (A_0 - A - A^{**})}{K_{m2} + (A_0 - A - A^{**})}. \tag{5.9}$$

Furthermore, the input is $u = E_1$ and the output $y = A^{**}$. One can determine the sign of the Jacobian (see Equation 3.32)

$$J = \frac{\partial f}{\partial \vec{c}} = \begin{bmatrix} - & - \\ - & - \end{bmatrix}, \tag{5.10}$$

together with

$$J_{input} = \frac{\partial f}{\partial \vec{u}} = \begin{bmatrix} - \\ + \end{bmatrix} \tag{5.11}$$

and

$$\frac{\partial y}{\partial \vec{c}} = [0 \ +]. \tag{5.12}$$

With this information the incidence graph can be easily set up (see Figure 5.5). As can be seen in both the structure of the Jacobian and in the graph, there is no negative feedback and the path from the input to the output is sign-constant. The system is thus strongly I/O monotone[7].
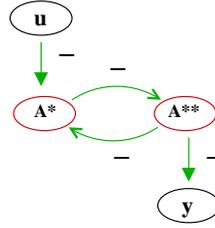
Figure 5.5: Incidence graph of the motif C3sr (see Figure 4.5(d)). Since there is no negative cycle and the path from the input *u* to the output *y* is sign-constant, the system is strongly I/O monotone[7].

One can proceed in a similar manner with the rest of the motifs defined in Chapter 4 (see Figures 4.2, 4.4, 4.5, and 4.6). The results are summarized in Table 5.2. Remarkably, with the exception of the motifs C3dp, C2C2 and C3C2 all motifs are monotone. Furthermore, for the latter examples, at least the characteristic curve is monotone (see Section A.3 for a proof). The fact that the characteristic curves are monotone will be exploited in the next section when these motifs are compared with Hammerstein modules.

5.3 Analysis of the Dynamics of Signaling Motifs

It is tempting to try to describe biological systems in technical terms. Motivated by the fact that the signaling motifs are monotone, and thus 'simple', one could start comparing the motifs to simple technical systems. In this chapter, we shall explore to what extent this is possible. First, the technical elements to be used will be presented, together with some useful definitions, in Section 5.3.1. Later, these technical systems will be compared with the biochemical motifs introduced in Chapter 4.

To keep the models as simple as possible, a Michaelis-Menten description will be used for the kinetics, as in Section 5.2. Furthermore, to minimize the number of parameters to consider, the equations are normalized with respect to the total concentration of protein and with respect of the kinetic constant of the first reaction. For example, for the motif C3sr, dividing Equations 5.8 and 5.9 by A_0 and k_1, and redefining $k_{-1} := \frac{k_{-1}}{k_1}$, $k_2 := \frac{k_2}{k_1}$, $k_{-2} := \frac{k_{-2}}{k_1}$, $t := \frac{t}{k_1}$, as well as $K_{m1} := \frac{K_{m1}}{A_0}$, $K_{m-1} := \frac{K_{m-1}}{A_0}$, $K_{m2} := \frac{K_{m2}}{A_0}$, and $K_{m-2} := \frac{K_{m-2}}{A_0}$, one obtains

$$\dot{A} = \frac{dA}{dt} = -\frac{E_1 \cdot A}{K_{m1} + A} + \frac{E_3 \cdot k_{-1} \cdot (1 - A - A^{**})}{K_{m-1} + (1 - A - A^{**})} \tag{5.13}$$

$$\dot{A}^{**} = \frac{dA^{**}}{dt} = -\frac{E_2 \cdot k_{-2} \cdot A^{**}}{K_{m-2} + A^{**}} + \frac{(1 - A) \cdot k_2 \cdot (1 - A - A^{**})}{K_{m2} + (1 - A - A^{**})}. \tag{5.14}$$

The equations for the rest of the motifs can be found in Section A.1.2. A number of methods, mainly arising from classical ideas from systems theory, will be applied to these simplified systems. In the following, they are succinctly presented.

5.3.1 Methods

5.3.1.1 Linear systems and the Hammerstein module

The standard linear transfer elements used in systems theory are temporally lagged proportional elements, derivative and integral elements, and combinations of these elements[82;94]. We shall start with a very brief presentation of such systems. A detailed study is out of the scope of this work, and the reader is referred to standard system-theory textbooks[82;94].

Simple linear systems. The simplest linear system is a proportional element (so-called P element), defined by the equation

$$y = Ku, \tag{5.15}$$

or, in the frequency domain[82;94] by its so-called transfer function

$$G(s) = \frac{y(s)}{u(s)} = K. \tag{5.16}$$

An *ideal* proportional system responds instantaneously to the input (see Figure 5.6(a)). Addition of time-lags (so-called T elements) delays the response of the system (and represents thus more realistic systems). Additionally, the curve of the response gets more sigmoidal the higher the order of the system is (see Figure 5.6(a)). For example, a temporally lagged element of first order (PT_1) is defined by

$$T\dot{y} + y = Ku \tag{5.17}$$

or its transfer function

$$G(s) = \frac{y(s)}{u(s)} = \frac{K}{Ts+1}. \tag{5.18}$$

The system's output response y to an input signal u can be derived either by solving Equation 5.17 or by evaluating $y = G(s)u$. The parameter K (amplification magnitude) determines whether the input signal is amplified ($K > 1$) or attenuated ($K < 1$). T (time constant) defines how fast the system responses to an input u (corresponds to the time required to reach 63.2% of $K^\dagger$, see also Section 5.3.1.2). Similar lagged systems with second (PT_2), third (PT_3), etc. order are also often used. The transfer function for a system of order n (PT_n), with all time constants equal, would then read

$$G(s) = \frac{y(s)}{u(s)} = \frac{K}{(Ts+1)^n}. \tag{5.19}$$

Also of interest in our context are systems containing a derivative element. For example,

---

†this value derives from the solution of Equation 5.17 for a unitary input 1 (or from computing the inverse Laplace function of the transfer function, $L^{-1}(G(s))$: $y(t) = 1 - e^{t/T} \rightarrow y(T) = 1 - e^{-1} \approx 0.632$.

(a) PT_n (b) DT_n (c) IT_n (d) PDT_n
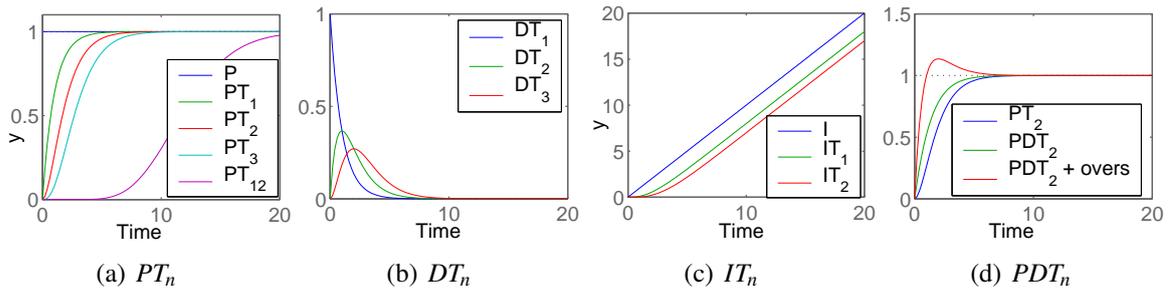
Figure 5.6: Typical step responses of simple linear systems. The simulations were performed with Matlab mapping all constants to 1.

the equation of a DT_2 system reads

$$T_1 T_2 \ddot{y} + (T_1 + T_2)\dot{y} + y = K\dot{u}, \tag{5.20}$$

and its transfer function

$$G(s) = \frac{Ks}{(T_1 s + 1)(T_2 s + 1)} \tag{5.21}$$

While a proportional element reacts according to the magnitude of input, a derivative element responds to *changes* in the input value. Therefore, the response to a step input has a transient form (see Figure 5.6(b)).

The third main type of elements are integrators. An integrator, as the name says, integrates the signal. Therefore, the response of an *ideal* integrator is a steadily growing signal (see Figure 5.6(c)), which is delayed by the addition of lag-elements. The equation of an IT_1 element, for example, reads

$$T\dot{y} + y = K \int_0^t u(\tau)d\tau, \tag{5.22}$$

and its transfer function

$$G(s) = \frac{K}{s(Ts + 1)}. \tag{5.23}$$

Additionally, we shall also consider combinations of these standard elements. A particularly important one will be a PDT_2 system, described by

$$T_1 T_2 \ddot{y} + (T_1 + T_2)\dot{y} + y = K(\dot{u} + u), \tag{5.24}$$

and with transfer function

$$G(s) = \frac{K(s + 1)}{(T_1 s + 1)(T_2 s + 1)}. \tag{5.25}$$

The response of this system is a certain combination of a proportional and derivative system. If the derivative element is important enough, a certain 'peak' (so-called overshoot) can be

seen (see Figure 5.6(d)).

The Hammerstein module. While linearized systems can provide useful insights into biochemical systems under certain conditions, the high non-linearity of the latter limits the applicability of the former. Therefore, it may be more convenient to extend the linearized systems with a non-linear static characteristic curve, resulting in the so-called Hammerstein module[38] (Figure 5.7). The characteristic curve describes the value of the output at steady
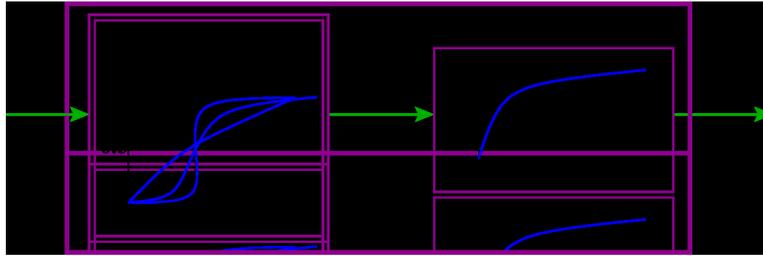
Figure 5.7: Hammerstein module with a Hill function describing its characteristic curve. The higher h, the more sigmoidal gets the characteristic curve. The dynamic part is described by a simple linear system such as a PT_1.

state as a function of a stationary input u, $y^{ss} = f(u)$. The modules analyzed here are monotone except three, and the static characteristic curve is monotone for all of them (see Section 5.2). A reasonable candidate for the function $y^{ss} = f(u)$ would be a normalized (since the amplification is already defined by the amplification magnitude K in the linear model, see Equation 5.19) Hill function

$$\widehat{u}(u) \;=\; \frac{u^h}{K_{0.5}^h + u^h},\tag{5.26}$$

where $K_{0.5}$ is the value of u for which $\widehat{u} = 1/2$ and can be considered as a threshold value (see Figure 5.7), and h is the Hill coefficient. If $h = 1$, then the curve is hyperbolic and is known as the Michaelis-Menten equation. If $h > 1$, then the curve shows a sigmoidal (also called ultrasensitive[96]) form. The higher the Hill coefficient, the more the curve tends to a step-form response. The Hill function is thus a suitable expression because it is monotone and has relatively few parameters which are easy to interpret. In addition, it is familiar to biologists, and was actually first introduced in the field of enzyme kinetics[235].

5.3.1.2 Characteristic parameters

It is convenient to use parameters that capture key features of a system to characterize them. In the analysis of signaling networks, special attention should be paid to their dynamic behavior, since the biological response is often determined by the transient characteristics of the output signal, such as how fast the system reacts (signaling time) and for how long (signal duration), rather than steady state properties[10]. Therefore, the classical steady state analysis

of such systems, though useful, might not provide sufficient insight into the properties of signaling pathways. Heinrich et al.[106] have introduced the parameters signal amplitude S, signaling time τ, and signal duration θ. For the output $y(t)$ of a module, these parameters can be calculated as

$$\tau = \frac{\int_0^\infty t\, y(t)\, dt}{\int_0^\infty y(t)\, dt}, \qquad \theta = \sqrt{\frac{\int_0^\infty t^2\, y(t)\, dt}{\int_0^\infty y(t)\, dt} - \tau^2}, \qquad S = \frac{\int_0^\infty y(t)\, dt}{2\,\theta}.$$

τ and θ (analogous to the mean value and standard deviation of a statistical distribution, respectively) represent the average time to activate the output element and the average time during which this output component is activated, respectively (see Figure 5.8). S gives the relationship between the total amount of output signal (the area under the curve), and the duration θ of the signal, hence providing a measurement of the average concentration of the output element[106]. These parameters provide meaningful information but, if the output signal
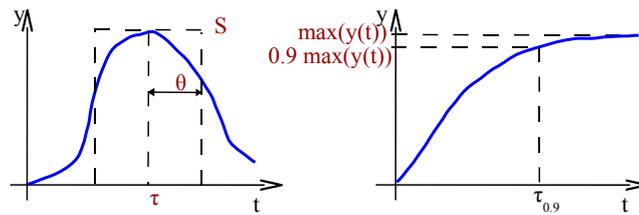
Figure 5.8: **Parameters to quantify the dynamic of a module.** The left plot shows the parameters S, τ, and σ as defined by Heinrich et al.[106] (see Equation 5.27), and the right one τ as defined by Saez-Rodriguez et al.[220] (see Equation 5.27).

does not return to zero after a certain time, τ and θ tend to infinity. A reasonable alternative to τ might be a parameter describing the time the system needs to reach a percentage of its maximal signal output. For example, $\tau_{0.9}$, defined as the time to reach 90% of the maximum (see Figure 5.8),

$$y(\tau_{0.9}) = 0.9\, max(y(t)), \tag{5.27}$$

which can be more generally applied[‡]. The value of the percentage adopted is somewhat arbitrary, and we shall use here the value $\tau_{0.9}$.

Additionally, if the signal does not return to the basal level, the signal amplitude S depends on the simulation time. Since S tends to the steady-state value multiplied by $\sqrt{3}$ as the time tends to infinity[220], in such cases S at $t = \infty$ should be used to have a unique S.

[‡]Note that $\tau_{0.632}$ (63.2%) would correspond to the time constant T of a first order system, see Section 5.3.1.1

5.3.2 Results

We shall now investigate to what extent the motifs introduced in Section 4.2 can be approximated by linear systems and, if not, by a Hammerstein module (see Section 5.3.1.1).

5.3.2.1 Approximation by linear systems

For extreme conditions, the simple motifs can be reduced to linear systems. Consider for example the motif C2s (Figure 4.3(a)). Its output A^* can be computed as

$$\frac{dA^*}{dt} = \frac{k_1[A]}{K_{m1}+[A]}u - \frac{k_2 A^*}{K_{m2}+A^*} = F_1([A])u - F_2(A^*). \tag{5.28}$$

(a)   (b)   (c)   (d)
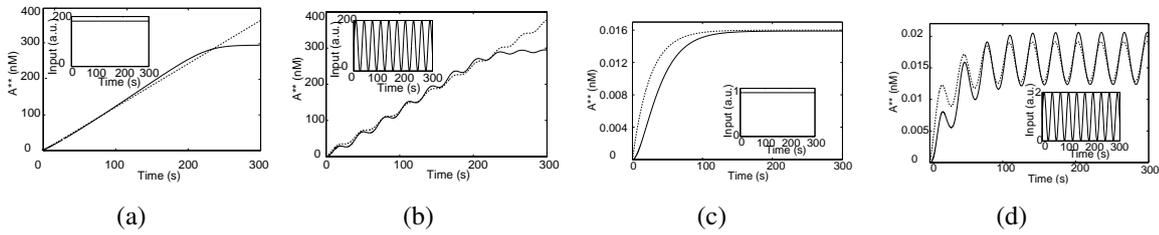
Figure 5.9: Comparison of the MAPK module to simple linear systems. The response of the MAPK module (solid line) is compared to a simple linear system (dashed line). An ideal integrator is able to reproduce the behavior of the MAPK module for high stimuli (Figures 5.9(a) and 5.9(b)) and a system with first order lag for low stimuli (Figures 5.9(c) and 5.9(d)). The input is shown in the inset figure.

It should be noted that A and A^* are coupled, since $A + A^* = A_0$. If the input is low, the conversion of A into A^* will be low. Considering the parameters of the model of the MAPK cascade of Kholodenko[134] (see Section 3.5(b)), the initial concentration of A is $A_0 = 100nM$ and $K_{m1} = 10nM$, then $A \gg K_{m1}$ holds approximately, and hence F_1 can be roughly estimated with k_1. Additionally, since the value of A^* is low (specifically, $A^* \ll K_{m2}$), $F_2 \approx (k_2/K_{m2})A^*$. Therefore, Equation 5.28 results in

$$\frac{dA^*}{dt} \approx k_1 u - (k_2/K_{m2})A^*, \tag{5.29}$$

which is the equation of a system with a first order lag.

On the other hand, if the input value u is high, $F_1 u \gg F_2$. The condition $A \gg K_{m1}$ is fulfilled for a certain period of time and hence $F_1 \approx k_1$, leading to the equation

$$\frac{dA^*}{dt} \approx k_1 u, \tag{5.30}$$

which is the equation of an integrator. If the assumption $A \gg K_{m1}$ is not fulfilled, the system behaves as an integrator, but with variable gain F_1. Since the output is limited by the amount of A (A_0), the system saturates at a certain time.

Analogous considerations can be applied to other modules such as C3dd (Figure 4.5(a)), leading to the same conclusion: behavior as an integrator for high inputs and as a proportional system with first order lag for low inputs, see Figure 5.9. However, for the intermediate range of inputs, none of these extreme cases hold, and thus the modules can not be simplified by linear systems. Thus, in the next section, we shall try to describe these modules with a Hammerstein module where we add a nonlinear characteristic curve to a linear system (see Section 5.3.1.1).

5.3.2.2 Approximation by a Hammerstein module

In this section we shall explore to what extent the motifs introduced in Chapter 4 can be replaced by a Hammerstein module. The goal here is twofold: (i) to gain insights into the behavior of the motifs (whether e.g. the dynamic is as simple as that of a PT_1, or on which kinetic parameters and how the characteristic curve depends), and (ii) to gauge the ability to replace biological by Hammerstein modules.

To characterize the motifs independently of a particular realization (parameterization), for each motif a large number (≈ 1000) of randomly chosen parameter sets was generated, the corresponding implementation of the motif analyzed, and the results evaluated statistically. The parameters were chosen from a log-normal distribution, which is a reasonable assumption for biochemical parameters[145]. A similar rationale was applied in Section 5.1 for the analysis of multistationarity, and in other works in different contexts[24;64;139;145].

The *modus operandi* reads as follows: for each parameter set, the steady state was computed and the parameters of the Hill function (see Equation 5.26) which fits the characteristic curve best were determined using the optimization methods of DIVA. Subsequently, a number (≈ 100) of different input values were chosen, and for each of them, the linear PDT_n system which fits the step-response of the biological system best was identified, using a method developed by Gayer[88]. The results were subsequently studied using different correlation and statistical tools. Here, we shall use the parameters introduced in Section 5.3.1.2 to characterize the dynamics of the response. Particularly revealing in this context is the signaling time τ.

To characterize the quality of the approximation, two errors, Δy_s and $\Delta \tau_{0.9}$, were defined. Δy_s is the relative difference between the steady-state characteristic curve of the biological motif $y^b_s s(u)$ and the Hammerstein module $y^H_s s(u)$ ($= K \widehat{u}(u)$), see Equation 5.26,

$$\Delta y_s = \frac{\int_0^\infty |y^b_s s(u) - y^H_s s(u)|}{\int_0^\infty y^b_s s(u)}, \tag{5.31}$$

and gauges thus the error of the steady-state error. Analogously, $\Delta \tau_{0.9}$ is defined as the rela-

tive difference of the signaling time $\tau_{0.9}$

$$\Delta\tau_{0.9} = \frac{|\tau_{0.9}^b - \tau_{0.9}^H|}{\tau_{0.9}^b}, \quad (5.32)$$

and gives a measure of the error of the dynamic response. With the identification method of Gayer[88], the linear system is identified whose response has the same point of inflexion and slope at this points as the biochemical motif. Therefore, $\tau_{0.632}$ will be useful to characterize the behavior of the motifs, but the error in the dynamics will be estimated via the parameter $\tau_{0.9}$, since $\tau_{0.632}$ is typically very close to the point of inflexion, where the biochemical and Hammerstein modules behave very similarly.

Since $\Delta\tau_{0.9}$ is an absolute value, it is convenient to assess the error but it does not capture the information about which of the systems compared is faster. As we are interested in this fact, a similar parameter will be computed, without the absolute value,

$$\widehat{\Delta\tau_{0.9}} = \frac{\tau_{0.9}^b - \tau_{0.9}^H}{\tau_{0.9}^b}, \quad (5.33)$$

The sign of $\widehat{\Delta\tau_{0.9}}$ will thus inform whether the biological or the Hammerstein system is faster (see Table 5.2). The thorough analysis of all motifs has been exhaustively described elsewhere[88]. Here, the analysis of one motif, namely C2s (see Figure 4.3(a)), will be exemplarily described . This example is deliberately chosen for being the simplest (and thus its results are the easiest to understand); it will however also illustrate the limitations of this approach. The results for the rest will be presented in a compact manner; for a detailed study the reader is referred to Gayer[88].

Analysis of C2s.   The steady state curve of C2s could be well described by a Hill function, with an error of around 2.0 % (see Table 5.2). Furthermore, a detailed analysis shows that the *steady-state* parameters $K_{0.5}$ and h strongly correlate with the *kinetic* parameters of the motif (see Figure 5.10). $K_{0.5}$ increases logarithmically with k_2 (activity of the backwards reaction) and K_{m1} (Michaelis-Menten constant for the forwards reaction), but it decreases with K_{m2}(Michaelis-Menten constant for the backwards reaction; see Section A.1.2 for the definition of the equations defining the motifs and the corresponding parameters).

In fact, the logarithm of $K_{0.5}$ correlates linearly well with the sum of the logarithms of k_2, K_{m1} and K_{m2} (see Figure 5.10(d)), in the form

$$log\, K_{0.5} = log\, k_2 + log\, K_{m1} - log\, K_{m2} = log\, \frac{k_2 \cdot K_{m1}}{K_{m2}}. \quad (5.34)$$

This connection has a simple explanation: the higher k_2, the stronger would be the backwards reaction r_2, and thus stronger should be the input to overcome it. Similar arguments hold for
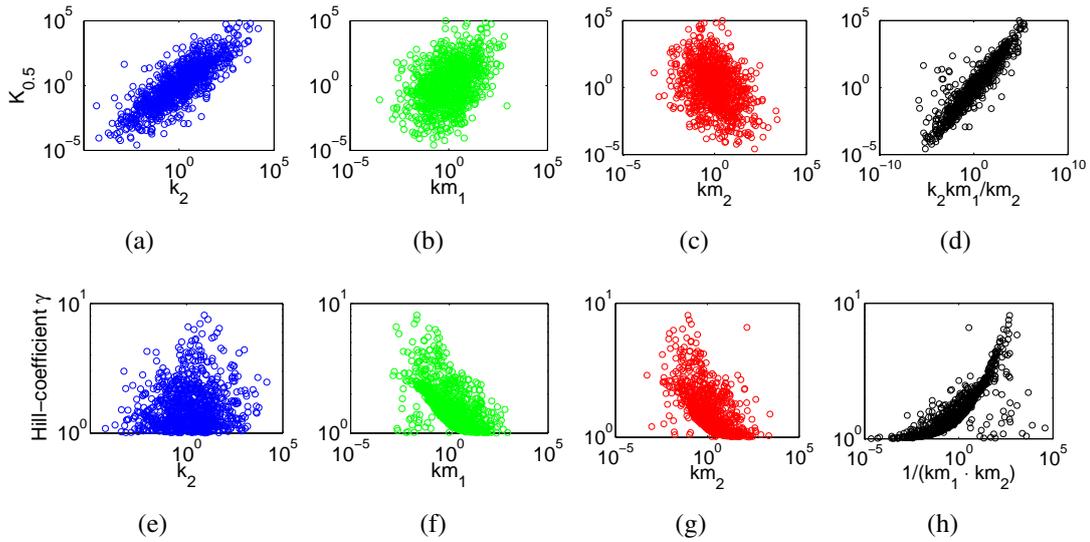
Figure 5.10: Correlation of $K_{0.5}$ and the Hill coefficient with the kinetic parameters of the motif C2s. The plots, in logarithmic scale, show the dependency of the constants $K_{0.5}$ and h of the Hill Equation (see Equation 5.26) on the kinetic parameters of C2s (see Section A.1.2 and Figure 4.3(a)). A linear dependency for $K_{0.5}$ as proposed in Equation 5.34 can be seen in the plot 5.10(d). For the Hill coefficient there is no simple linear dependency, and h depends in a non-linear fashion on $\frac{1}{K_{m1} \cdot K_{m1}}$, plot 5.10(h).

K_{m1} and K_{m2}: High K_{m1} values weaken the forwards reaction r_1 (and thus stronger inputs are required), and K_{m2} weakens r_2 (and weaker inputs are required). Besides, h is not correlated with k_2, but decreases with both K_{m1} and K_{m2}, being correlated in a non-linear fashion (see Figure 5.10(h)).

The situation is more complex with respect to the dynamics. The identification algorithm reveals that a simple PT_1 describes at best the dynamics. Furthermore, the error $\Delta\tau_{0.9}$ is relatively low ($\approx 10\%$). Interestingly, the mean of the parameter $\widehat{\Delta\tau_{0.9}}$ reveals that the Hammerstein module tends to require $\approx 10\%$ more time to reach the steady state than C2s. However, as Figure 5.11 shows, and in difference to a PT_1, $\tau_{0.632}$ is not constant with respect to the input. τ is indeed constant for low inputs (as was previously demonstrated analytically, see Equation 5.29), but after a certain value the signaling time decreases, eventually after a transient 'peak' (see Figure 5.11). This shows that, even for the simplest system, only a linear system with an input-dependent τ can approximate a biochemical motif well. Here, also correlations of the response time with the kinetic parameters can be retrieved. For example, the system responds faster for lower k_2 (as one would expect since the lower k_2, the weaker the 'brake' that the backwards reaction makes on the signal transfer).

General results. The linear systems which were most commonly identified as the equivalent to the different biochemical motifs are summarized in Table 5.2, together with the mean and standard deviation for both the steady-state Δy_s and dynamical $\Delta\tau_{0.9}$ error. The analy-
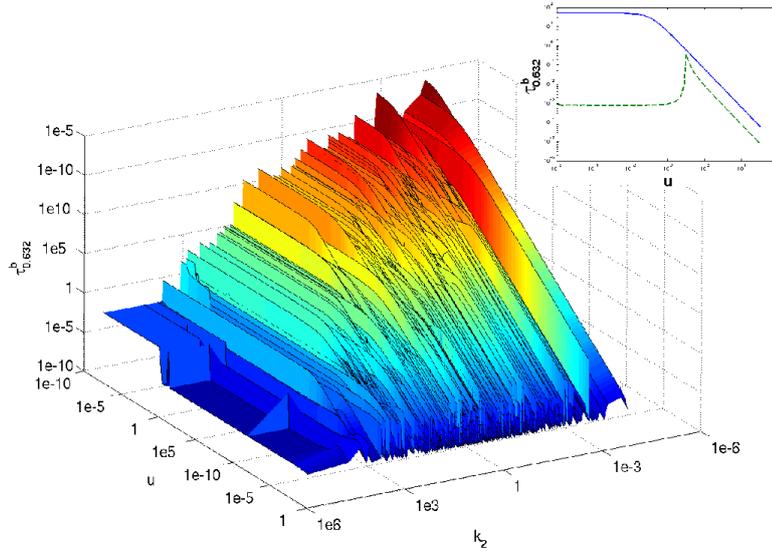
Figure 5.11: Dependency of $\tau_{0.632}$ on the input and k_2 for the motif C2s. The three-dimensional plot shows the dependency on both k_2 and the input u. Two characteristic plots of $\tau_{0.6321}$ vs. u (one with peak for input values around $K_{0.5}$ and one without) are depicted the upper-right corner.

ses were not performed for the motifs B1, C2B1, and C2B2 because they posed enormous numerical problems that were impossible to overcome. Note also that C3di is in this context equivalent to C3dd (compare Figures 4.5(a) and 4.5(b)). Besides, for C3ii (see Figure 4.5(c)) only co-stimulation via E_1 and E_2 can lead to an output signal (A^{**}). This would require to consider a multiple-input system, which is out of the scope of this work. Therefore, both inputs were considered to have the same value, leading thus to an equivalent system to C3dd.

We had seen in Section 5.1 that some of the motifs (specifically, C2p, C3dd, C3di, and C3sr) could show multistable behavior under a mass-action law description. During the exploration of the parameter space performed here, and even though we now simplified the enzymatic kinetics into a Michaelis-Menten expression, we found for all of them (and only for them) cases of bistability. These results thus are in agreement with those of Section 5.1 (note that with the approach described in this Section we could *not* exclude multistability for the cases where we could not find it, something we could do with the methods of Section 5.1).

In general, the approximation of the characteristic curve by the Hill function was very satisfactory (see Table 5.2), being the average error in the worst case (C2n) $\overline{\Delta y_s} \approx 5.0\%$. Furthermore, the standard deviation of the error $\sigma_{\Delta y_s}$ was also relatively narrow for most cases, laying in the range $0.5 - 18\%$. The widest $\sigma_{\Delta y_s}$ was again for the motif C2n, implying that in this case there are a relevant (but not dramatic) number of cases far from the average, and thus with the particular structure of C2n (probably its negative feedback) lead in some cases to a characteristic curve difficult to describe with a Hill function. An inspection of the corresponding characteristic curves shows that, for large inputs ($u > K_{0.5}$), the curve is quite 'sigmoidal' (large h), but, in contrast, less sigmoidal (lower h) for low inputs ($u < K_{0.5}$), a

Table 5.2: Summary of the analysis of the monotony and dynamics of the motifs depicted in Figures 4.2, 4.4, and 4.5. The monotony was determined according to Angeli et al.[7]. The second column shows the linear system which most commonly was identified as the closest to the motif[87]. Additionally, the mean and standard deviation for both the steady-state Δy_s (Equation 5.31) and dynamic $\Delta\tau_{0.9}$ (Equation 5.32) error for each motif, as well as $\widehat{\Delta\tau_{0.9}}$ (Equation 5.33) are summarized. The later analyses were not performed for the motifs C2B1, and C2B2, B1 due to numerical problems.

Motif	S. I/O monotone	Closest lin. system	$\overline{\Delta y_s}$	$\sigma_{\Delta y_s}$	$\overline{\widehat{\Delta\tau_{0.9}}}$	$\sigma_{\widehat{\Delta\tau_{0.9}}}$	$\overline{\Delta\tau_{0.9}}$	$\sigma_{\Delta\tau_{0.9}}$
C2s	$\checkmark$	PT_1	0.0016	0.0089	-0.0923	0.2496	0.1037	0.2451
C2p	$\checkmark$	PT_1	0.0252	0.1566	-0.1385	0.1895	0.1484	0.1819
C2n	$\checkmark$	PDT_2	0.0512	0.4305	-0.1588	0.1286	0.1596	0.1276
B1	$\checkmark$	-	-	-	-	-	-	-
C2B1	$\checkmark$	-	-	-	-	-	-	-
C2B2	$\checkmark$	-	-	-	-	-	-	-
C3dd	$\checkmark$	PDT_2	0.0009	0.0254	0.1349	0.2098	0.2023	0.1459
C3di	$\checkmark$	PDT_2	0.0009	0.0254	0.1349	0.2098	0.2023	0.1459
C3ii	$\checkmark$	PDT_2	0.0009	0.0254	0.1349	0.2098	0.2023	0.1459
C3sr	$\checkmark$	PDT_2	0.0266	0.1288	0.2262	0.2752	0.2714	0.2308
C3dp	X	PDT_2	0.0019	0.0056	0.9963	0.0374	0.9976	0.0185
C2C2	X	PT_1	0.0349	0.1731	-0.0481	0.2737	0.1197	0.2308
C3C2	X	PT_1	0.0258	0.1551	0.2125	0.2333	0.2597	0.1793

shape that can not be captured with a Hill function, since the shape of the curve is symmetric with respect to $K_{0.5}$.

The approximation of the long-term response (recall that for short-time the identification algorithm guarantees a good approximation) is also quite satisfactory ($\approx 10.0 - 25.0\%$), with the notorious exception of C3dp. This motif shows a clear difference in both $\Delta\tau_{0.9}$ and $\widehat{\Delta\tau_{0.9}}$ of ≈ 100 and at the same time a remarkably narrow set of values. This result suggest that the corresponding hammerstein motif is always much faster for time greater than $\tau_{0.632}$. Also noteworthy is the fact that the motifs set up exclusively single cycles (C2s, C2p, C2n, C2C2) are slightly faster than the corresponding Hammerstein module, while the rest, which involve double cycles are slower.

Interestingly, in many cases there was a peak in the signaling time τ (see e.g. Figure 5.11, upper-right corner, green line) for the region of input values around $K_{0.5}$ (see Equation 5.26 and Figure 5.7), where the steady-state curve shows the largest slope (see Figure 5.12). This means that the system gets 'sluggish' in this region. A simple explanation for this fact reads as follows: at the points at the steady state curve it holds $\vec{c} = 0$. For inputs slighltly higher than $K_{0.5}$, the system runs for a while close, almost parallelly to this curve, in regions where $\vec{c}$ is small, and thus it needs a long time to evolve.
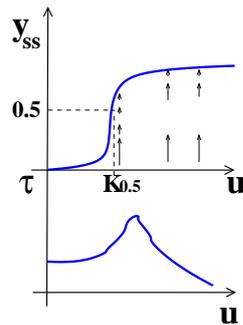
Figure 5.12: Correlation observed between y_{ss} and τ and a simple explanation. The figure is an scheme resembling typical plots of the characteristic curve and τ (ccf. Figure 5.11). For inputs u slightly higher than $K_{0.5}$ the system evolves in a region where the field is very weak, and gets thus more sluggish.

5.4 Conclusions

In this chapter, the motifs presented in Section 4.2 have been thoroughly analyzed with respect to three essential properties: multistability, monotony and input/output behavior. The results provide a first step towards a characterization of signal transduction building blocks.

To explore the multistability, a number of methods (bifurcation analysis, identification of positive feedback loops, Feinberg's Chemical Reaction Network Theory (CRNT), and a novel method to characterize the parameter space where multistability is possible[46]) have been applied, and their results compared. It was shown that CRNT provides a fast and reliable method to uncover mechanisms leading to multistability (limited by the number of compounds that the toolbox can handle). To be more precise about the nature of the multistationarity, bifurcation methods are required: here, they revealed that bistability (a particular form of multistationarity of high biological relevance) was present in all motifs showing multistationarity. Furthermore, the statistical analysis of the bifurcations of these motifs for different parameter sets illustrates the new possibilities which opens a method as that described in[46], which allows to analytically characterize the parameter space where multistationarity takes place.

On the other hand, the presence of positive feedback in the Jacobian matrix, even though being a condition for multistationarity, does not seem to be a conclusive tool since positive feedbacks are present in the Jacobian matrix of virtually any signal transduction system. A more sophisticated analysis of the nature of the feedback loops, however, could help to determine which positive feedbacks have potential to induce multistability[237].

The analysis shows under which conditions multistationarity can be expected and suggests that, in general, it may not be very robust against concentration fluctuations. Since the noisy nature of signaling and gene expression processes is a well-established fact[127;200;209], this results may pose a word of caution with regard to the biological relevance of single-motif multistability, particularly in the light of recent experimental data in human cells[240],

yeast[208], and bacteria[66]: the cell-to-cell variation in the protein level within a cell population would shift many cells out of the bistability range of a certain switch.

It may be however that single-motif switches, when embedded in cascades with positive feedback regulation, expand their parameter space of bistability[168]. Alternatively, such a switch would be plausible if the concentration of the corresponding protein would be subject of strict regulation. Additionally, it would be interesting to know to what extent the parameter sets leading to a large Δ_{A_T} are physiologically plausible; if they were, it would be tempting to speculate whether they are *de facto* implemented in cellular networks.

The analysis also suggests that autocatalytic reactions are stronger promoters of multistationarity than the competition between substrates and enzymes in distributive double-step activation processes. This would argue for a more common use of the former mechanism to obtain robust bistable behavior. Actually, cases of bistability controlled by autocatalytic steps are well-described[16;192], while those mediated by distributive double-step activation processes have been recently discovered and only theoretically.

Thereafter, it was explored, using a Michaelis-Menten formalism, the monotony and input/output dynamic behavior. Remarkably, all motifs revealed to be monotone except one, which at least has a monotone characteristic curve. Finally, the motifs were analyzed with regard to the input/output dynamic behavior. Since the non-linearities of these systems are notorious, a combination of a non-linear characteristic curve and a simple linear system was chosen as a reasonable compromise between simplicity and reality. For the characteristic curve, and motivated from the results on the monotony which guarantee that all motifs have a monotone characteristic curve, we adopted a sigmoidal curve in the form of the Hill function, which proofed to describe well the steady state behavior of the motifs. As far as the dynamics are concerned, all motifs seem to have a simple behavior, mainly analogous to that of a simple first-order system (PT_1) or a second-order system with differential part (PDT_2). However, an important drawback for the analysis is that the time-constant depends on the input.

Natural future steps of this work would be to analyze how these properties evolve upon the aggregation of motifs. For example, whether a cascade comprising a bistable motif is also bistable, and how the robustness of this switch differs from that of the single motif. Another interesting question would be to explore to what extent real models set up as a combination of these monotone motifs are themselves monotone, or what part of them[67]. This would allow to determine which subnetworks are monotone and thus well-characterized in a mathematical sense[242]. Regarding the input/output behavior, it would be particularly interesting to model a realistic signaling network with Hammerstein modules, and compare the behavior of the resulting model with a model set up using the classical reaction-network formalism, and with experimental data. In case this formalism would prove applicable, one would benefit from the structure of the model (a pure combination of non-linear characteristic curves and linear

dynamical systems), what would facilitate the analysis enormously.

It is well known that coupling of modules can lead to the appearance of new, emergent properties[20]. Therefore, probably unexpected and exciting discoveries wait for us along this path. A mouthful of them will be presented in Chapter 7.

Chapter 6

Structural Analysis
of Large Signaling Networks

Once a model has been decomposed into modules, one has to define the modeling approach to use for the analysis. There is a large array of possibilities. The choice depends mainly on the question to be answered, the quantity and quality of the data available, and the complexity of the system[123;242]. Some of them are depicted in Figure 6.1. Generally, as the rigor of the approach increase, the level of detail and potential insights increase, but at the cost of higher experimental and computational requirements, shrinking thus the size of the system to be analyzed. For reviews on different approaches the reader is referred to[56;123;250]. In this thesis, the application of two approaches will be illustrated: first, in this chapter, a rather rough approach, based on an interaction-based logical description, will be used. In Chapter 7, a detailed approach, based on a mechanistic description using ordinary differential equations (ODEs) as formalism, will be applied.

Kinetic models - either described as ODEs or as stochastic systems (see Figure 6.1) - provide a suitable formalism to study cellular networks. Actually, it is arguably the most applied in the field. It has proved successful in unraveling certain properties of networks of moderate size[2;4;226;275]. In Chapter 7, and using the methods introduced in Section 5.3.1.2, different aspects will be investigated, ranging from the new properties emerging from the concatenation of simple motifs to a modularity-based model reduction.

The mathematical structure of kinetic models can be more or less obtained from biochemical maps but they require for their parameterization a huge amount of quantitative data which is currently not available for large-scale networks. Furthermore, even supposing one would have such an enormous amount of data, the challenges it would pose on steps such as parameter estimation and model discrimination are far beyond the current state of the art[147].

Therefore, at the present stage, methods which require only information on the topology (structure) of the network are very useful, particularly considering the current considerable efforts to reconstruct maps of signaling networks, relying on a number of novel experimental and theoretical methods[187;198].

	Interaction-based			Constraint-based	Mechanism-based	
	Undirected graph A—B, C	Directed graph A→B, C	Logical network A→•→B, C	Stoichiom. matrix $0 = N\,v$ $A + B \rightleftharpoons C$	Kinetic chemical network $A + B \underset{k\text{-}1}{\overset{k1}{\rightleftharpoons}} C$	
Stoichiometry	N	N	N (Y logic)	Y	Y	Y
Dynamic	N	N	N (Y logic)	N	Y	Y
Quantitative	N	N (Y bayes.)	N/discrete	Y	Y	Y
Parameters	N	N(+/-)	+/-	N	Y	Y
Stochasticity	N	N (Y bayes.)	N	N	N	Y

+ detail
+ knowledge to win
+ difficult to get data
- size of model

Figure 6.1: Some mathematical approaches for the analysis of signaling networks. As one moves from the left (undirected graph) to the right (stochastic, master-equation-based modeling), the detail and potential insights increase, but at the same time more data is required and thus smaller systems can be analyzed. For more details on modeling approaches see[56;123;250].

Several methods relying on graph-based statistical analysis have been applied to signaling networks[21;124;165;286] and other types of protein interaction networks[125]. These studies are important for examining statistical properties and for understanding the global organization of a certain network, but they are limited for providing insights into the *modus operandi* of the network. To address these functional questions, a large corpus of methods has been developed for metabolic networks relying mainly on the constraint-based approach[201;247;250]. However, methods allowing a functional analysis of signaling networks have been applied to a much lesser extent[198].

Arguably, the shortage of such analyses could be mainly due to two reasons: On the one hand, the absence of enough information about signaling networks (which is however changing in recent years). On the other hand, the intrinsic dynamic behavior of signaling processes limits the applicability of constraint-based methods (which have proved to be successful for metabolic networks), since they assume steady state conditions. However, there have been some efforts to apply constraint-based methods to signaling networks[52;195;196]. Even though this approach proved able to undercover interesting properties, it is limited to networks that show an actual mass flow through the network[196], or need a hand-made modification to consider effects such as activation and inhibitions[52], ubiquitous in signaling networks.

Due to these drawbacks, a different approach was followed here. As shown in Section 5.2, the recurring set of motifs present in signaling networks are monotone, and their characteristic curve (see Section 5.3) can be in most cases estimated with a Hill curve. The Hill curve tends to adopt the form of a switch as the Hill coefficient (i.e., the ultrasensitivity) increases

(see Figure 5.7). It is therefore tempting, and to a certain extent justified, to try to describe these motifs as simple switches which can be on or off. Admittedly, this representation is an enormous simplification but, on the other side, leads to a huge increase of the size of the models which can be handled, as will be illustrated in Section 6.3, where a model including 94 different compounds is analyzed. Particularly important is the fact that one does not require values of parameters, but only information on the structure of the network. Therefore, a *structural* analysis can be performed.

The problem of applying a mass-flow-based framework (the constraint-based methods) to a signal-flow system (signaling networks), is not present in the approach used here, which is specifically tailored to signal transfer processes. However, a part of the methods are adapted from structural analyses of metabolic networks (but using a new formalism), and take thus advantage of the developments there. For example, concepts such as elementary flux modes are meaningful for signaling networks.

In Section 6.1, the foundations for a structural analysis of large signaling networks using a logical (Boolean) framework will be briefly introduced, suitable tools will be presented in Section 6.2 and its applicability illustrated by a realistic model of T-cell activation in Section 6.3. The usefulness was demonstrated by its ability not only to reproduce known data, but also and more importantly, by its predictions of unexpected results which where subsequently tested experimentally.

6.1 Methodological foundations

6.1.1 Representation of a signaling network as a logical hypergraph

The starting point for the analysis is an interaction graph, which is very similar to the typical biochemical maps (see Figure 6.2(c)): nodes representing molecules (e.g. an adaptor protein or a kinase) are connected via directed arcs indicating that a certain compound has an effect upon another one. These arcs are signed: they have either positive (activation, for example mediated by phosphorylation) or negative (inhibition, e.g. mediated by dephosphorylation) nature. Formally, an interaction graph (IG) is a signed directed graph $G = (V, A)$, where V is the set of vertices or nodes (species) and A the set of labeled directed edges (arcs)[56]. An arc from vertex i to j with sign s is denoted by an ordered tuple (i, j, s) with $i, j \in V$ and $s \in \{+, -\}$.

Apart from based on a map such as Figure 6.2(a) (e.g. from a database), which provides directly the IG, one can set up the model from the mechanistic description of a signaling network. If the kinetic model is modularly structured as described in Section 3, the conversion into the IG is particularly straightforward: one just need to consider which modules influence which ones, and whether it is a positive or a negative influence. Consider the toy system depicted in Figure 6.2(a). Two Receptors (R1 and R2) can bind via the corresponding domains

(Rb1 and Rb2) to 2 different ligands (L1 and L2). Binding of ligand promotes a conformational change at a second domain of the receptors (Rp1 and Rp2, respectively), allowing the binding of the kinase K1 (there are 2 pools of K1, one related to each receptor). Binding of K1 to Rp localizes K1 to the proximity of the membrane, allowing it to phosphorylate the membrane-bound kinase K3. K1 bound to the Rp1 can, in addition, phosphorylate the kinase K2, which is constitutively bound to R1. Both K2 and K3 phosphorylate the transcription factor TF1, activating it. They also phosphorylate a second transcription factor TF2 but at different sites; therefore, the joint action of K2 and K3 is required to active TF2. K2, in addition, activates a phosphatase Ph which in turn dephosphorylates (and thus deactivates) K2. This information can be easily coded in the IG depicted in Figure 6.2(c).

(a) Biological map

(b) Modular kinetic model

(c) Interaction graph

(d) Logical interaction hypergraph
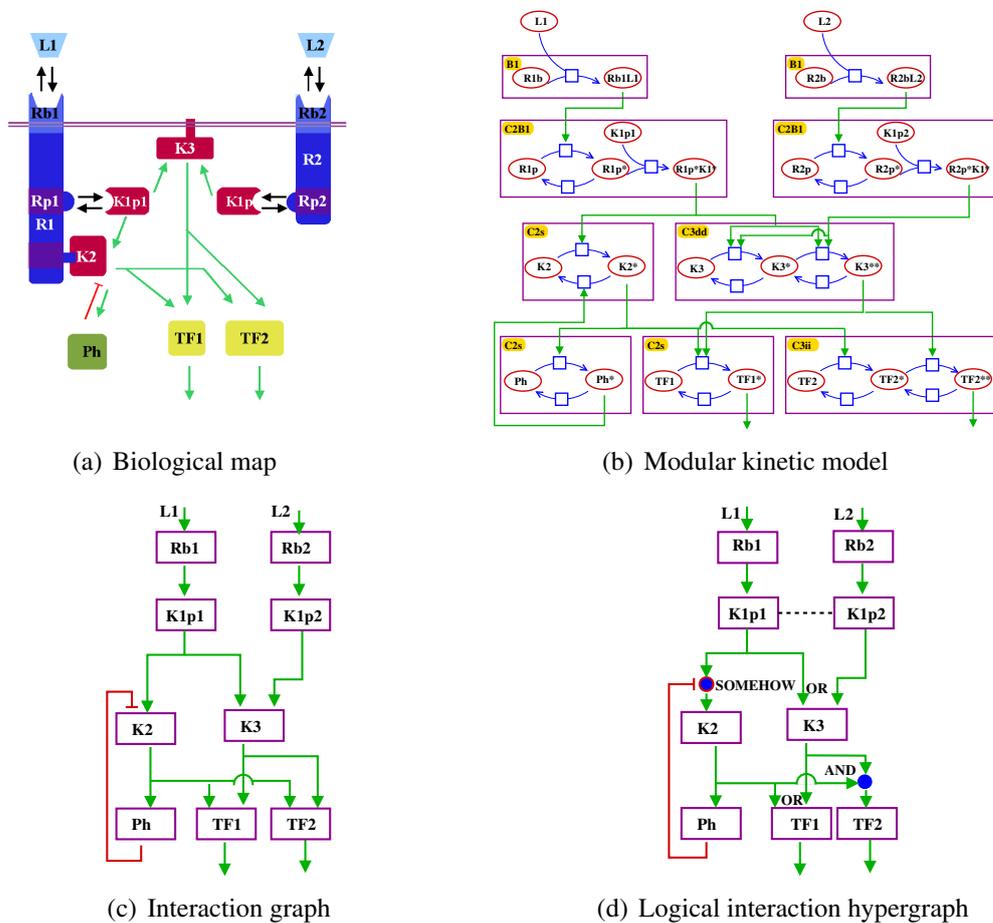
Figure 6.2: Formulation of a simple network as a kinetic model, an interaction graph (IG), and a logical interaction hypergraph (LIH). A typical biological cartoon is depicted in (a). Given a modular, mechanistic description (b), the conversion of the kinetic model into the IG (c) is relatively straightforward. Adding logical operators to the IG, one obtains the LIH (d). Here additional mechanistically information as in the IG is required. Sometimes the connection is not clear and a SOMEHOW (incomplete truth table) may be applied, see text. The dashed line denotes that both pools depend on a common reservoir. The labels in the golden boxes of the modules in (a) correspond to the modules defined in Chapter 4.

An IG provides useful information (and in fact part of the analyses presented here are based on it), but is somewhat ambiguous: in the example in Figure 6.2(c), it is for example not clear whether one requires K2 AND K3 to activate TF1, or any of them alone is enough.

To obtain a more precise description, Boolean operators (AND, OR and NOT) are introduced. AND-relationships are not possible in graphs but in hypergraphs. Similar to a directed graph, a directed hypergraph $H = (V, A)$ consists of a set V of nodes and a set A of hyper-arcs. A hyperarc a connects two subsets of nodes: $a = \{S, E\}, S, E \subseteq V$. S comprises the start nodes and E the end nodes of the connection. S and E can comprise an arbitrary number of nodes (actually a graph is a special case of a hypergraph where all arcs consist of one input and one output nodes). Specifically, a special representation of Boolean functions known as disjunctive normal form (DNF, also called 'sum of product' representation) is used, which uses exclusively AND, OR and NOT operators. Therewith, one can describe any Boolean operator in an hypergraph. We shall call the hypergraph corresponding a Boolean network with Boolean functions in DNF form logical interaction hypergraph (LIH)[143]*.

Now, it is clearly defined that e.g. both K2 AND K3 are required to activate TF2, while in the case of TF1, either K2 or K3 are enough (see Figure 6.2(d)). More mechanistical information - which is generally not provided by the biological map (Figure 6.2(a)) - is required to define the LIH than to define the IG: when two compounds can fulfill the same task (e.g. phosphorylate a certain substrate, as K2 and K3 in Figure 6.2(b)), this process is considered as a logical OR. When both are required such as when the effect of two kinases is required to activate a substrate (e.g. K2 and K3 to activate TF2 in Figure 6.2(b)), an AND connections is used.

However, the mechanistic details are not always enough to determine the logical operators. A recurring example occurs when the phosphorylation of a protein is dependent on a kinase and its counteracting phosphatase (e.g. K1 and Ph acting on K2 in Figure 6.2(b)). Here, it is clear that if the kinase is active and the phosphatase is not active, the substrate would be phosphorylated; vice versa, if the kinase is inactive but the phosphatase is active, it will not get phosphorylated. However, what occurs when either both are active or both inactive is not clear, and depends on the amount of active kinase and phosphatase and their relative activities. If detailed experimental information is available, it might be possible to assign a logical operator. However, this is often not the case, and thus an incomplete truth table (where only the known effects are determined) of the form

K1	1	0	1	0
Ph	1	0	0	1
K2	?	?	1	1

is required, leading to the definition of the operator 'somehow'.

---

*Note that, each LIH has a unique underlying interaction graph (which can be easily derived from the LIH representation) whereas the opposite is, in general, not true.

There is an additional relationship which should be considered: the 2 pools of K1 are activated independently (upon binding to R1p and R2p, respectively), but they are actually set up of equivalent molecules. Therefore, a perturbation such as a genetical knock-out would affect both pools. To model this circumstance, we introduce an additional state, K1r (reservoir of K1), which is by default active and is required to activate both pools of K1 (i.e., is AND connected to Rb1 and Rb2, respectively). We shall not represent explicitly the reservoirs, and draw a dashed line connecting them to the corresponding pools; in this way, the mathematical information (there is a reservoir present) is maintained in the graphical representation, but coded in such a way that it is not confused with other kind of information. This and other visualization considerations are conveniently handled in ProMoT's framework for logical models, which will be presented in Section 6.2.

By converting the kinetic model into a LIH, each module is substituted by a logical switch. Generally, each module (component) can be either on ("1") or off ("0"). We consider a compound to be on if it is fully activated and able to trigger downstream events properly; otherwise it is off. While the models and analyses described here are based on this binary description, the formalism and methods can be easily extended to multiple levels, e.g. off ("0"), weakly activated ("1"), and fully activated ("2").

The analysis of regulatory mechanisms using Boolean formalisms is a commonly used technique[256]. However, it has been applied so far only to systems of moderate size, e.g.[39;133;174]. Here, in contrast, we aim to handle large signaling networks. The following subsections will introduce succinctly the methods used.

6.1.2 Analysis of the interaction graph

The structure of a signed graph can be stored conveniently by an $m \times q$ incidence matrix $B \in \{0, 1, -1\}^{m \times q}$ in which the columns correspond to the q arcs (interactions) and the rows to the m nodes (species), similar to the stoichiometric matrices of metabolic reaction networks[107]. In fact, considering the graph as a reaction network with the arcs being irreversible mono-molecular reactions, the incidence matrix would be equivalent to the stoichiometric matrix. For the k-th arc (i, j, s) a (-1) is stored in the k-th column of B for the tail vertex i and $(+1)$ for the head vertex j of arc k, i.e. $B_{ik} = -1$, $B_{jk} = +1$, and $B_{lk} = 0$ ($l \neq i, j$). For storing the signs, a q-vector s is introduced so that $s_k = +1$ if arc k is positive and $s_k = -1$ if k is negative.

From the incidence matrix of an interaction graph we can identify important features like feedback loops as well as signaling paths and network-wide interdependencies between pairs of species (e.g. perturbing species A may have no effect on B as there is no path connecting A to B).

Feedback loops are of major importance for the dynamic behavior of the network[44;214;256].

Positive feedbacks, as pointed out in Section 5.1[†], are responsible for multiple steady state behavior[244], which plays a central role in differentiation and other decision-making processes[65;277]. In contrast, negative feedback loops are essential for homeostatic mechanisms (i.e. for adjusting and maintaining levels of system variables) or for generating oscillatory behavior[257].

The computation of all the paths between a pair of species i and j allows the classification of i with respect to j as

- **activator**, if i is involved in at least one positive path and in no negative path to j,

- **inhibitor**, if i is involved in at least one negative path of and in no positive path to j,

- **ambivalent factor**, if i is involved in at least one negative and one positive path to j, and

- **non-influencing**, if there is no path from i to j.

In the case of activator and inhibitor, a subtle additional distinction should be made: i is a **total activator** (total inhibitor) if there is no path from i to j that touches a negative feedback circuit. This definition is required since the contact to a negative feedback loop hinders predictions on the effect of a perturbation (e.g. increasing the activity of a total activator of a species s would unambiguously lead to an increase of s, but in the case of a non-total activator it could decrease after some time). The influences among species can be represented in a compact manner via the dependency matrix D, where the element D_{ij} denotes the nature of i (activator, inhibitor, etc.) with respect to j[‡].
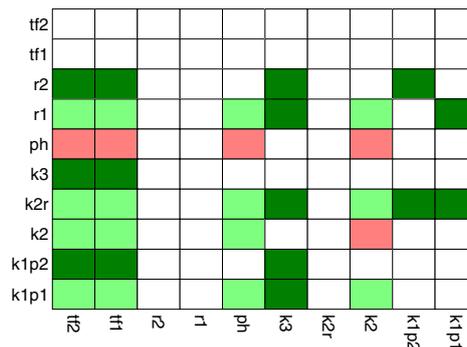
Figure 6.3: Dependency matrix of the simple Boolean model of Figure 6.2(d). The color of a matrix element D_{ij} has the following meaning[143]: (i) dark green: i is a total activator of j; (ii) light green: i is a (non-total) activator of j; (iii) dark red: i is a total inhibitor of j; (iv) light red: i is a (non-total) inhibitor of j; (v) white: i does not influence j.

[†]Note that in Section 5.1 we were interested in positive feedback inside a module, while here we are looking for network-wide loops.

[‡]Actually, to classify the pairwise effects among the nodes and construct the dependency matrix, it is not necessarily required to compute all paths between all species i and j. Computing the shortest positive and negative path, which is a much less computationally intensive task, is enough[143].

Remarkably, the determination of feedback cycles and of all the signaling paths between any pair of species from the incidence matrix is equivalent to the computation of elementary modes known from metabolic networks [143]. This allows to use the highly optimized algorithms for computing elementary modes. Other concepts of metabolic network analysis, such as Minimal Cut Sets can also be applied [143].

Even though important properties can be obtained from the IG (such as identification of all or particular signaling pathways, feedback loops and crosstalks, and network-wide interdependencies between network elements), as argued above, this description is limited. For example, in the simple model of Figure 6.2(c), one could not know whether, upon addition of ligand, TF1 would get activated. Therefore the LIH must be analyzed, as will be explained in the next section.

6.1.3 Analysis of the logical interaction hypergraph

The structure of the LIH can be stored, similarly to the IG, conveniently by an $m \times q$ incidence matrix $B \in \{0, 1, -1\}^{m \times q}$ so that $B_{ik} = -1$ if i is contained in the set of start nodes of a hyperarc k, $B_{jk} = +1$ if i is the endpoint , and $B_{ik} = 0$ if it is not involved. Analogously to the vector s in IG, a matrix $U \in \{0, 1\}^{m \times q}$ is defined so that $U_{ik} = 1$ if and only if i enters the hyperarc k with its value negated.

One can compute in the LIH the effect of defining a set of input stimuli on downstream signaling by computing the resulting logical steady state [143]. Sometimes a logical steady state is not unique or does not exist due to the presence of feedback loops. However, many feedback loops become active only in a longer time scale justifying setting them "off" in the first wave of signal propagation (allowing them to be switched "on" for the second timescale). This motivates the introduction of time-scales at which the different interactions can be active [143].

The effect of knocking-out a species can be easily tested by re-computing the logical steady state for the respective stimuli. Thus, one can check the plausibility and consistency of the network structure by confronting the predictions of the model, for both 'wild type' (unmodified network structure) and 'knock-out' (changed structure) with experimental data (as done in Section 6.3).

A useful concept for the analysis of LIHs is that of Minimal Intervention Set (MIS) [143]: a minimal collection of components that must be *externally* activated or deactivated to fulfill a certain task such as the permanent activation/deactivation of certain compounds. This allows to identify key elements for a certain process: for example, elements recurrently present in the MISs for the activation of a certain protein are important for its activation. This idea can be used, for example, to identify potential oncogenes, by analyzing the MISs that lead to the activation of the transcription factors controlling proliferation of a cell (see Section 6.3). The analysis of MISs is thus closely related to structural network properties like redundancy

and robustness. MISs can also help to find targets and intervention points in the network for repressing or provoking a certain behavior or response, useful for drug target identification. Furthermore, if a model fails to reproduce a certain experiment, MISs can be applied to provide candidates to fill this gap in the network structure. An example illustrating this use can be found in Section 6.3.

6.2 ProMoT and CellNetAnalyzer: Tools to setup and analyze logical models of large signaling networks

The methods described above have been embedded in *CellNetAnalyzer* (CNA), a toolbox for Matlab[170] facilitating, in an interactive and visual manner, a comprehensive structural analysis of metabolic, signaling and regulatory networks[144]. All the analyses described here were performed in CNA.

In CNA, the user should provide a graphical map of the network, a mathematical (textual) input of the network structure, and a mapping from the latter to the former. However, the setting up procedure for large-scale networks by hand, of both the graph and text, can be a cumbersome and error-prone task. There are many tools available to set up models describing signaling networks as a biochemical reaction network[85;95]. However, there is currently no tool available that allows the visual setup of large logical networks, and has the ability to export both the mathematical model and the graphical representation together.

Therefore, the abilities of the modeling tool ProMoT[95] (whose original application is the set up dynamical models, see Sections 3.3.2.3 and 4.3) were extended to fill this gap[223]. First, a library of basic logical elements (`compound`, `and`, `not`, etc.) was set up, which also possesses properties that contain additional information. Subsequently, we developed advanced visualization techniques that allow the user to customize the representation of the logical model. Finally, an export to CNA (and other formats, such as a Matlab function expressing the Boolean network) was implemented, and other exports can be easily implemented[223].

In the following, ProMoT's main properties with regard to Boolean models will be described.

6.2.1 Definition of a library of basic elements in ProMoT

There are two main classes in the modeling library, `compound` (representing a state), and `gate` (defining a logical interaction between compounds). Applying ProMoT's object-oriented modeling paradigm[95], subclasses of a certain class can be easily defined. For example, we have defined subclasses of the class `compound` (e.g. `receptor`, `kinase`, and `reservoir`), which are all mathematically equivalent to `compound` but can be specifically considered in the later visualization process, for example, using different colors for the different subclasses.
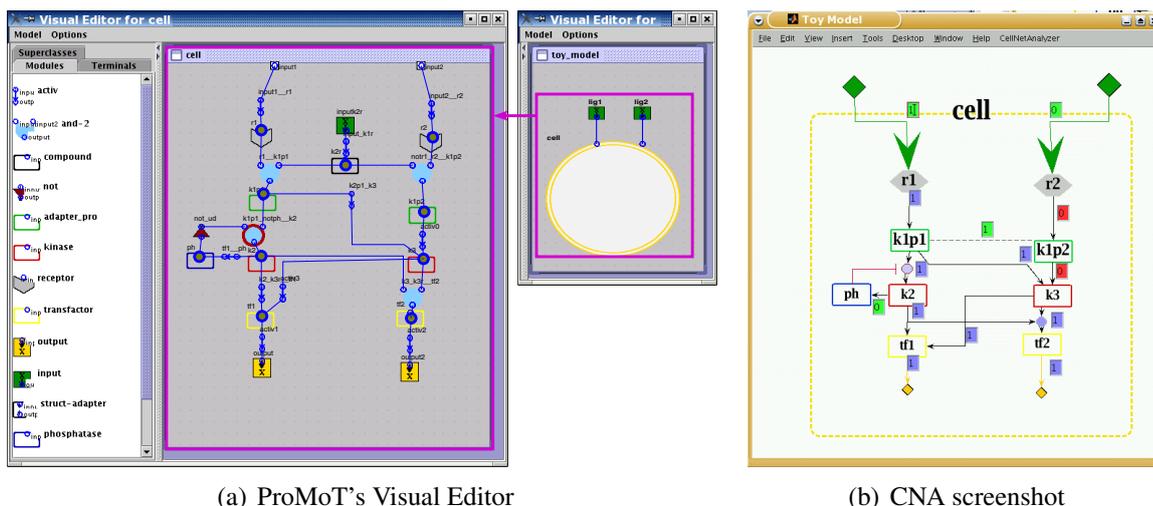
(a) ProMoT's Visual Editor

(b) CNA screenshot

Figure 6.4: Screenshot of the visual editor of a toy model in ProMoT (a) and of its visually processed export to CNA (b). The model corresponds to the LIH depicted in Figure 6.2(d). In Figure 6.4(a), a palette containing some of the modules used to set up logical models can be seen, as well as the model for the whole network (small box) and its module describing the content of the cell. This model was treated using a convenient visual scenario, and exported to CNA. Figure 6.4(b) shows a screenshot upon the computation of a logical steady state when only the ligand L1 is present. Note how, the feedback Ph →K2 was switched off, since it has a time-scale $\tau=2$ and here only the early events ($\tau=1$) are considered. Small text boxes display the signal flows along the hyperarcs (green boxes: fixed values prior to computation; blue boxes: hyperarcs activating a species; red boxes: hyperarcs which are not active).

Thereby, additional documentation of the model is provided with the map (Figure 6.4).

To define different logical connections among the elements we subclassed the class `gate` into `activ` (to describe a causal one-to-one relation between two compounds), `and` (to define the requirement of several elements to activate a certain compound), and `not` (to express a negative effect, i.e., an inhibition). An OR gate can be implemented by including several `activ` elements pointing at a certain compound. Since any logical connection can be described as a combination of ANDs, ORs, and NOTs (see Section 6.1), the set of basic elements described above allows to set up any logical network of arbitrary size. In addition, to describe cases where the logic is unclear (see Section 6.1), we have also included the class `somehow` which represents logical gates with partially incomplete truth tables.

Finally, the classes `input` and `output` allow to define the incoming and outgoing signals of the model, respectively.

Properties can be easily added to the different classes. For example, we have defined parameters for the default value and time-scale, which are exported with the model. Multiple levels (i.e., discretizing the states into more than two (0,1) levels) are also implemented using the properties of the gates[223]. Additionally, all elements have a documentation, which can also be exported with the model.

6.2.2 Exploiting ProMoT's modularity and visualization techniques

ProMoT allows to set up models in a modular manner[95]. This unique property can be used to define different submodules, either physically delimited (e.g. a cell, see Figure 6.4(a)) or comprising a particular submodel (e.g. the cell cycle regulation). For example, in the toy model introduced in the previous section, one module is defined to model the content of the cell (see Figure 6.4(a)).

The elements described above enable to precisely set up a mathematical model, but the representation is not biologically intuitive. This issue is tackled with scenario-based visualization techniques (compare Figure 6.4(a) and Figure 6.4(a)). A visual scenario - which can be easily edited - describes a set of mapping functions defining the visual properties of the different elements. Therewith, biologically important elements are emphasized and elements that are non-relevant are de-emphasized or hidden. For example, in the visual editor, inhibition is encoded by a `not` element between the compound and the gate (Figure 6.4(a)) and, after applying the visual scenario, it is represented by a single red-colored connection line. Additionally, the direction is indicated by an arrow symbol (different for activation and inhibition), which is implicitly defined in the mathematical description. Also the elements of the class `reservoir` are hidden (e.g. `k1r` in Figure 6.4). As pointed out it in Section 6.1.1, reservoirs are mathematical entities with no biological meaning and a very particular interpretation. Therefore, they are hidden and the information of their presence is coded via a dashed line (compare Figure 6.4(a) and Figure 6.4(b)).

Furthermore, setting the property `visibility`, the content of a particular module can be shown in the graph or not. The earlier might be desired if the module comprises different elements of biological interest (e.g. a module describing the nucleus), while the latter might be useful for modules describing a particular logical operation as a set of basic elements.

The ability of ProMoT to set up a large logical network was demonstrated with a comprehensive model of T-cell signaling, which will be described in the following section.

6.3 Structural Analysis of T-cell signaling

The methodology is applied to a curated model of T-cell signaling, which was introduced in Section 2.2.2. The signaling network was constructed, as much as possible, from data collected from primary T-cells (from the literature and own experiments), and only well-established connections were considered. Several players, in particular, some whose role and activation is not completely understood, are not included in our model and thus their effects are not considered or lumped with others. Additionally, in several, currently still controversial cases, we have assumed one of the possible hypotheses; however, this does not mean that we propose this to be the correct description of the TCR-induced signaling network; we just want to demonstrate the applicability of our approach on a realistic, complex case. The logical model (Figure 6.5 and Tables A.1 and A.2) describes the main events and elements connecting the T-cell receptor (TCR), its coreceptor CD4, and the costimulatory molecule CD28, to the activation of transcription factors, e.g. AP-1, NFAT and NFκB, that determine T-cell function. In general, it includes: the activation of the Src kinases Lck and Fyn, followed by the activation of ZAP70 and the subsequent formation of the LAT signalosome (which in turn triggers MAPKs and calcium cascades)[118]. Additionally, it includes the activation of the PI3K/PKB pathway that regulates many aspects of cellular activation and differentiation, particularly survival.

The high number of kinases, phosphatases, adaptor molecules and their interactions give rise to a complex map (94 compounds and 123 interactions) which cannot be interpreted via pure intuition (see Figure 6.5).

Note that a species can represent different states of a molecule: for example, CD45 refers to the availability of CD45 to act on its substrates (Lck and Fyn), PLCg(bind) refers to PLCγ1 bound to LAT, and PLCg(act) to the active (bound to LAT and phosphorylated) form of PLCγ1. It is also important to realize that several steps can be lumped together or expressed in higher detail; for example, the two steps of c-Cbl's effect (ubiquitination and degradation) are lumped in the hyperarcs pointing to its targets ZAP70 and TCR.

Two time-scales (see Section 6.1)[143] are considered: early (τ=1) and late (τ=2), involving processes occurring during or after the first minutes of activation, respectively (see Table A.2). Key regulatory processes such as the degradation of signaling proteins mediated by c-Cbl occur after a certain time, and are thus assigned τ=2.

This will be important in Section 6.3.2, where the analysis of signal propagation during the early events reveals which elements get activated (Table 6.1), and the consideration of the late events allows a rough approximation to the dynamic behavior (sustained vs. transient) of the network (Figure 6.9). We shall start first with some analyses based on the interaction graph. All the investigations performed on the T-cell model were performed with CNA.
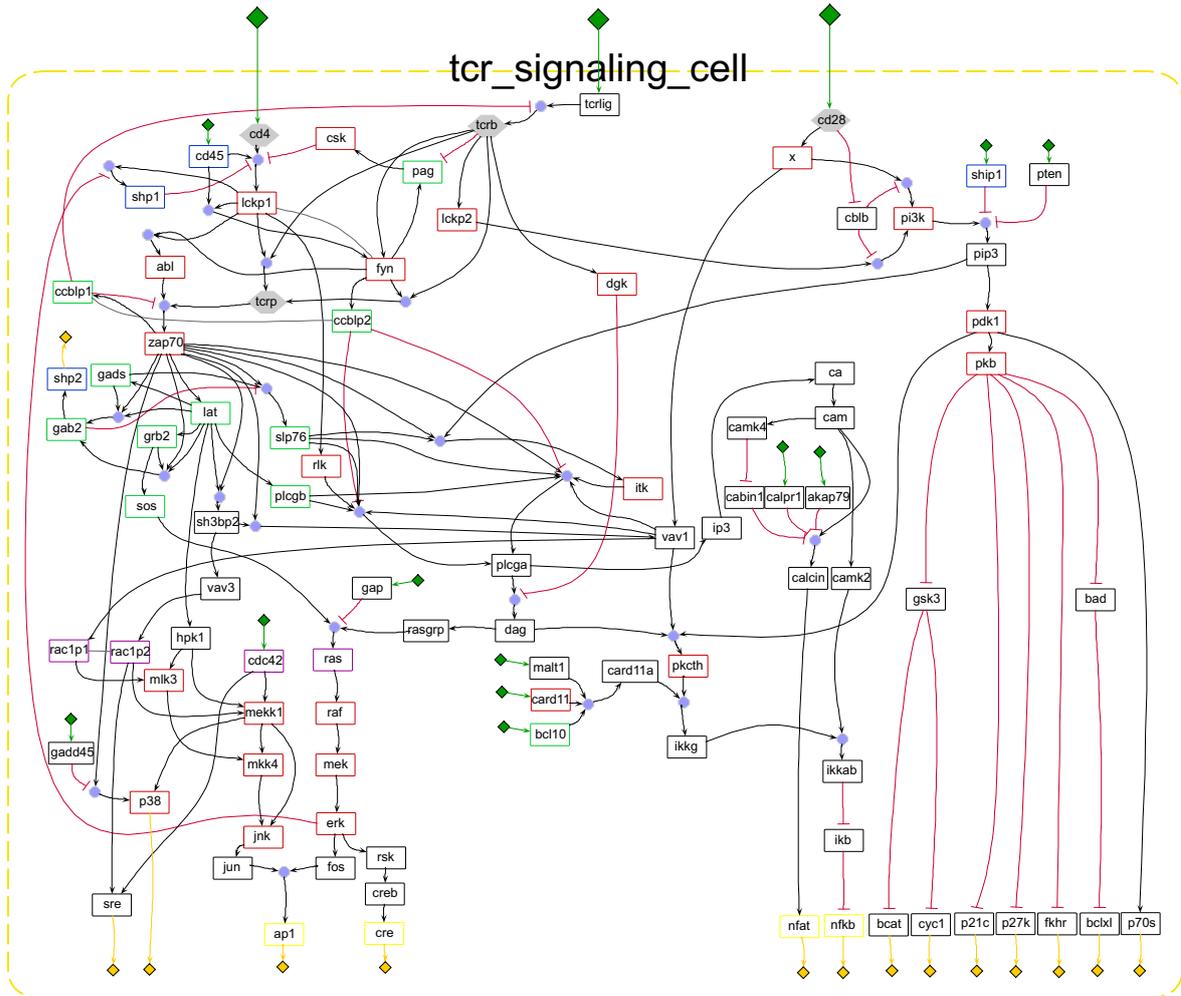
Figure 6.5: Logical model of T-cell activation. The model was implemented in ProMoT (see Section 6.2). Each arrow pointing at a species box is a hyperarc representing one possibility to activate that species (see Section 6.1). All the hyperarcs pointing at a particular species box are OR connected. Yellow species boxes denote output elements, while green ones represent (co)receptors.

6.3.1 Interaction-graph-based analysis

The interaction graph underlying the logical model has 172 feedbacks, thereof 89 negative. As mentioned above, feedback loops are of major importance for the dynamic behavior of the network. Here, the feedback loops are only active in the second time-scale because each loop contains at least one process of the second time-scale. The elements of the MAPK cascade are involved in most (92%) feedbacks. This is due to the fact that there is a connection ERK→SHP-1 (from bottom to top of the network)[4], and the resulting feedback can return to ERK via many different paths leading to a high number of loops. If the ERK→SHP-1 connection is not considered, the number of loops is reduced dramatically to 13 (11 negative), located in the upper part of the network. c-Cbl is involved in ≈85% of them, underscoring its importance in regulating signaling[61].

There are 4538 paths, each connecting a compound from the input layer (TCR, CD4 and
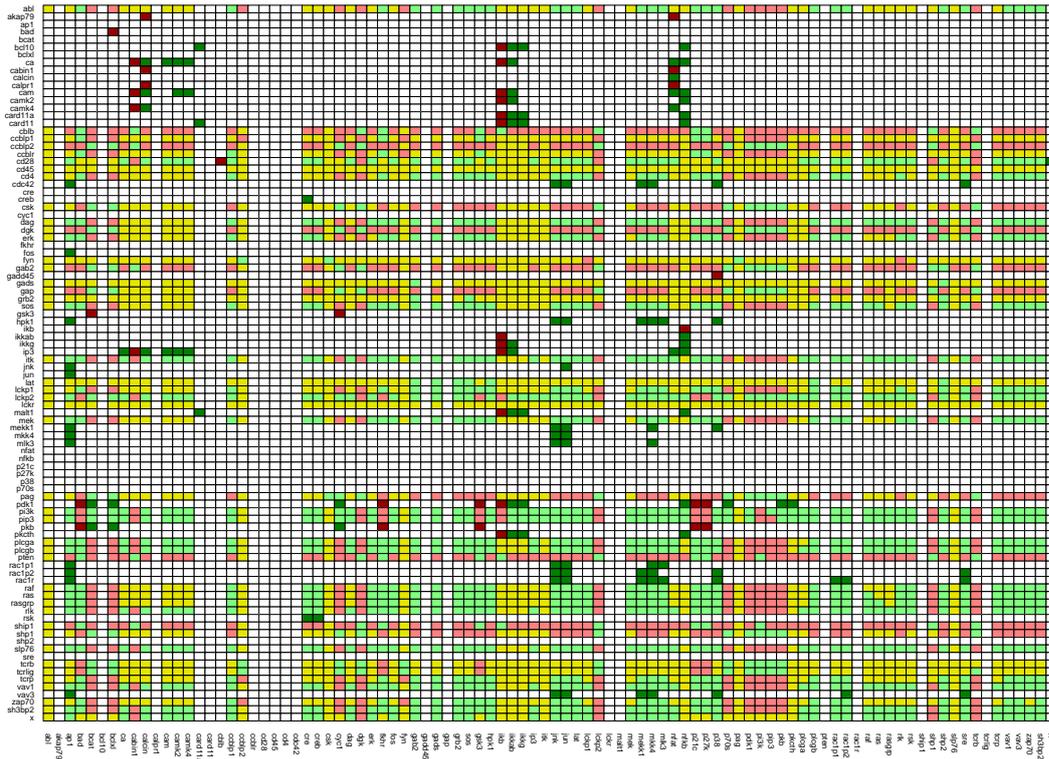
Figure 6.6: Dependency matrix of the logical T-cell signaling model when all interactions are active. The color of a matrix element D_{ij} has the following meaning[143]: (i) dark green: i is a total activator of j; (ii) light green: i is a non-total activator of j; (iii) dark red: i is a total inhibitor of j; (iv) light red: i is a non-total inhibitor of j; (v) yellow: i is an ambivalent factor for j; (vi) white: i does not influence j.

CD28) with a compound in the output layer (transcription factors and other elements controlling T-cell activation). The high number of negative paths (2058) can be traced back to the presence of two negative connections (via DGK and Gab2). In fact, considering the early events, where these mechanisms are not active, the number of paths is reduced to 1530, with only 6 of them being negative. These paths are from the TCR and CD28 to negative regulators of the cell cycle (p21, p27 and FKHR), having thus a positive effect on T-cell proliferation. These and other effects can be graphically inspected via the dependency matrix[143] (Figures 6.7 and 6.6). Remarkably, when considering time-scale τ=1, there is no ambivalent effect between any pair of species (e.g. TCR has only a positive effect on AP1, see Figure 6.7).

6.3.2 Logical-Interaction-hypergraph-based analysis

Using the logical model, we first analyzed the activation pattern of key elements upon different stimuli (activation of the TCR and/or CD4 and CD28). The model was able to reproduce data from literature and own experiments, providing a holistic and integrated interpretation for a large body of data.

Furthermore, the model predicted a non-obvious signaling event: activation of the costim-
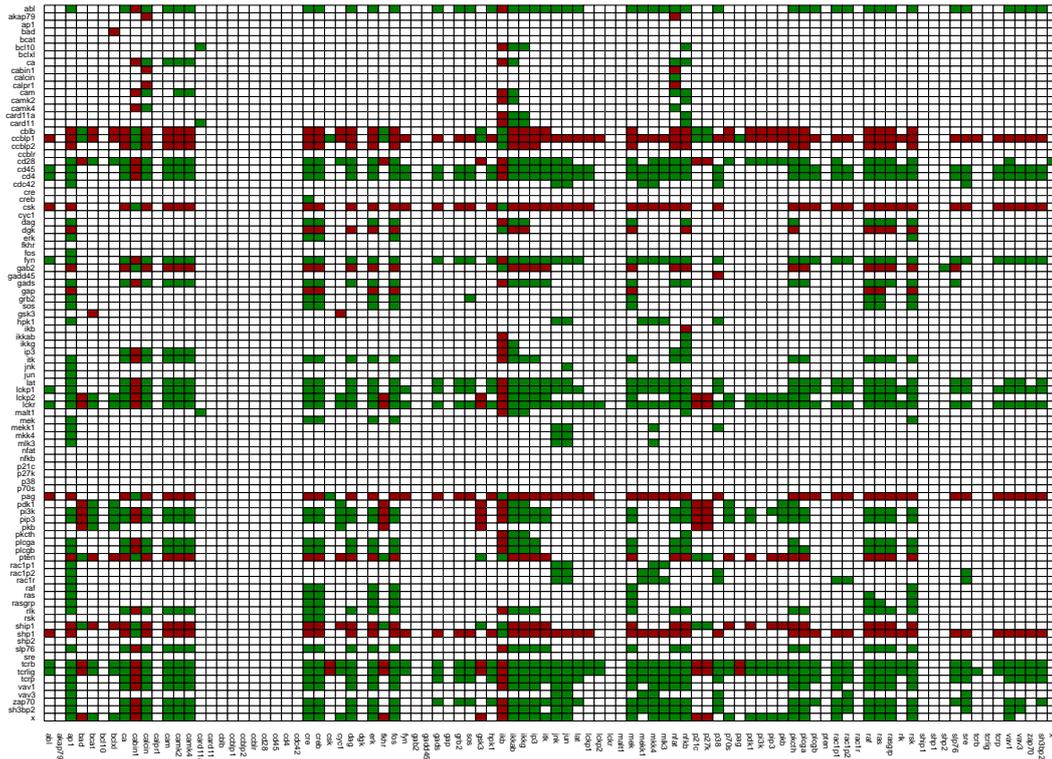
Figure 6.7: Dependency matrix of the logical T-cell signaling model for the early events scenario (τ=1). The color of a matrix element D_{ij} has the following meaning[143]: (i) dark green: i is n total activator of j; (iii) dark red: i is a total inhibitor of j; (vi) white: i does not influence j. Note that there are no non-total activators/inhibitors or ambivalent factors in the early events scenario (compare to Figure 6.6).

ulatory molecule CD28 alone leads not only to the activation of PI3K (which is expected), but also to activation of JNK (but not ERK) via Vav1. The published data supports the view that CD28 stimulation alone without TCR crosslinking is not sufficient for[102;253] (or induces only weak[98]) JNK activation. However, the model clearly predicted that we should observe JNK activation and provided a holistic explanation for it (see Figure 6.8).

Driven by this surprising observation, we performed the corresponding experiments in vitro. As shown in Figure A.7A, stimulation of mouse primary T-cells with a CD28 antibody induced an evident and sustained JNK phosphorylation (independent of PI3K, Figure A.8), thus confirming almost perfectly the predicted binary response. The discrepancies with the literature could be due either to the different cellular systems (primary T-cells in our case versus T-cell lines in the literature) or to the different stimulation conditions (e.g. different antibody concentrations, different medium, etc).

The nature of the kinase involved in CD28-mediated signaling remains unclear. Src kinases (Fyn and Lck) have been proposed to mediate CD28 signaling[113]. To test this hypothesis in our experimental system, we analyzed the effect of a Src kinase inhibitor (inactivating Lck+Fyn) on CD28-dependent events. The experiment showed that Src kinases are dispensable for CD28-mediated signaling (Figure A.9). Therefore, the model contains a
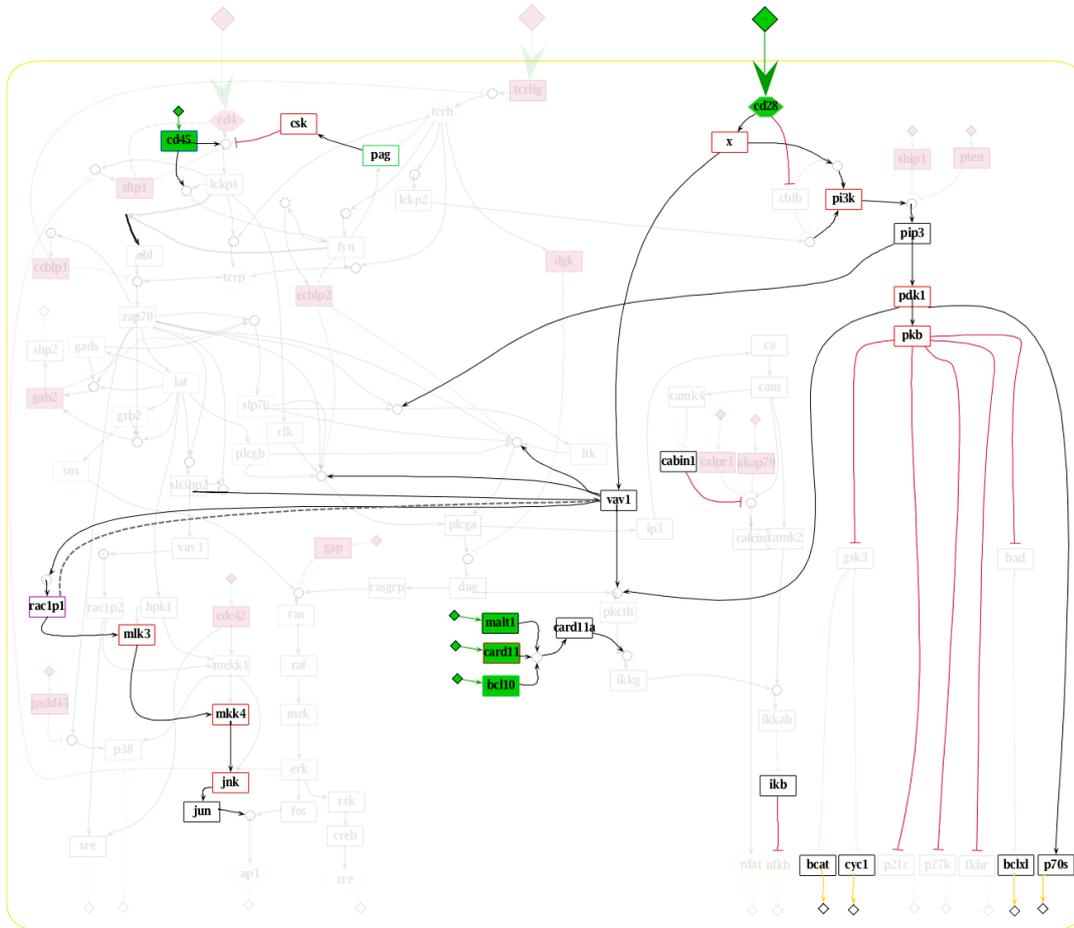
Figure 6.8: Paths activated upon CD28 stimulation in the logical model of T-cell activation. The analysis was performed in CNA and the results imported in the visual navigator of ProMoT for representation (see Section 6.2). Full boxes in green and red represent states fixed to 1 and 0, respectively. Transparent elements are not activated upon CD28 stimulation, while non-transparents are.

to-be-identified kinase X connected to CD28 (Figure 6.5), probably of the Tec family.

The model was also challenged to predict results under different knock-out conditions, and the results were compared with published data (Table 6.1). The model could reproduce the phenotype of several experimental knock-outs. Furthermore, as with the wild type upon CD28 stimulation, it also reported an unexpected result: activation of the TCR in Fyn-deficient cells triggers the PI3K/PKB pathway. This prediction was tested using Fyn-deficient mice, corroborating the model result (Figure A.7B). However, there was a result which the model could not reproduce experimentally: TCR-mediated JNK activation is blocked by PI3K inhibition (Figure A.7C). This result is not in accordance with the current network because PI3K has no influence upon JNK (see dependency matrix, Fig. S1).

To identify potential connections which could explain the data, the concept of Minimal Intervention Sets (MISs, see Section 6.1) was applied to determine which minimal combinations of elements should be directly or indirectly affected by PI3K to produce the observed

Table 6.1: Summary of the activation pattern predicted by the logical model for T-cell signaling upon different stimuli and knock-out conditions. The "KOs" row denotes the perturbed (switched-off) element. In the case of PI3K and Lck&Fyn, the perturbation was done via a chemical inhibitor, and for the rest it was through a genetic knock-out (see A.5). The "Input" rows show the stimuli, and "Output" the predictions of the model for key elements of the network. Therein, green boxes denote results corroborated by published data, while blue ones where confirmed by our own data (see line Fig./Refs.). The red box shows a discrepancy between model and experiment (see text).

KOs	WT	WT	WT	PI3K	PI3K	PI3K	SLP 76	Fyn	Fyn	Fyn	Rlk & Itk	Lck & Fyn	Lck & Fyn	Lck & Fyn
Input														
TCR	1	0	1	1	0	1	1	1	1	1	1	1	0	1
CD4	0	0	0	0	0	0	0	0	1	0	0	0	0	0
CD28	0	1	1	0	1	1	0	0	0	1	0	0	1	1
Output														
ZAP	1	0	1	1	0	1	1	0	1	0	1	0	0	0
LAT	1	0	1	1	0	1	1	0	1	0	1	0	0	0
PLCga	1	0	1	0	0	0	0	0	1	0	0	0	0	0
ERK	1	0	1	0	0	0	0	0	1	0	0	0	0	0
JNK	1	1	1	1	1	1	1	0	1	1	0	1	1	1
PKB	1	1	1	0	0	0	1	1	1	1	1	0	1	1
AP1	1	0	1	0	0	0	0	0	1	0	0	0	0	0
NFKB	1	0	1	0	0	0	0	0	1	0	0	0	0	0
NFAT	1	0	1	0	0	0	0	0	1	0	0	0	0	0
Fig./Ref.	A.7	A.7, A.8	A.8	A.7	A.8	A.8	254	A.7,[254]	A.7,[254]	A.7	228	A.9	A.9	A.9

result. The MISs from Table 6.2 provide a list of minimal combinations of elements that should be affected (directly or indirectly) by PI3K to explain the block of the TCR-dependent activation of JNK upon inhibiting PI3K .

Some of them are obvious, e.g. the first MIS in Table 6.2 suggests that JNK activation could be directly interacting with PI3K or downstream elements (e.g. PIP3). However, to the best of our knowledge, there is no experimental evidence for a direct effect of PI3K on JNK. An effect on the activation of Vav (both 1 and 3 isoforms) seems to be a good mechanism to include, since it appears in some of the MISs. Vav has a PH domain which can bind to PIP3, and this mechanism could be important for Vav activation[264], thus making it a reasonable extension of the model. Another element of importance is HPK1. Interestingly, HPK1 is phosphorylated by Protein Kinase D1 (PKD1)[9], a kinase dependent on PKC (in turn dependent on DAG, downstream of PI3K) for activation. Since the regulation and role of both PKD1 and PKC (with the exception of the θ isoform) are not well established in T-cells, we did not include them into the model, but a connection DAG→ PKC→ PKD1→ HPK1 would be plausible. An alternative could be a Rac dependent activation of HPK1[105]; however, this is a not-well-established connection.

Definitely, the model requires a direct or indirect connection from PI3K to JNK and testing the mechanisms proposed here in experiments could be a way to reconcile the network

Table 6.2: Application of the Minimal Intervention Sets (MISs) to identify candidates to fill the gap between PI3K and JNK. The MISs of maximal size 3 to obtain JNK off under the conditions (i) TCR on, (ii) PI3K off, and (iii) ZAP70 on (as shown in the experiment, see Fig. 2A and Table 6.1) were computed, setting the rest of conditions to the standard values for the early events. Here, each MIS represents one set of molecules that should be influenced by PI3K in order to be consistent with the fact that PI3K inhibition blocks JNK activation.

MIS	MIS
jnk	hpk1 mkk4 rac1p2
zap70	hpk1 mlk3 vav3
hpk1 rac1r	hpk1 rac1p1 rac1p2
hpk1 sh3bp2	hpk1 rac1p1 vav3
mekk1 mkk4	hpk1 rac1p2 vav1
mekk1 mlk3	hpk1 vav1 vav3
hpk1 mekk1 rac1p1	hpk1 mlk3 rac1p2
hpk1 mekk1 vav1	

structure with the PI3K-inhibitor data. This example illustrates another useful application of our approach: the model not only reveals that a link is missing, but also suggests candidates that can be verified experimentally.

Table 6.1 summarizes the results of 14 different scenarios, in which the logical model predicted 126 states. For 44 of them, experimental data was available (15 from literature and 29 from own experiments) confirming the predictions except the case discussed above: the model was able to predict that a PI3K inhibitor would block ERK activation, but not that it would also inhibit JNK activation (Figure A.7C).

Table 6.3: Minimal Intervention Sets (MISs) to produce the full activation pattern in T-cells. The MISs of maximal size 3 that induce sustained full activation (namely: ap1, bcat, bclxl, cre, cyc1, nfkb, p70s, sre and nfat are on, whereas fkhr, p21c and p27k are off) of T-cells without external stimuli. The exclamation mark denotes "deactivation"; species without this symbol have to be activated (constitutively). Interestingly, the compounds involved in these MISs are involved in oncogenesis. Note that, since PIP3 is a second messenger and not 'mutable', for the purpose of this analysis the MISs involving its activation can be considered equivalent to those involving its activator PI3K (i.e., these MISs are equivalent).

MIS
!gab2 pi3k zap70
!gab2 pip3 zap70
pi3k plcga zap70
pi3k slp76 zap70
pip3 slp76 zap70
pip3 plcga zap70
pdk1 plcga zap70

As an additional application of MISs, we computed combinations of failures (constitutive activation or deactivation of elements caused e.g. by mutations) which lead to sustained

T-cell activation without external stimuli. These failure modes would cause uncontrolled proliferation and thus may be connected to diseases such as leukemia or autoimmunity. Interestingly, the components occurring in the MISs with few elements (Table 6.3), are in fact involved in oncogenesis: ZAP70[108], PI3K[36], Gab2[19], PLCγ1[186] (and SLP76 is directly involved in PLCγ1 activation, see Figure 6.3 and Table A.2) .

We have focused here mainly on the analysis of which elements would get activated upon signal triggering (i.e. for the first time scale τ=1), for which a larger corpus of data is available (see Table 6.1). However, the model was also able to roughly predict the dynamics upon different stimuli and conditions (see Figure 6.9). This sort of analysis not only shows the ability of the Boolean approach to reproduce the key dynamic properties (transient vs. sustained) of a signaling process, but also underscores the redundancy of negative feedbacks in switching off signals: KOs of different key regulatory mechanisms would not have a definite effect on the dynamics of key output signals, and only a multiple KO could lead to a clear effect. It is well known that the form (transient versus sustained) of e.g. the activation of ERK is decisive in the cellular outcome[169;181;226]. Therefore, the effect on the dynamics should be carefully analyzed in order to understand the effects of a certain intervention. Thereby, a complementary approach would be to consider a time-dependent (synchronous or asynchronous) approach[56], where properties such as the robustness[39] of the dynamic behavior in signaling networks of moderate size could be studied .
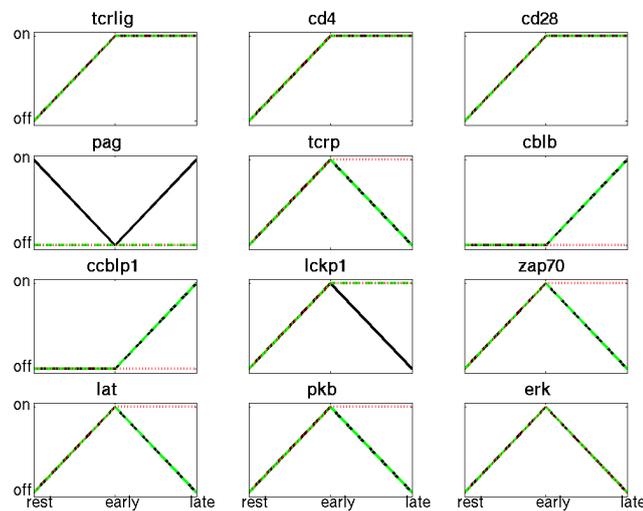
Figure 6.9: Rough description of dynamics considering different time scales. The activation of key elements upon activation of the TCR, the coreceptor CD4, and the costimulatory molecule CD28 is represented at the resting state, τ=0 (no inputs); early events, τ=1 (input(s), no feedbacks); and later-time events, τ=2 (input(s), feedbacks). The black lines correspond to a wild type while the green ones to a PAG KO. Note that the absence of PAG has no effect on key downstream elements of the cascade, due to the redundant role of other negative regulatory mechanisms (degradation via c-Cbl and Cbl-b, Gab-2 mediated inhibition of PLCγ1). Multiple knock-out of these regulatory molecules leads to sustained activation of key elements (red lines).

6.4 Conclusions

In this section, it was shown how large signaling networks can be analyzed using a Boolean formalism. One can perform logical steady state analyses unraveling the processing of signals and the global input-output behavior. Moreover, by converting the logical model into an interaction graph, one can extract further important features, such as feedback loops, signaling paths, and network-wide interdependencies between species. Furthermore, the concept of intervention sets allows one (a) to identify missing links in the network, (b) to reveal failure modes that can explain the effects of a physiological dysfunction or disease, and (c) to search for suitable intervention strategies, while keeping track of potential side-effects, which is valuable for drug target identification.

The methodology was illustrated with a comprehensive model describing T-cell signaling. While in the model set up we only considered *local* information about the interaction of signaling molecules (e.g. kinase *A* phosphorylates adaptor *B*), the model recapitulates *global* published data upon different stimuli (e.g. activation of receptor *R* lead to activation of transcription factor $TF1$) and knock-out conditions for both natural and perturbed conditions (e.g. knock-outs, inhibitors, mutations, etc.). More importantly, the model could predict unexpected results that were then verified experimentally.

At last but not least, convenient tools for the setup (ProMoT) and analysis (*CellNetAnalyzer*) within this Boolean framework have been developed.

This approach may surely not replace a detailed analysis based on a mechanistic description, and can certainly not answer the same questions. However, requiring a relatively small amount of quantitative information, it proved to be a promising in-silico tool for the analysis of large signaling networks. Particularly, it may be of great use in foreseeing the effects of drugs, mutations, and other network modifications. The following section is devoted to illustrate the use of a mechanistic-based approach where more subtle questions regarding the dynamics and quantitative differences can be precisely addressed.

Chapter 7

Dynamic Analysis of Modularly Structured Signaling Networks

In Chapter 6 a Boolean framework was presented that allows to handle large signaling networks at the price of neglecting details, particularly concerning quantitative and dynamic aspects. In this chapter, we shall discuss the analysis of signaling networks using kinetic, ODE-based models from a modular perspective. This framework is particularly suitable to analyze the aspects which a Boolean approach can not (but in a smaller scale), and is thus a complementary approach to the former.

First, it will be explored how the connection of simple motifs as those presented in Section 4.2 leads to the appearance of new, emergent properties. We will start with an open-loop concatenation of modules (the MAPK cascade, see Sections 2.2.1 and 3.4.1), which is also extended to consider a closed-loop system incorporating a negative feedback. As a second example, in Section 7.1.2 it will be shown how a simple feedback system can explain a non-trivial dynamic behavior observed upon activation of T-cells. Finally, the model for the EGF-induced MAPK cascade of Schoeberl et al.[234] (see Section 2.2.1) will be considered as a case of a given, non-modularly setup model. Using the decomposition outlined in Section 3.4.4, in Section 7.2 it will be shown how a modular approach can unravel properties of the whole network and facilitate a model reduction. We shall start by considering how the properties of modules change as they are aggregated into more complex models.

7.1 Aggregation of modules and emergent properties

A thorough analysis of signaling modules as the one presented in Chapter 5 is certainly a reasonable first step towards understanding signaling systems. However, as introduced in Chapter 1, signaling networks are characterized by a high number of crosstalks and feedbacks among the modules. In fact, aggregation of modules with different connections can lead to the appearance of properties not present at all in the single modules, so-called emergent properties[20]. Therefore, only by analyzing how these units work in an orchestrated

manner can we attempt to understand biological signal processing. This is, however, an extremely challenging task: as stated in the motivation (Section 1.1), the size and complexity of the systems to analyze is discouraging. Here, we shall illustrate the appearance of emergent properties with two simple examples, a concatenation of motifs to set up a MAPK cascade, and a simple positive feedback able to reproduce a complex behavior from T-cell signaling.

7.1.1 MAPK cascade

Let us consider a cascade of 3 motifs, one of the type C2s and two of the type C3dd (see Section 4.2). A prominent example of this structure is the MAPK cascade, introduced in Section 2.2.1.1 (see Figure 2.2), whose system-theoretical properties have been studied extensively. To illustrate its analysis, we shall take, from the variety of models available, the implementation of Kholodenko[134], which has also been discussed with regard to its modularity in Section 3.5(b).

The characteristic parameters for the signal amplitude S and signaling time $\tau_{0.9}$, as defined in Section 5.3.1.2, are computed for the three modules of the MAPK cascade (MAPKKK, MAPKK, and MAPK) and for the whole cascade operating in open and closed loop, see Figure 7.1, using steps of different magnitude as inputs.
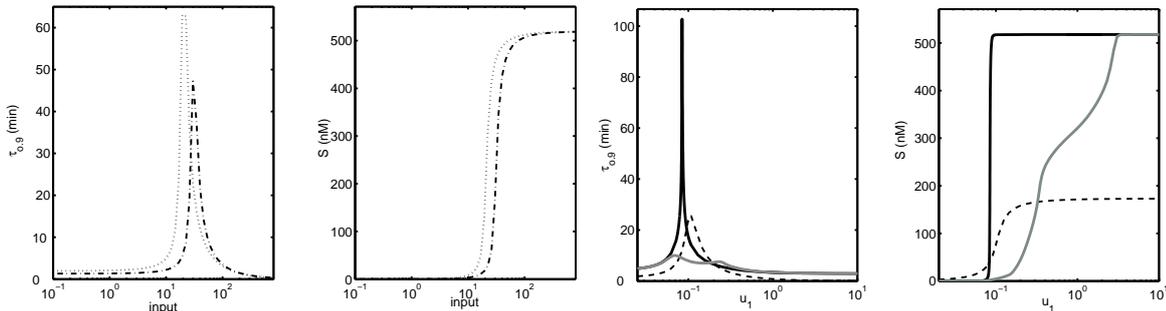
(a) $\tau_{0.9}$ and S for the MAPKK (dash-dot line) MAPK (dotted line) modules.

(b) $\tau_{0.9}$ and S for the MAPKKK module (dashed line), as well as the whole MAPK cascade in open (solid black line) and closed (solid grey line) loop.

Figure 7.1: Signaling time and signal amplitude for the MAPK cascade and its subunits. All the subunits show an ultrasensitive behavior and a peak of the signaling time around the threshold value for the signal amplitude. The entire cascade combines the sigmoidity of the 3 subunits, and the negative feedback decreases the signal amplitude (see Figure 2.2).

The MAPK cascade is a paradigm of a modular system: through 3 subunits and eventually a feedback loop, the MAPK is able to perform several tasks. Probably the most evident property of such a three-step structure is its potential to amplify the input signal[75]. In our case study, however, since the concentration of the three proteins is the same, there is no such amplification. Additionally, the characteristic curve of the MAPK cascade provides a sigmoid (ultrasensitive) I/O behavior[75] (see Figure 7.1). The three modules show ultrasensitivity, with Hill coefficients (see Section 5.3.1.1) of 4.0 for MAPKKK and 6.6 for MAPKK

and MAPK, due to the saturation of the enzymes in the case of MAPKKK, and, in the case of MAPKK and MAPK, due to the dual-phosphorylation mechanism and to the saturation of the enzymes[75]. Moreover, the different levels add their ultrasensitivity[76], producing high steepness in the curve of the total signal amplitude (Hill coefficient of 111).

Interestingly, the threshold 0.085 of the cascade is close to, but slightly lower than, the threshold 0.103 of the first module, meaning that the system does not need the complete activation of the first module to reach full activation. Additionally, the maximal signal amplitude of the whole cascade corresponds to the maximal possible signal amplitude of the last module (Figure 7.1), meaning thus that the MAPK module can be fully activated while connected to the others.

(a) Simulation results   (b) Bifurcation analysis
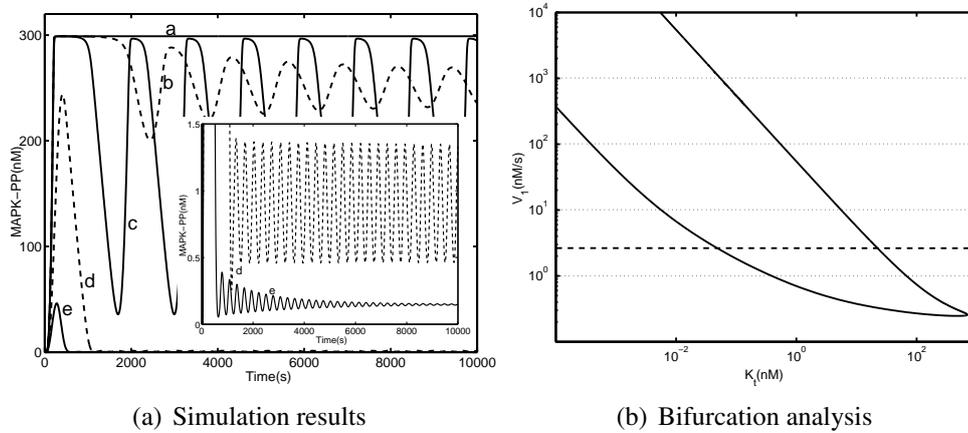
Figure 7.2: Oscillations in the MAPK cascade. The different simulation results were obtained vary-
ing the feedback strength of the model of Kholodenko[134]: (a) $K_t=10^5$; (b)$K_t=23.5$; (c)$K_t=9$;
(d)$K_t=6\,10^{-2}$; (e)$K_t=10^{-3}$. The input is $V_1 = 2.5nM/s$ for all the cases. In the bifurcation graph
(Figure 7.2(b)), the black line shows Hopf bifurcation points as a function of the feedback
strength (K_t value) and input (V_1 value). For values between both lines sustained oscillations
occur. The dashed line shows the input value (V_1) used by Kholodenko[134] (Hopf bifurcations
points at $K_t=0.0532$ and $K_t=23.6$).

The three modules and the whole cascade show a sharp deceleration of the response around the threshold value, as shown by the parameter $\tau_{0.9}$. This result was already observed in Section 5.3 in the systematic analysis of different signal transduction motifs (which included the motifs C2s and C3dd corresponding to Raf and MEK/ERK, respectively).

The peak is higher and narrower for the total cascade than for the first module parameter (Figure 7.1). Far from this peak value, the whole cascade is mostly slower than the single modules. This is an intuitive result; however, it does not need to be the case and, depending on the parameter values, cascade of motifs can accelerate signal propagation[106]. In fact, for a narrow range of inputs ($\simeq$ 0.1-0.3 nM), the whole cascade is faster than the first module.

Several MAPK cascades have been found to be embedded in feedback loops, both positive and negative (see Figure 2.2), enriching the versatility of the MAPK cascades. A positive feedback can, together with the inherent ultrasensitivity of the MAPK cascade, produce a

bistable system and, if the feedback is strong enough, the system is able to give an *irreversible* on/off response to a transient continuous stimulus[77;78;277].

Negative feedback can potentially drive the system to return to the basal state after a transient response to a constant input, a phenomenon known as adaptation[11]. Additionally, it can introduce sustained oscillations[134]. In the model of Kholodenko[134], the inclusion of a negative feedback loop decreases the response time, but only around the peak, and a decrease in the signal amplitude (Figure 7.1). Depending on the strength of the feedback, sustained and damped oscillations can be observed[134]. A bifurcation analysis shows that, over a wide range of feedback strengths determined by the value of K_t, there is a range of the values of the input V_1 for which the system shows sustained oscillations[220] (Figure 7.2). For strong feedback values, the oscillations disappear and the output signal decreases to almost zero.

7.1.2 Identification of the minimal realization to describe the dynamics of the TCR-induced MAPK cascade

In this section, a simple model is used to tackle an interesting question arising from a set of experimental data upon activation of T-cells. We shall start describing the experiments, and subsequently the model will be described and analyzed.

7.1.2.1 Dynamics of the TCR-induced MAPK cascade: experimental results

As outlined in Section 2.2.2, antibodies have been extensively used in T-cell signaling research, even though they trigger an unphysiologically strong stimulation. In the last years, the use of streptamers (or tetramers) loaded with MHC-peptides has provided a more realistic system[115], closer to the real activation of a T-cell by an APC (Antigen Presenting Cell, see Section 2.2.2). Importantly, streptamers produce a costimulation of CD4/CD8, and allows the use of different ligands with different affinities (agonist, antagonists, etc.).

To understand the dynamics of the TCR-induced MAPK cascade, experiments were performed by X. Wang at the Institute of Immunology (University of Magdeburg) with naive primary T-cells extracted from the spleen of OT-1 transgenic mice (see Section 2.2.2). In the experiments, T-cells were stimulated either with antibodies (against both the CD3 subunit of the TCR and the coreceptor CD8), or with streptamers loaded with the agonist for the TCR of the OT-1 mouse. Thus, one can compare the effect of two different inputs: the moderately strong one driven by the physiological affinity of the agonist for the TCR, and the extremely strong one triggered by the unphysiological affinity of the antibodies.

Upon activation, the time-dependent phosphorylation pattern of some key elements (using specific antibodies against ZAP70 at position p_{319}, $LAT\,p_{171}$, PLCγl p_{783}, $ERK\,pp$, and $AKT\,p_{473}$) was quantified relative to the maximal value (see Figure 7.3). Additionally, the blots revealed a 'ladder' structure in the phosphorylation of ZAP70, characteristic of ubiquitination processes, which lead to degradation[61]. Therefore, the total amount of ZAP70

was quantified (via an antibody which binds to ZAP70 independently of its phosphorylation state). The amount of ZAP70 was normed with the amount of ZAP70 in the resting state. Subsequently, the difference with respect to the total amount was computed and, indeed, a significant degradation could be observed (see Figure 7.3). The experiments were performed three times, and the mean, together with the standard error of the mean, were computed.
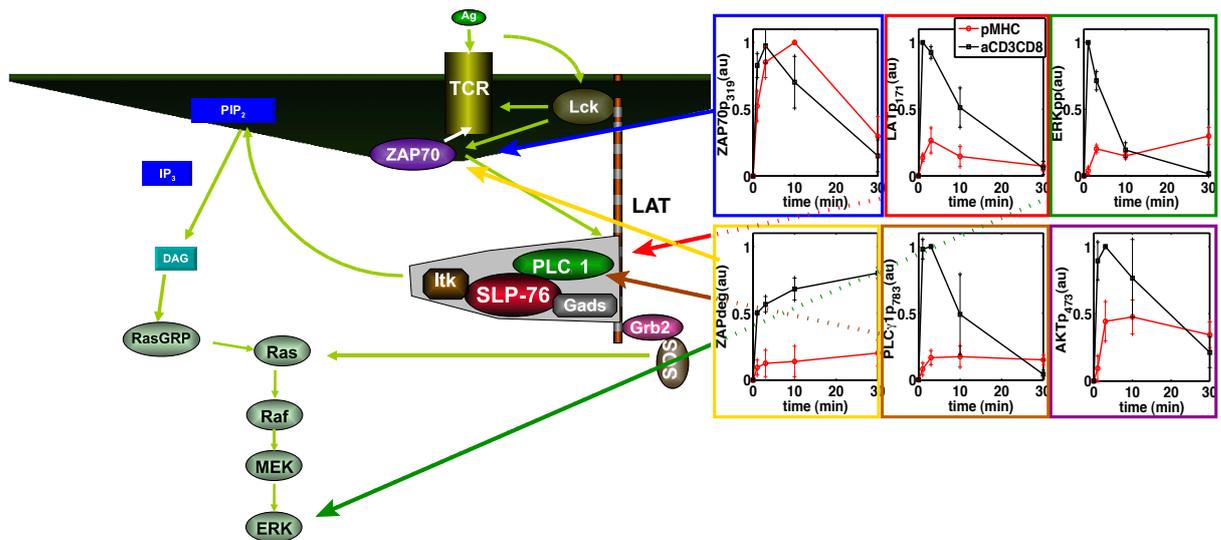
Figure 7.3: Data on the dynamics of key molecules of the TCR-induced MAPK cascade. The time-dependent phosphorylation pattern of key molecules in primary T-cells is plotted upon stimulation with CD3+CD8 antibodies (black lines) and streptamers loaded with the agonist for the OT-1 TCR (red lines). The experiments were performed by X. Wang at the Institute of Immunology, University of Magdeburg.

The experimental results show that a strong input (antibodies) leads to the propagation of a strong, transient signal throughout the cascade, while a weak input (streptamers) triggers a moderate, sustained signaling. Furthermore, T-cells proliferate vigorously upon streptamers stimulation, but weakly upon antibody perturbation (data not shown). Collectively, these data suggest that antibody-mediated (=strong) signals initiate negative feedback loops (probably involving ZAP70 degradation) which could serve to prevent inappropriate T-cell activation.

A reasonable next step would be to try to describe this data with a mathematical model, to confirm that ZAP70 degradation can lead to the observed behavior (the experiments show a correlation, but not a proof of causality), providing thus support to the hypothesis that ZAP70 is responsible for this behavior. This goal will be pursued in the following section.

7.1.2.2 Failure of a bottom-up approach drives to adopt a top-down perspective

First, we tried to explore the utility of a bottom-up perspective to tackle this question. Even though a bottom-up approach is arguably the most common one while facing questions such as ours, it is mostly convenient when there is a relative large amount of data and knowledge. In our case, however, we have only information about 6 states to build up a model describing 60; we decided to test whether this was a tractable problem at all. Therefore, using the approach outlined in Section 4.3, a detailed kinetic model describing the TCR-induced MAPK cascade was set up in ProMoT. It was trained with the experimental data depicted in Figure 7.3. The model, however, was not able to reproduce the data (even though it possesses enough complexity to do so). This was the case probably due to parameter estimation problems: the computational effort and complexity of the estimation problem grow exponentially with the number of parameters. Even though a battery of methods was used[*], no successful fit could be found. Since the number of parameters to fit is around one hundred, it is reasonable to suspect that those algorithms could not find the global optimum.

This result confirmed our hypothesis that it was not the most convenient approach to tackle the analysis of the data. Accordingly, an opposite, 'top-down' approach was adopted. The idea is to perform the following steps:

1. recognize the key behavior emerging from the data,

2. determine what is needed *mathematically* to achieve this behavior, and

3. identify these mathematical ingredients in the *biochemical* network.

Once the minimal realization has been found, it should be straightforward to expand it into a detailed, 'bottom-up' model. First, to identify the key property, the experimental results should be evaluated from a general, abstract perspective. As explained above, the general pattern of the experimental data is that the strong input triggers a response with a high maximum but transient signal down to ERK (and thus with a low long-term value), while the weaker input leads to a not-so-high maximum but a sustained signal (with a higher long-term value). Supposing that the latest value available resembles the steady-state value, this could be interpreted in system-theoretical terms as that the characteristic curve is non-monotone (see Figure 7.4).

Importantly, a body of experimental evidence supports the view that sustained ERK activation leads to a different physiological outcome than a transient activation of ERK. First observations on this concern can be found in a seminal paper by Marshall[169]. Later on, the molecular mechanisms responsible for this fact have been uncovered[181]: the level of activation of ERK serves as an input signal which is *integrated* over the time at the level of the

---

[*]The methods include: genetic algorithms implemented in Diana, an evolutionary algorithm[215] based on the ideas of Rechenberg[211], implemented both in DIVA and Matlab[249], simulated annealing and several local optimizers implemented in the Matlab Systems Biology Toolbox[232].
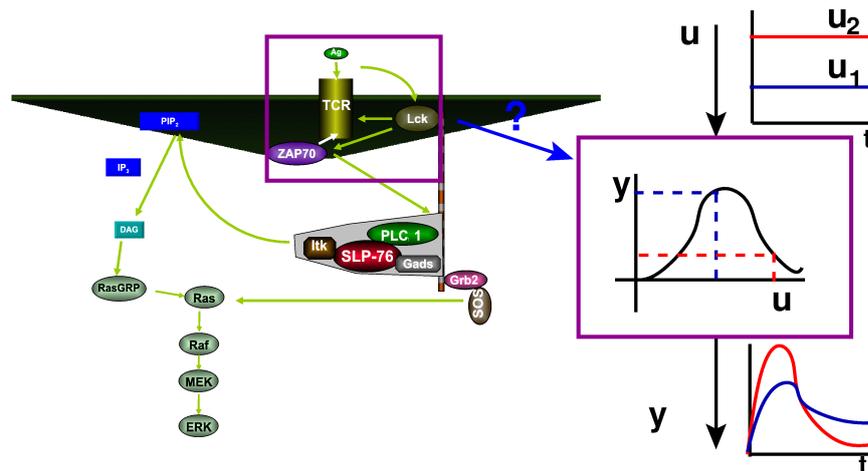
Figure 7.4: System-theoretical interpretation of the biological data and hypothesis. Somewhere in the early events close to the binding of ligand to the TCR, strong inputs are converted into strong but transient signals with a low asymptotic value, while moderate inputs are converted into mild and sustained outputs, with a higher asymptotic value. The characteristic curve of such a system is therefore non-monotone.

early gene expression: phosphorylated (active) ERK controls the activation of the transcription factor c-fos, and only a long-enough ERK signal leads to the formation of active and stable c-fos. Furthermore, in the context of T-cell signaling, it has been shown that sustained ERK activation upon moderate inputs leads to survival of lymphocytes in the course of T-cell development, while a transient ERK signal associated to strong inputs is correlated with strong apoptosis and absence of survival[51;271]. These results are very similar to those with naive T-cells presented in Section 7.1.2.1. Therefore, it is reasonable from a biological point of view to propose the steady-state value of ERK activation as a key property of this system and what our data suggests:

- strong input →transient ERK →death

- weak input →sustained ERK →survival.

An additional observation is that this pattern is present throughout the cascade, since it can be already observed for LAT phosphorylation. Therefore, the key mechanism responsible is upstream of LAT, involving early events of T-cell signaling.

Once the key behavior has been characterized, one should try to identify a system, as small and simple in mathematical terms as possible, which can explain it. The next section is devoted to this goal.

7.1.2.3 Determination of the minimal system able to reproduce the data

How should such a system look like? A first intuitive idea would be that a negative feedback is required, as the system is able to convert a sustained signal (addition of a ligand) into a

transient output (Figure 7.4). Arguably, the minimal system one could think about would consist of 2 states operating in closed loop. Such a system, depicted in Figure 7.5(a), is of

(a) Simplest model

(b) Biochemical implementation. The names of the functions as defined in Equations 7.9 and 7.10 are shown

(c) Block diagram of the simplest model
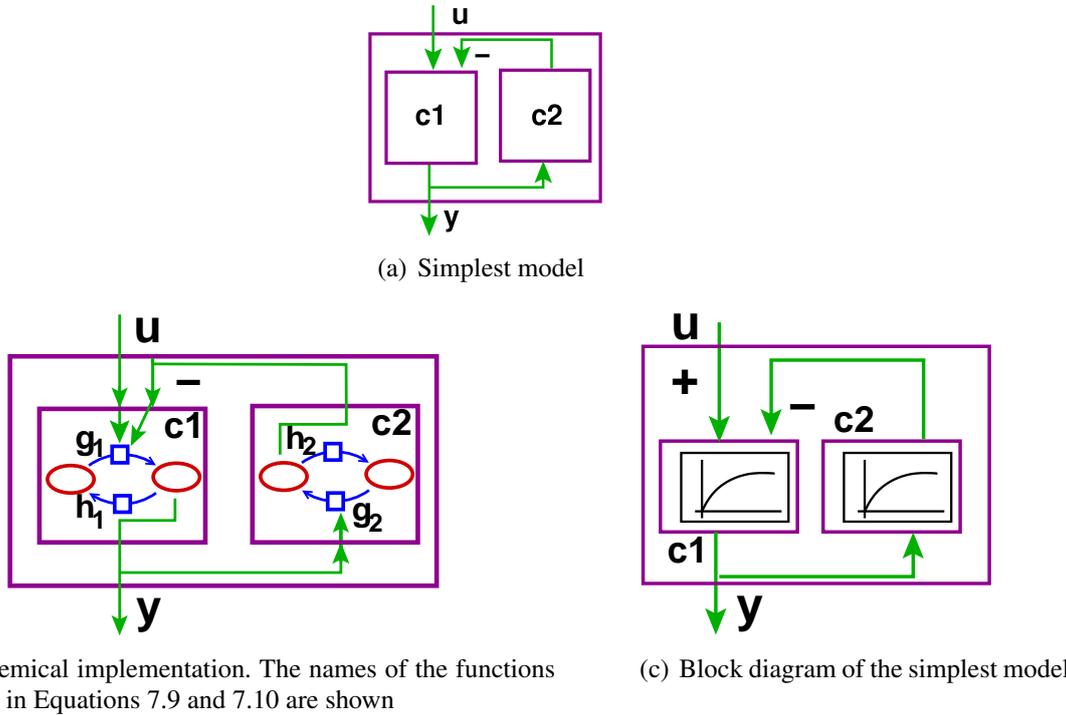
Figure 7.5: Simplest feedback model. The structure depicted in Figure 7.5(a) is arguably the simplest able to describe the behavior shown in Figure 7.4, and the scheme in Figure 7.5(b) shows a plausible biochemical implementation. If not marked otherwise, all arrows denote a positive effect. Figure 7.5(c) shows a block diagram of this system.

the general form of Equation 3.2

$$\dot{\vec{c}} = \frac{d\vec{c}}{dt} = f(\vec{c}, \vec{u}, \vec{p}). \tag{7.1}$$

For the sake of compactness, we shall remove the parameters $\vec{p}$ from the notation, since they are not relevant for the analysis hereafter. Additionally, and to avoid unnecessary notation, we shall refer to both the state and module with c_i. Accordingly, the system depicted in Figure 7.5(a) can be specifically defined in mathematical terms by the equations

$$\dot{c}_1 = f_1(c_1, c_2, u) \tag{7.2}$$

and

$$\dot{c}_2 = f_2(c_1, c_2), \tag{7.3}$$

where u is the input to the system, and the output $y = c_1$.

As explained in Section 5.2, monotone systems are 'simple' in a mathematical sense. Since we are trying to find the simplest system to explain the behavior, let us assume that c_1 and c_2

are both monotone or, more strictly, strongly input/output (I/O) monotone. Since both systems only comprise one state, it is enough that J_{input}, the partial derivatives of f with respect to the inputs, have definite sign to guarantee strong I/O monotony[7] (see Section 5.2.1.1). Furthermore, we would like to have a positive effect of c_1 on c_2, a negative effect of c_2 on c_1 (to produce a negative feedback) and a positive effect of the input u on c_1. Therefore, the partial derivatives must fulfill that

$$\partial f_1/\partial c_2 < 0 \; \forall c_2, \tag{7.4}$$

$$\partial f_2/\partial c_1 > 0 \; \forall c_1, \text{ and} \tag{7.5}$$

$$\partial f_1/\partial u > 0 \; \forall u. \tag{7.6}$$

In addition, we would like to have biologically realistic systems where the states (concentrations) have a limited value. To this end, the growth of each state is limited by its own value,

$$\partial f_1/\partial c_1 < 0 \text{ and} \tag{7.7}$$

$$\partial f_2/\partial c_2 < 0. \tag{7.8}$$

To be able to perform simulations, but without losing generality, a concrete form has to be assigned to f_1 and f_2. We shall try to define a 'biologically inspired' implementation, as we are interested in finding afterwards an analogous realization in a real biological system.

Recall from Section 5.2.2 that most signaling motifs are monotone, and that all have a monotone characteristic curve. Thus, we could consider c_1 and c_2 equivalent to any of the motifs described there. As we are seeking the simplest, we shall consider the motif C2s (see Figure 4.3(a) here). Here, a Hill function (Equation 5.26) shall be used for the kinetics, as a generalization of the Michaelis-Menten used in Sections 5.2 and 5.3[†].

In addition, one also has to model in biochemical terms the connections between c_1 and c_2. The connection $c_1 \rightarrow c_2$ (with $\partial f_2/\partial c_1 > 0 \; \forall c_1$) could be simply realized by assuming c_1 being the enzyme catalyzing the activation of c_2. Besides, two simple alternatives could be envisioned for the connection $c_2 \rightarrow c_1$ (with $\partial f_1/\partial c_2 < 0 \; \forall c_2$): a multiplicative, deactivating effect counteracting the activating input (which would resemble an inhibitor), and an additive effect diminishing the presence of c_1, mimicking the effect of a degradation. Note that with the interconnections defined as above, a coupling free of retroactivity (see Section 3.2) is guaranteed.

According to the premises described above, the equations for a biochemical realization for

---

[†]It should be noted that the results hereafter hold qualitatively also for Michaelis-Menten kinetics; the use of Hill kinetics has as only reason more degrees of freedom to obtain a better fit to the time courses.

the system defined by Equations 7.2 and 7.3 read

$$\dot{c}_1 = u g_1(c_1, c_2) - h_1(c_1) = u \frac{(c_1^0 - c_1 - k_{ac} \cdot c_2)^{h1}}{Km_1^{h1} + (c_1^0 - c_1)^{h1}} - \frac{kr_1 c_1^{h1r}}{Kmr_1^{h1r} + c_1^{h1r}}, \tag{7.9}$$

$$\dot{c}_2 = g_2(c_1, c_2) \quad - h_2(c_2) = c_1 \frac{k_2(c_2^0 - c_2)^{h2}}{Km_2^{h2} + (c_2^0 - c_2)^{h2}} - \frac{kr_2 c_2^{h2r}}{Kmr_2^{h2r} + c_2^{h2r}}, \tag{7.10}$$

with[‡]

$$\partial g_1/\partial c_1 < 0 \; \forall c_1, \; \partial h_1/\partial c_1 > 0 \; \forall c_1, \quad \partial g_1/\partial c_2 < 0 \; \forall c_2, \tag{7.11}$$

$$\partial g_2/\partial c_2 < 0 \; \forall c_2, \; \partial g_2/\partial c_1 > 0 \; \forall c_1 \text{ and } \partial h_2/\partial c_2 > 0 \; \forall c_2, \tag{7.12}$$

including the aforementioned additive mechanism for the connection $c_2 \to c_1$, whose strength is controlled by the parameter k_{ac} [§].

This simple model defined with Equations 7.9 and 7.10 was trained with the experimental data. Recall from Section 7.1.2.2 that this small system describes the early events. Therefore, the output $c_1 = y$ was mapped to the phosphorylation of LAT (Figure 7.3), an event immediately downstream of the early events (see Figure 7.4).

The model was able to successfully fit the experimental data for LAT phosphorylation. Additionally, the state c_2 (*not* fitted to any data) behaved qualitatively similar to the degradation of ZAP70 (see Figure 7.3), supporting thus the hypothesis that ZAP70 is involved in the feedback responsible for the behavior of LAT. Furthermore, when the model was trained with both the LAT phosphorylation and ZAP70 data (for c_1 and c_2, respectively), the fit was remarkably good.

As mentioned above, the characteristic pattern spreads downstream of LAT to PLCγ1 and ERK (see Figure 7.3). It therefore seems that, as far as from this set of data can be extracted, there are no additional complex mechanisms between LAT and ERK[¶]. Therefore, the model depicted in Figure 7.5(b) was extended with the minimal elements to fit the data for PLCγ1 and ERK: two additional modules, c_3 and c_4 (corresponding the C2s motifs modeled as the ones above, with Hill kinetics) added in series to the output of c_1 (see Figure 7.6(a)). The resulting model provided a good fit to the data mapping $c_1 \leftrightarrow LATp$, $c_2 \leftrightarrow ZAP_{deg}$, $c_3 \leftrightarrow PLC\gamma - 1p$, and $c_4 \leftrightarrow ERKpp$ (see Figure 7.6(b) and Section A.7.1 for details). The parameters varied in a range of $\approx 10^{-1} - 10^4$. The range is narrow enough to consider the

---

[‡] strictly, $\partial g_1/\partial c_1 < 0$ holds only for $c_1^0 - c_1 - k_{ac} \cdot c_2 > 0$ or for $h1 = 1, 3, 5, ...$ We shall consider here thus $h1 = 1$.

[§] If one would like to include the multiplicative feedback, one would just replace u with $\frac{u}{(1 + k_{mc} \cdot c_2)}$. In the following, the additive feedback will be used, but the results hold also for the multiplicative.

[¶] A detailed inspection reveals that there is a subtle difference at the level of ERK, where the signal upon streptamers stimulation increases, decreases, and then increases again. This suggests the presence of oscillations, sustained or dumped. As no further experimental data is available, and thus no clear statement could be made, this point was not further pursued.

(a) Extended model   (b) Fit to the data
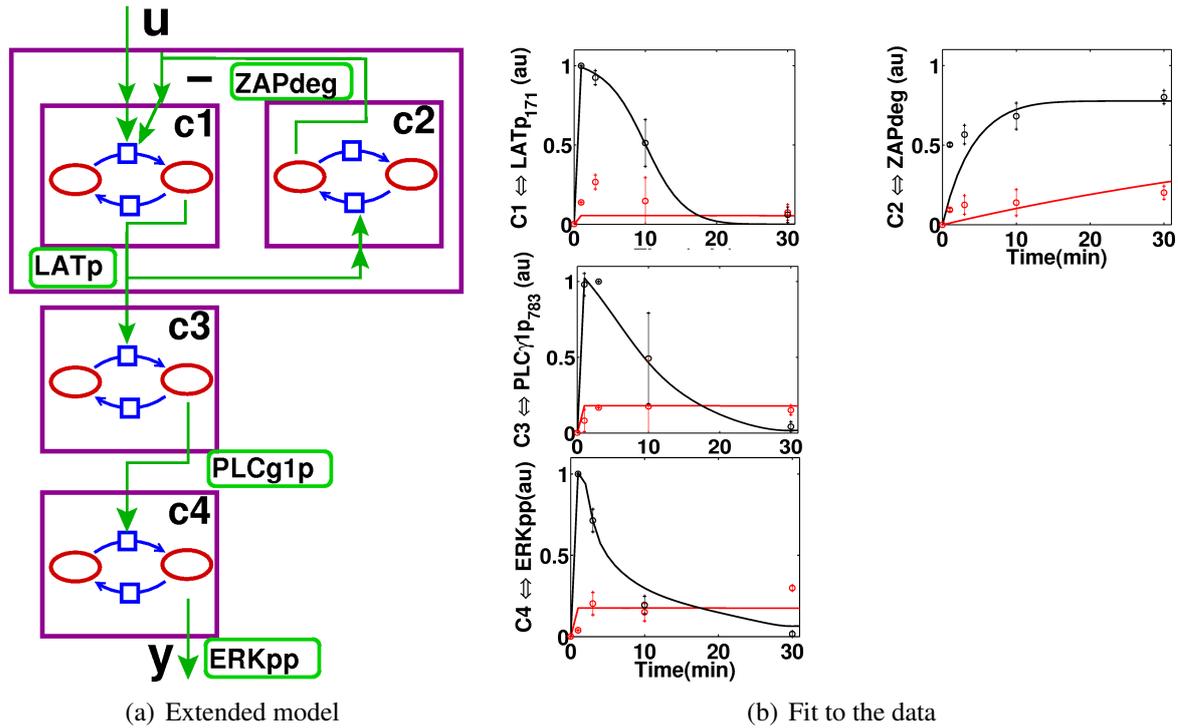
Figure 7.6: Extension of the simplest feedback model to consider PLCγ1 and ERK and its fit of the data. The scheme in Figure 7.6(a) shows a plausible biochemical implementation of the extended model. If not marked otherwise, all arrows denote a positive effect. Figure 7.6(b) shows the result of the fitting to the experimental data of Figure 7.3 (red: streptamers stimulation; black: antibody stimulation). The dots with bars correspond to the experimental values (mean $\pm$ standard error of the mean), and the continuous lines to the simulation results.

parameters as plausible; too different values would be impossible *de facto* due to physical limitations such as diffusion coefficients, etc. It is important to stress that better fits than those of Figure 7.6(b) could be found, at the expense of introducing additional parameters into Equations 7.9 and 7.10. As the subject of this analysis is the *qualitative* behavior rather than the *quantitative* fit, the results depicted in Figure 7.6(b) are used as a balance between good fit and simplicity of the model.

Additionally, including a supplementary block with two modules c_5 and c_6 connected in the same way as c_1 and c_2, the data of PKB could be fitted.

In summary, a simple model (Figure 7.6(a)) has been found which can reproduce the experimental data. This means that the degradation of ZAP70 *could* be the mechanism responsible for the observed behavior, but it is, however, not a proof for it, as we can not rule out other mechanisms to be involved. Actually, as it will be seen in the next section, a deeper analysis reveals that the behavior of the system is not the desired one.

7.1.2.4 Analysis of the steady state behavior of the simple model

In order to gain more insight into this control system, which may help to decide whether it is plausible that this mechanism is behind the behavior, its system-theoretical properties

will be analyzed, in particular its steady-state behavior. The plots of Figure 7.6(b) *suggest* that the steady-state behavior is as desired: higher value for moderate inputs than for strong inputs (recall Figure 7.4). However, if one runs the simulations plotted in Figure 7.6(b) for a longer time, the curves eventually cross again at a certain time (Figure 7.7). Accordingly, computation of the characteristic curve reveals that it has a monotone shape (Figure 7.8).
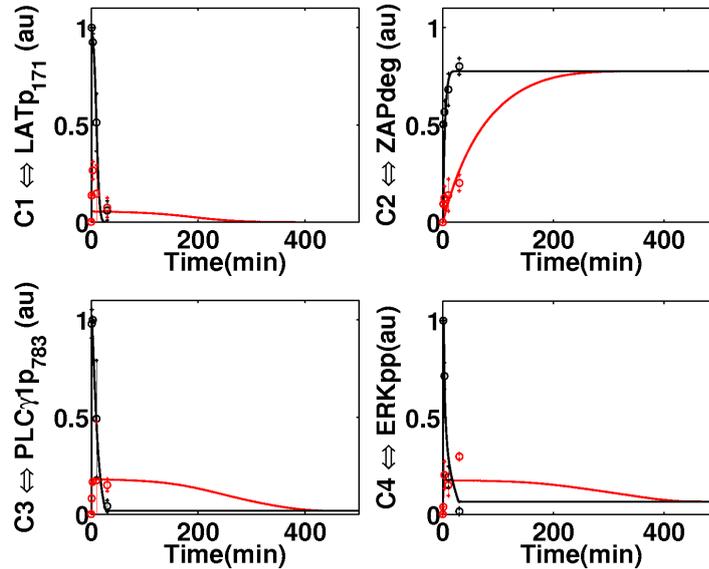
Figure 7.7: Simulation of the model trained to the data for a longer time. Extending the simulation (see Figure 7.6(b)) until 200 minutes reveals that the system tends to an asymptotic state which is *not* higher for the stronger input.
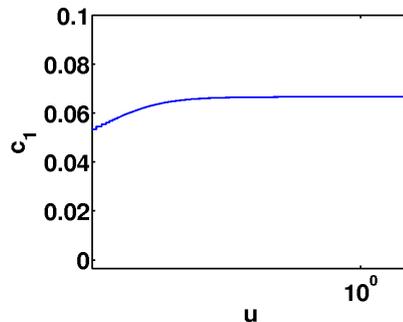
Figure 7.8: Steady-state characteristic curve for the simplest feedback model. The parameter values correspond to those obtained from the fit of the data (see Figure 7.6(b) and Section A.7.1)The characteristic curve has a monotone form.

It may be that this is the case for this particularly set of parameters, or an inherent property of this structure. To clarify this point, a deeper mathematical analysis will be performed. We shall consider solely c_1 and c_2 (Figure 7.5(a)), as this is the part where the interesting dynamic behavior takes place. Recall from Section 5.2.1.1 that monotone systems have a monotone steady-state characteristic curve. Therefore, if the system defined by Equations 7.9 and 7.10 is monotone, its characteristic curve will be monotone, which could then *not* have the desired form (Figure 7.4).

If one computes the Jacobian, one gets

$$J = \frac{\partial f}{\partial \vec{c}} = \begin{pmatrix} \frac{\partial f_1}{\partial c_1} & \frac{\partial f_1}{\partial c_2} \\ \frac{\partial f_2}{\partial c_1} & \frac{\partial f_2}{\partial c_2} \end{pmatrix} = \begin{pmatrix} u\frac{\partial g_1}{\partial c_1} - \frac{\partial h_1}{\partial c_1} & u\frac{\partial g_1}{\partial c_2} - \frac{\partial h_1}{\partial c_2} \\ \frac{\partial g_2}{\partial c_1} - \frac{\partial h_2}{\partial c_1} & \frac{\partial g_2}{\partial c_2} - \frac{\partial h_2}{\partial c_2} \end{pmatrix} \tag{7.13}$$

which, considering the conditions in Equations 7.11 and 7.12, leads to

$$J = \begin{pmatrix} (+)(-) - (+) & (+)(-) - 0 \\ (+) - 0 & (-) - (+) \end{pmatrix} = \begin{pmatrix} - & - \\ + & - \end{pmatrix} \tag{7.14}$$

and

$$J_{input} = \frac{\partial f}{\partial \vec{u}} = \begin{pmatrix} + \\ 0 \end{pmatrix}. \tag{7.15}$$

Therefore, with the method presented in Section 5.2, one can not guarantee that the system as a whole is monotone, because it is sign-definite but there is one negative feedback in the incidence matrix associated to the Jacobian[242].

However, the steady-state characteristic curve $c_1^{ss}(u)$ is monotone. The demonstration is similar to the one employed for the motifs C3dp, C2C2 and C3C2 (see Sections 5.2.2 and A.3): since c depends on u, from Equation 7.1 in the steady state it holds

$$0 = f(\vec{c}_{ss}, u). \tag{7.16}$$

Deriving with respect to u and applying the chain rule, since $\vec{c}_{ss}$ depends on u, we obtain

$$0 = \frac{\partial f(\vec{c}^{ss}(u), u)}{\partial c} \frac{\partial c}{\partial u} + \frac{\partial f(\vec{c}^{ss}(u), u)}{\partial u} = J \cdot c_u + f_u, \tag{7.17}$$

from where it leads

$$c_u = -J^{-1} f_u \tag{7.18}$$

with $J = \frac{\partial f}{\partial c}$, $c_u = \frac{\partial c}{\partial u}$, and $f_u = \frac{\partial f}{\partial u}$. Since

$$J^{-1} = \frac{-1}{J_{22}J_{12} - J_{21}J_{11}} \begin{pmatrix} J_{22} & -J_{12} \\ -J_{21} & J_{11} \end{pmatrix} = \frac{-1}{(-)\cdot(-) - (+)\cdot(-)} \begin{pmatrix} - & -(-) \\ -(+) & - \end{pmatrix} = \begin{pmatrix} - & + \\ - & - \end{pmatrix},$$

it results from Equation 7.18 that

$$c_u = \begin{pmatrix} \frac{\partial c_1^{ss}}{\partial u} \\ \frac{\partial c_2^{ss}}{\partial u} \end{pmatrix} = -J^{-1} f_u = -\begin{pmatrix} - & + \\ - & - \end{pmatrix} \begin{pmatrix} + \\ 0 \end{pmatrix} = \begin{pmatrix} + \\ + \end{pmatrix}. \tag{7.19}$$

That is to say, both characteristic curves $c_1^{ss}(u)$ and $c_2^{ss}(u)$ are monotone. This result is unexpected when compared to the fitting of data (Figure 7.6(b)): it seemed that the steady state

value for the strong input would be *lower* than for the weak input. However, the rigorous result of Equation 7.19 states that, without question, the characteristic curve is monotone for any parameter values.

As we are looking for a system where the steady state value is higher for weak inputs than for strong ones, the system defined by Equations 7.9 and 7.10, even though able to fit the experimental data (available for a short range of time), is not a suitable candidate. In the next section an extension will be presented to fulfill the desired criterion.

7.1.2.5 Extension of the model to provide reasonable steady-state behavior

The analysis of the previous section demonstrates that two monotone systems connected as in Figure 7.5(a) must have a monotone characteristic curve. If we want to keep the structure simple, a reasonable way to circumvent this problem is to simply make the element J_{22} in the Jacobian non-definite, since from Equation 7.19

$$\frac{\partial c_1^{ss}}{\partial u} = J_{11}^{-1}(+) + J_{12}^{-1}0 = \frac{J_{22}}{J_{22}J_{12} - J_{21}J_{11}}(+). \tag{7.20}$$
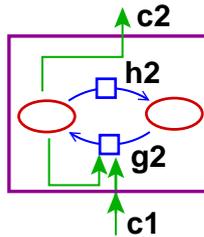
(a) Autocatalytic module
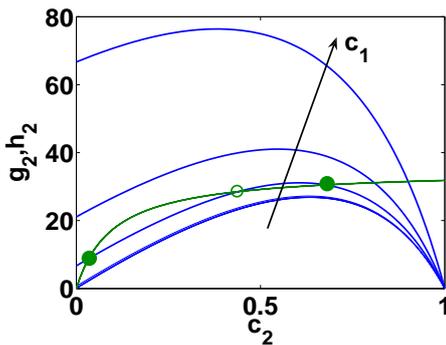
(b) Forward (g_2, blue) and backward (h_2, green) reactions (Equations 7.9, 7.10, and 7.21) as a function of c_2. Full points denote stable steady states, and the empty circle the unstable one.
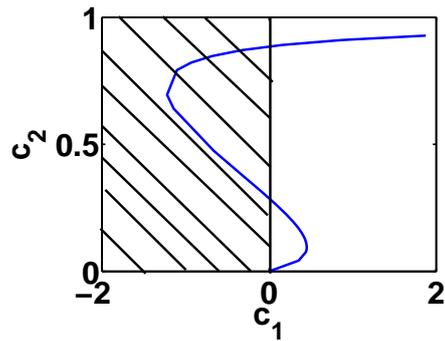
(c) Characteristic curve of the module c_2

Figure 7.9: Emergence of bistability via an autocatalytic step. For a range of values of the input c_1, the forward reaction $g_2(c_2)$ crosses the backward reaction $h_2(c_2)$ at three points, 2 stable and one unstable, leading to bistability (Figure 7.9(b)). Since negative values for c_1 are physically not possible, when the input c_1 is large enough, creating an irreversible switch (Figure 7.9(c)).

For example, since $\partial g_2/\partial c_2 < 0$, by somehow adding a term which leads to a positive effect

of c_2 on g_2, this goal could be achieved. In biochemical terms, an element has a positive effect on itself if there is a sort of autocatalytic regulation. One can e.g. extend g_2 (see Equation 7.10) to

$$g_2(c_1, c_2) = \frac{(c_1 + k_{af} c_2^n) k_2 (c_2^0 - c_2)^{h2}}{Km_2^{h2} + (c_2^0 - c_2)^{h2}}. \tag{7.21}$$

Now, the sign of the element of the Jacobian $J_{22} = \partial f_2 / \partial c_2$ is not determined. Therefore, and operating as above with Equation 7.18, it results that the sign of $\frac{\partial c_1^{ss}}{\partial u}$ is *not* definite, i.e., the characteristic curve is not necessarily monotone. Furthermore, c_2 is converted from a C2s into a C2p motif (see Figure 7.10(a) and Section 4).

This motif can show bistability[||] and, thus, its characteristic curve can show a sharp, switch-like form (see Figure 7.9). Hence, it is a promising extension as it could bring this switch-like form in the characteristic curve of the whole system, while keeping the desired fit to the data of the simplest model. In fact, the form of such a system can have a non-monotone characteristic curve with an abrupt decay (see Figure 7.10(c)).
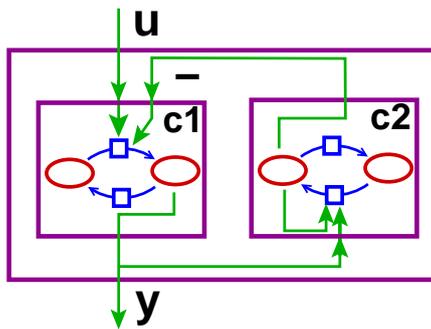
(a) Biochemical implementation of the extended model including an autocatalytic step
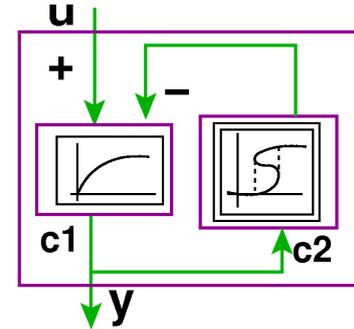
(b) Block diagram of the extended model including an autocatalytic step
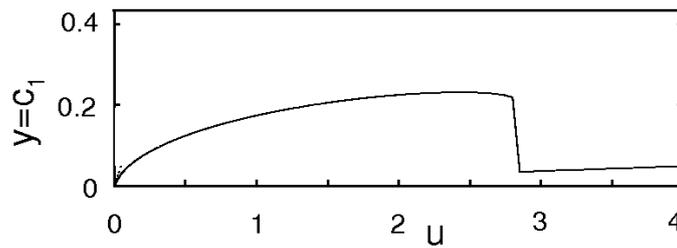
(c) Steady-state characteristic curve for $y = c_1$

Figure 7.10: Model including an autocatalytic step. The simplest model of Figure 7.5 is extended with an autocatalytic step (Equation 7.21) to obtain the switch-like behavior (Figure 7.10(c)).

However, a further analysis reveals that, after a certain value of the input u, c_1^{ss} grows again monotonely with the input u:

$$\exists u_i : \quad \frac{\partial c_1^{ss}}{\partial u} > 0, \forall u > u_i \tag{7.22}$$

[||]This result was already demonstrated in Section 5.1.

First, consider the nullclines[**] of the system:

$$\dot{c_1}^{ss} = 0 \rightarrow ug_1(c_1^{ss}, c_2^{ss}) = h_1(c_1^{ss}) \Leftrightarrow u\frac{(c_1^0 - c_1^{ss} - k_{ac} \cdot c_2^{ss})^{h1}}{Km_1^{h1} + (c_1^0 - c_1^{ss})^{h1}} = \frac{kr_1 c_1^{ss^{h1r}}}{Kmr_1^{h1r} + c_1^{ss^{h1r}}} \quad (7.23)$$

$$\dot{c_2}^{ss} = 0 \rightarrow g_2(c_2^{ss}, c_1^{ss}) = h_2(c_2^{ss}) \Leftrightarrow \frac{k_2(c_1^{ss} + k_{af})(c_2^0 - c_2^{ss})^{h2}}{Km_2^{h2} + (c_2^0 - c_2^{ss})^{h2}} = \frac{kr_2 c_2^{ss^{h2r}}}{Kmr_2^{h2r} + c_2^{ss^{h2r}}}. \quad (7.24)$$

Both nullclines together are sufficient to characterize the steady state values and it can be seen that the input u is involved in the nullclines. Therefore, the steady state values of both c_1^{ss} and c_2^{ss} depend on the input. Furthermore, studying the form of the nullclines, it can be shown that the values of c_1^{ss} and c_2^{ss} grow with the input u (see Figure 7.11). Therefore, the characteristic curve grows steadily with the input and, for strong inputs, the steady state value can be similar to the one of the 'optimal', moderate inputs.
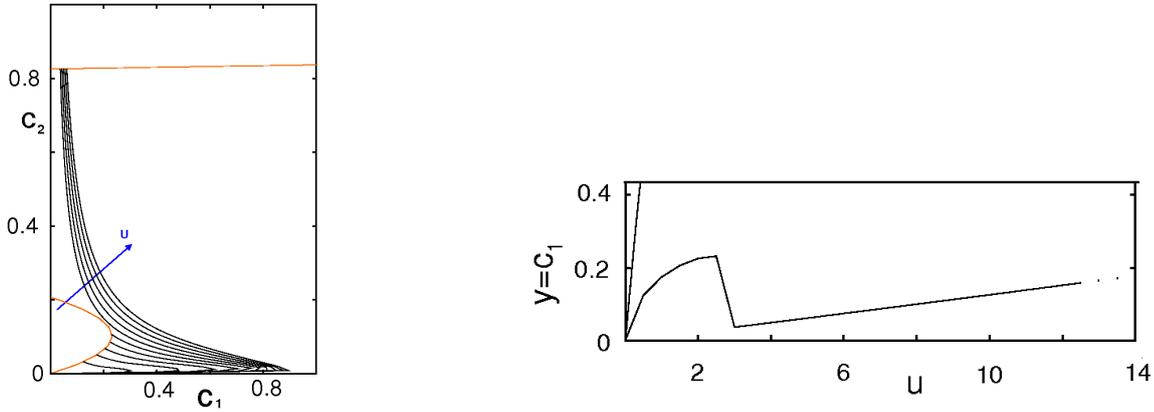
(a) Nullcline for c_2 (red, Equation 7.28), and trajectories for different values of u

(b) Steady-state characteristic curve for high inputs ($y = c_1$)

Figure 7.11: Nullclines and characteristic curve for high inputs for the model including an autocatalytic step. In Figure 7.11(a), the red line represents the null-cline for c_2^{ss}, and the black lines different trajectories corresponding to different values of u. It results from Equation 7.23 that, the higher the input, the higher is the nullcline for c_1^{ss}, and thus, at a higher point meets the c_2^{ss} nullcline. Therefore, for strong input values, the steady state values become inexorably high.

This result means that, even though inputs moderately higher than the agonist (optimal stimulus) could be shut down, extremely strong inputs would not (see Figure 7.11(b)). This is not reasonable from a biological point of view: if e.g. upon a certain mutation a peptide is generated which has an extraordinary affinity for the TCR, it would lead to incorrect T-cell activation. Additionally, we have seen in Section 7.1.2.1 that antibodies (which have a huge affinity for the TCR) do not lead to high, but rather very low, long-term ERK activation. Therefore, the extension of the model with the positive feedback, even if it takes us closer

---

[**]Nullclines are curves along which the vector field is either completely horizontal or vertical, i.e., where $\dot{c_i} = 0$[273]. The intersections between nullclines correspond to points of equilibrium.

to the goal, is not the answer. We are looking for a system so that the output reaches a high value if the input is below a certain threshold u_t but, for $u > u_t$ it should remain low and independent of the input value.

It is well-known in control theory that a hallmark of an integral feedback system is that the output in steady-state is independent of the input and always zero. The key property of an integrator is its ability to adapt to the strength of the signal and thus to switch off signals of different strength. Interestingly, integral feedback can be found in biological control systems[121]. It seems thus reasonable to consider an integral element as an extension of our system. Importantly, we shall consider a *real* instead of ideal integrator (see Section 5.3.1.1) for 2 reasons: (i) purely *ideal* integrators are not realizable and (ii) an ideal integrator would lead to an output signal which is *always* (including for low inputs) zero and would 'destroy' the bistable system generated by the autocatalytic step.
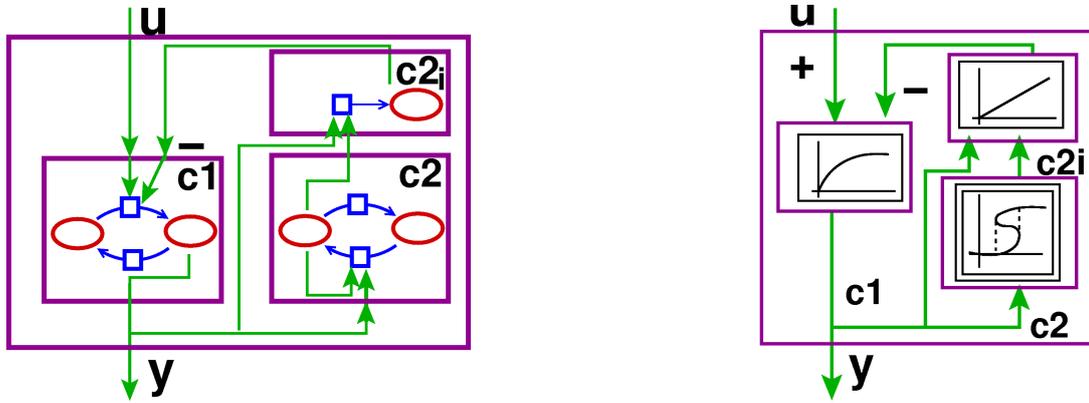
(a) Biochemical implementation of the extended model including a real integrator

(b) Block diagram of the extended model including a real integrator

Figure 7.12: Model including a real integrator. The simplest model of Figure 7.5 is extended with an autocatalytic step (Equation 7.21) and an integrator (Equation 7.25) to obtain the switch-like behavior (Figure 7.13(b)).

We shall consider a biologically inspired integrator of the form

$$\dot{c_{2i}} = f_{2i}(c_1, c_2) = (D_c + k_{2i} \cdot c_2)\frac{c_1}{Km_d + c_1} - F \qquad (7.25)$$

and subsequently rewrite g_1 (see Equation 7.9) to

$$g_1(c_1, c_{2i}) = u \frac{(c_1^0 - c_1 - k_{ac} \cdot c_{2i})^{h1}}{Km_1^{h1} + (c_1^0 - c_1)^{h1}}. \qquad (7.26)$$

Note that for $D_c = 0$ and $F = 0$, when $c_1 \gg Km_d$ (saturation), Equation 7.25 converts into $\dot{c_{2i}} = k_{2i} \cdot c_2$, which is the equation of an ideal integrator.

The form of Equation 7.25 can be motivated as follows: it describes a real, biological integrator, where c_{2i} is a degradation process which depends on the amount of protein c_1 in a
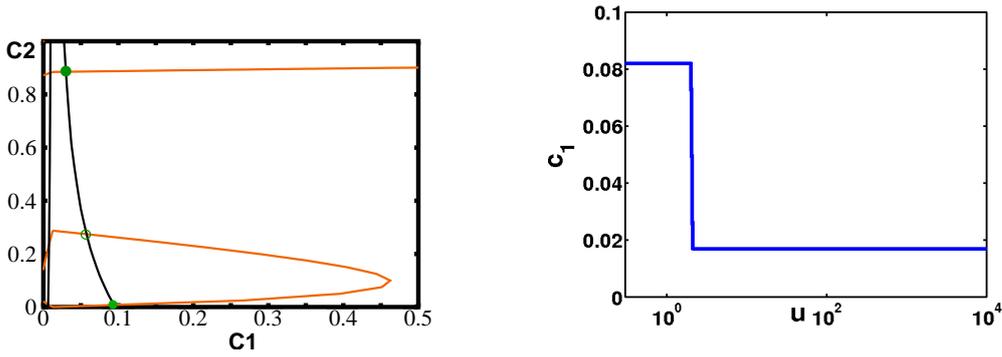
(a) Nullclines for c_2 (red, Equation 7.28) and c_2i (blue, Equation 7.29). Full points are the stable steady states, and the empty circle the unstable one

(b) Steady-state characteristic curve for $y = c_1$

Figure 7.13: Nullclines and characteristic curve for the model including a real integrator. The parameter values correspond to those obtained from the fit of the data (see Figure 7.14(b) and Section A.7.2). Note that the nullcline for c_2 in Figure 7.13(a) represents at the same time the characteristic curve of $c_2(c_1)$.

Michaelis-Menten manner, as well as on the amount of 'enzyme' c_2 catalyzing the degradation. The terms D_c and F represent the constitutive degradation and formation, respectively. In steady-state, from Equations 7.2, 7.3, and 7.25, the nullclines read

$$\dot{c}_1^{ss} = 0 \rightarrow u g_1(c_1^{ss}, c_{2i}^{ss}) = h_1(c_1^{ss}) \Leftrightarrow u \frac{(c_1^0 - c_1^{ss} - k_{ac} \cdot c_{2i}^{ss})^{h1}}{Km_1^{h1} + (c_1^0 - c_1^{ss})^{h1}} = \frac{kr_1 c_1^{ss^{h1r}}}{Kmr_1^{h1r} + c_1^{ss^{h1r}}} \quad (7.27)$$

$$\dot{c}_2^{ss} = 0 \rightarrow g_2(c_1^{ss}, c_2^{ss}) = h_2(c_2^{ss}) \Leftrightarrow \frac{k_2(c_1^{ss} + k_{af} c_2^n)(c_2^0 - c_2^{ss})^{h2}}{Km_2^{h2} + (c_2^0 - c_2^{ss})^{h2}} = \frac{kr_2 c_2^{ss^{h2r}}}{Kmr_2^{h2r} + c_2^{ss^{h2r}}} \quad (7.28)$$

$$\dot{c}_{2i}^{ss} = 0 \rightarrow f_{2i}(c_2^{ss}, c_1^{ss}) = 0 \qquad \Leftrightarrow (D_c + k_{2i} \cdot c_2^{ss}) \frac{c_1^{ss}}{Km_d + c_1^{cc}} = F. \quad (7.29)$$

Equations 7.28 and 7.29 define a system of 2 equations and 2 variables (c_1^{ss} and c_2^{ss}). Therefore, the concentration in steady-state of both c_1 and c_2 is fixed, i.e. only dependent on the kinetic parameters but not on the input (see Figure 7.13(a)). Equation 7.27, in turn, defines c_{2i}^{ss} as a function of c_1^{ss}, c_2^{ss} and the input u. As expected, the characteristic curve of such a system has a clean switch-like form (see Figure 7.13(b)). Thus, a design was found that possesses all the properties required to be an effective 'fuse' of T-cell activation.

The model with the real integrator was extended with two additional modules (analogously as done in Section 7.1.2.3) to train it with the experimental data (see Figure 7.14(a)). First, parameters for the module c_2 where found which lead to an irreversible switch (see Figure 7.13(a)). Subsequently, the fit to the experimental data was used to determine the rest of parameters (Section A.7.2). Remarkably, the extended model is able to reconcile the fit of the data (Figure 7.14(b)) with the desired asymptotic behavior (Figure 7.13(b)).

(a) Extended model   (b) Fit to the data
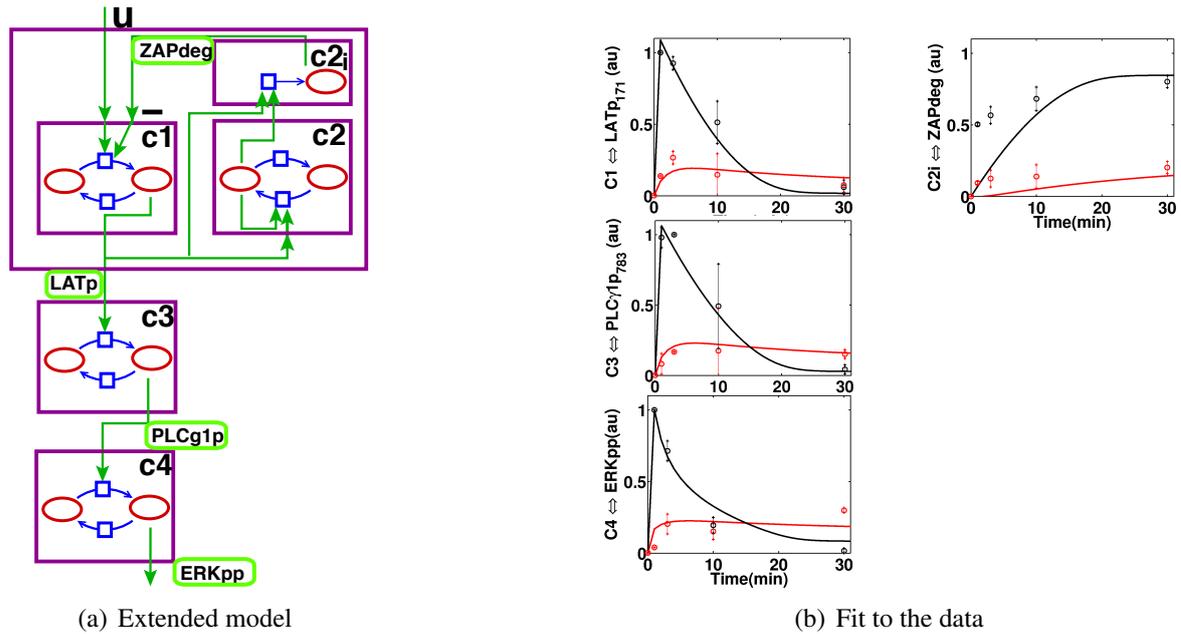
Figure 7.14: Extension of the model including the integrator to consider PLCγ1 and ERK and its fit of the data. The scheme in Figure 7.14(a) shows a plausible biochemical implementation of the extended model. If not marked otherwise, all arrows denote a positive effect. Figure 7.14(b) shows the result of the fitting to the experimental data of Figure 7.3 (red: streptamers stimulation; black: antibody stimulation). The dots with bars correspond to the experimental values (mean $\pm$ standard error of the mean), and the continuous lines to the simulation results.

7.1.2.6 Identification in the signaling network of the mathematical ingredients

In the previous section, a system able to produce the desired behavior was characterized. In summary, 3 mathematical elements are required:

- A negative feedback

- an autocatalytic step, to break the monotony and also to create the bistable system essential for the switch and

- an integrator, to repress strong inputs and make the output independent of the input.

Interestingly, these ingredients are present in the molecule ZAP70 (see Figure 7.15)[118;204]: Upon Lck[††]-mediated phosphorylation at the residue p_{493}, ZAP70 becomes activated. Recall from Section 2.2.2 that Lck is directly downstream of the TCR and the first element activated in the TCR-dependent signaling cascade (see Figure 7.4). Activated ZAP70 autophosphorylates at positions p_{292}, p_{315}, and p_{319} in a trans manner[162]. cCbl can bind at the phosphorylated position p_{292}, leading to an ubiquitination and subsequent degradation of ZAP70. Recall that the experimental data shows that the degradation seems to be involved in the shut-off of the too strong signal (see Figure 7.3), supporting thus the involvement

---

[††]and probably Fyn, see Section 2.2.2.
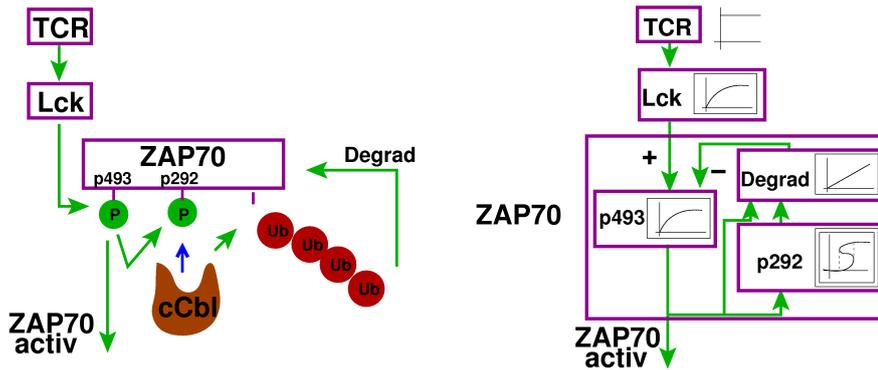
Figure 7.15: Identification in the protein ZAP70 of the required mathematical ingredients. The kinase ZAP70 has the 3 ingredients required to generate a system of the form of Figure 7.12(a): activation by phosphorylation, and feedback loop comprising an autophosphorylation and an integrator (degradation).

of ZAP70 and, furthermore the training of the model to the data also suggests that ZAP70 degradation could be the state in the feedback.

7.1.2.7 Interpretation of the results in the context of T-cell signaling

T-cell activation is a complex process, and so it has to be, as it is a critical decision which, if wrongly taken, risks the whole organism. Accordingly, it must be precisely regulated. Furthermore, T-cell activation does not only depend on the affinity of a particular ligand for the TCR, but also on other context-dependent factors, including signals from other immune cells. Therefore, and not surprisingly, a large array of factors have been proposed to play a role in T-cell activation. A particularly fascinating fact is that T-cells must be precisely controlled to respond only to an antigen characterized by a specific affinity. If one would plot the response (which, as discussed above in Section 7.1.2.2, can be roughly mapped to the ERK activation) vs. affinity, one would expect a pulse form characterized by two sharp switch-like transitions (see Figure 7.16).

The mechanism underlying T-cell specificity is still matter of intense debate. The mechanism firstly proposed is kinetic proofreading[173]: a series of biochemical processes must take place during binding of ligand to the TCR to lead to full activation, and the longer the chain of processes, the sharper (more ultrasensitive) is the activation switch. Therefore, only ligands with enough affinity to bind the TCR for a sufficient time (more precisely, enough lifetime of the complex pMHC:TCR) lead to full T-cell activation. A counteracting phenomenon is serial triggering[266]: since the density of pMHC on the surface of the APC is low, the lifetime of the pMHC:TCR must be short enough to allow a single peptide to engage different TCRs. Kinetic proofreading together with serial triggering leads to the existence of an optimum half-life of the complex pMHC:TCR, in agreement with experimental and modeling[49] studies.

However, a number of recent experimental results show an all-or-none response in T-

(a) Characteristic curve   (b) Molecular mechanisms involved
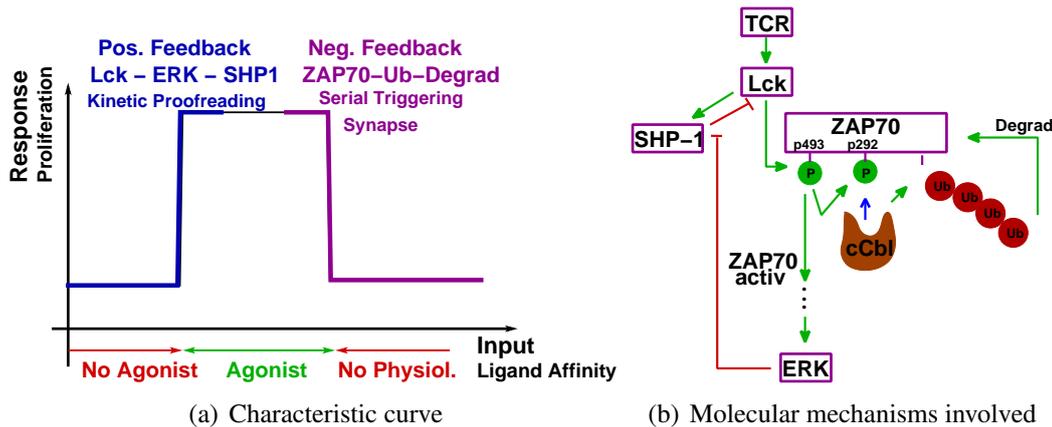
Figure 7.16: Combination of positive and negative feedback gives rise to a double switch mechanism guaranteeing the specificity of T-cell activation. On the one hand, the positive (to be more precise, double negative) feedback Lck →...→ERK ⊣ SHP-1 ⊣ Lck [4;245] gives rise to the first switch. On the other hand, the mechanism of autophosphorylation+degradation of ZAP70 (Figure 7.16(b)) could produce the second switch, avoiding activation upon unphysiologically strong inputs.

cell activation [3]. This, together with the fact that the kinetic-proofreading schemes can only provide adequate ligand discrimination at the expense of sensitivity or speed of response, in addition to the lack of robustness of kinetic-proofreading [32;34;203] pose serious doubts on the actual importance of classical kinetic-proofreading and serial triggering.

As an alternative, it has been proposed that feedback regulation in the signaling cascade downstream of TCR activation can provide this specificity and sensitivity [3;32]. Chan et al. [33] explored how a simple system based on the interplay of a positive and a negative feedback (implemented by an autocatalytic kinase and a phosphatase operation in closed loop) can lead to a switch-like response, assuring that too weak ligands do not activate T-cells. Previous, similar work can be found in [132].

More recently, Altan-Bonnet and Germain [3] developed a comprehensive model, based also on the interplay of a positive and a negative feedback. In their model, the key mechanism comprises Lck-mediated activation of SHP-1, which in turn inhibits Lck. Since Lck is directly downstream of the TCR (and above ZAP70), this would create a negative feedback at the level of the early events. To counteract this negative regulation ERK, downstream of Lck and ZAP70, can inhibit the effect of SHP-1 on Lck, creating a double negative feedback (Lck →...→ERK ⊣ SHP-1 ⊣ Lck) which has the net effect of a positive feedback (see Figure 7.16(b) and also Section 2.2.2). This model relies on well-founded biochemical evidence [245] and ad-hoc wet-lab experiments [3], and seems to be a good candidate for the first switch, i.e., the one responsible to avoid T-cells to get activated by too weak inputs (Figure 7.16).

Another complex phenomenon controlling T-cell activation is the immunological synapse, a highly structured junction between T-cell and APC surfaces. It has been shown, both ex-

perimentally and mathematically, that this structure works as an adaptative controller: it concentrates the players to boost the signaling process but, at the same time, attenuates strong signals by promoting receptor degradation[154]. The latter effect is very similar to the one described in this work, both regarding its goal (block of too strong signals) and the methods used (ubiquitination and degradation). However, in our experimental system there is no synapse, as we stimulate T-cells with streptamers and not APCs (see Figure 2.3). Therefore, the effect of the synapse may probably complement the one proposed here.

Without challenging the importance of this and other mechanisms, the analysis presented here tries to make the point that the negative feedback involving ZAP70 autophosphorylation plus degradation, is responsible, or at least involved, in the generation of the second switch to avoid activation by too-strong, unphysiological signals (see Figure 7.16). Also of biological relevance is the irreversibility of this mechanism: once the strong input signal triggers the switch (c_2), it remains active. This would block activation upon a posterior input. This property makes sense in biological terms: if a T-cell meets such an extremely high input, it means that something unnormal is happing either in the cell or in its environment. Therefore, as a safety measurement, should be permanently switched off.

In summary, even though the experimental data *per se* suggested the involvement of ZAP70 degradation in the switch off of strong signals, only together with a mathematical model was it possible to demonstrate that ZAP70 degradation can be responsible for this behavior (since a simple model including only this regulatory mechanism can fit the data). Furthermore, a detailed analysis of the simple model provides insights into the systems properties of this control mechanism which can be seen as an explanation of why it is an 'expensive' degradation process and not a simple inhibition via e.g. dephosphorylation what is included into the design of this biochemical fuse.

7.2 Modular analysis of the EGF-induced MAPK cascade

To illustrate the applicability of the modular approach and the definitions introduced in Section 5.3.1.2 to a large signaling network, these concepts will be applied to the model of the EGF-induced MAPK cascade of Schoeberl et al.[234] (see Section 2.2.1) , which was decomposed into modules in Section 3.4.4.

One hallmark of this signaling system is a remarkable independence on the concentration of the ligand[234]. The amount of EGF can vary over a wide range of biologically significant values without major effect on the output signal, the double phosphorylated form of ERK (Figures 7.17(a) and 7.17(b)). For any value higher than approximately 0.1 nM, neither the amplitude (Figure 7.17(c)/7.18(a)) nor the signaling time (Figure 7.18(b)) vary.

Interestingly, the output of the module MEK depends on the concentration of the ligand (Figure 7.17(d))[220]. Therefore, it is the ERK module that produces independence from the ligand concentration. This phenomenon is due to the sigmoid input/output relationship of the module ERK (Hill coefficient of 2.44), and to the fact that the threshold value of the ERK module ($K_{0.5}^h \approx 3100$ molecules per cell (molec/cell)) is reached for low stimuli, $\approx$ 0.007 nM. Hence, the output (ERK-PP) shows little variation for input values above 0.1 nM (Figure 7.17(e))[220].

(a) Three relevant stepwise stimuli

(b) Response of ERK to the inputs depicted in Figure 7.17(a)

(c) S(ERK-PP) vs EGF

(d) S(MEK-PP) vs EGF

(e) S(ERK-PP) vs MEK
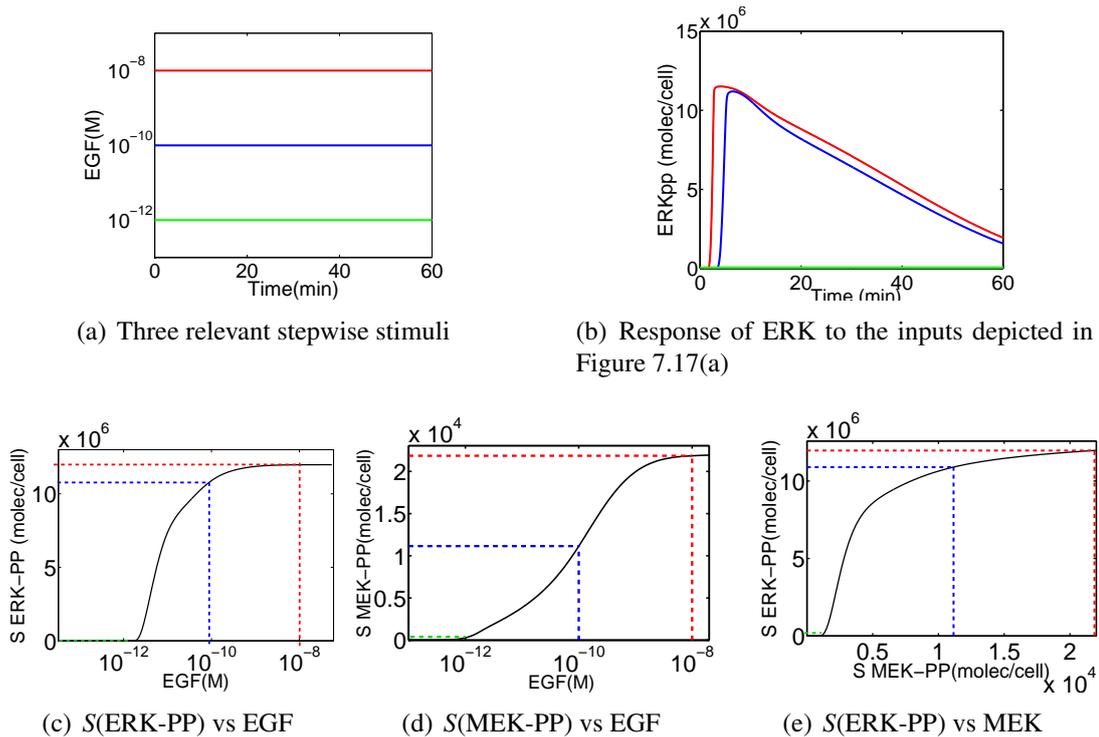
Figure 7.17: Analysis of the input/output behavior of the EGF induced MAPK cascade. For a wide range of inputs (blue to green line in Figure 7.17(a)), the output (ERK-PP) remains almost unaltered (Figure 7.17(b)). This information can be represented in a compact manner with a characteristic curve plotting the signal amplitude S of ERK-PP vs. input[220] (Figure 7.17(c)). Since MEK does not show this saturation (Figure 7.17(d)), this property can be traced back to the ERK module[220] (Figure 7.17(e))

Additionally, the contribution of the internalized receptors is important only for input values below 0.1 nM[234], since for higher values the amplitude of the signal for the internalized pathway is negligible, see Figure 7.18.

Another point still unclear is the role of the adaptor molecule Shc, which provides a second mechanism for activating Ras. The Shc-dependent pathway is redundant and seems to be preferentially used[97]. When this pathway is disabled - by setting the concentration of Shc to zero, simulating an Shc knock-out, - the output signal is slightly lower for high EGF concentrations, higher than roughly 0.1 nM, and slightly higher for lower values, between approximately 0.01 nM and 0.1 nM (see Figure 7.18(a))[220]. For EGF concentrations under

(a) S(ERK-PP)

(b) $\tau_{0.9}$(ERK-PP)
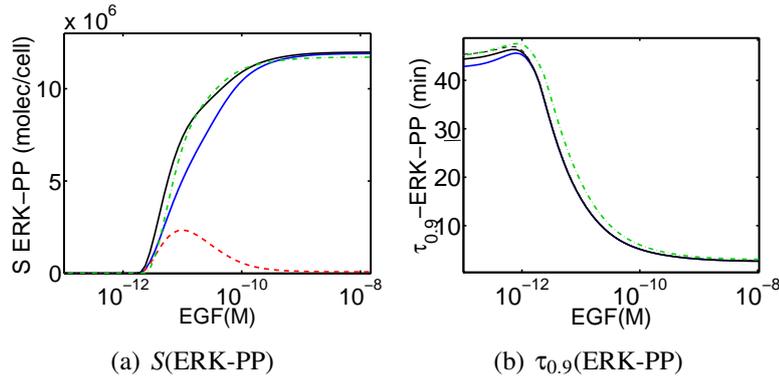
Figure 7.18: Signal amplitude and signaling time in the EGF-induced MAPK cascade. Signal amplitude S and signaling time $\tau_{0.9}$ for (i) the total output of the complete model (solid black line), (ii) the total output of the model without Shc (dashed-dot green line), (iii) the output of the complete model due to the receptors on the surface (solid blue line) and (iv) the output of the complete model due to the internalized receptors (dashed red line) from the EGF model [220;234]. The internalized receptors are important only for input values below 0.1 nM. The adaptor molecule Shc accelerates the response and increases the signal amplitude moderately for input values below 0.01 nM.

0.01 nM the signal without Shc is again lower, but in this case the difference is relatively more important, see Figure 7.18(a). The system reacts faster if Shc is present, although the difference is small (Figure 7.18(b)). Therefore, Shc seems to play an important role only at low EGF concentrations [233].

This subsection illustrates how the parameters introduced in Section 5.3.1.2 help to describe in a compact manner key properties of signaling networks (and thus facilitate their analysis), and how a modular analysis can uncover the elements responsible for a certain property (e.g. that the low sensitivity to ligand concentration of the EGF-induced MAPK cascade can be traced back to the saturation of the ERK module [220]). In the following section it will be shown that a modular decomposition also facilitates model reduction.

7.2.1 Model reduction of the EGF-induced MAPK cascade model

An advantage of a decomposition into modules without retroactivity is that, if a less complex model can be found which retains the essential input/output behavior of the original module, it can directly replace the original one, leading to a model reduction [47]. In the case of (weak) retroactivity, a certain difference may appear.

In Section 5.3.2 it was shown that, for extreme cases, simple motifs can be approximated by linear systems or, for more general conditions, by a combination of a non-linear characteristic curve and a linear system (i.e., by a Hammerstein module). For more complex modules, however, finding such a simple replacement may be probably not possible. Nevertheless, one is also interested in reducing such modules - probably more than in the case of the simple motifs, since the more complex the module, the larger the potential reduction is and, besides, the
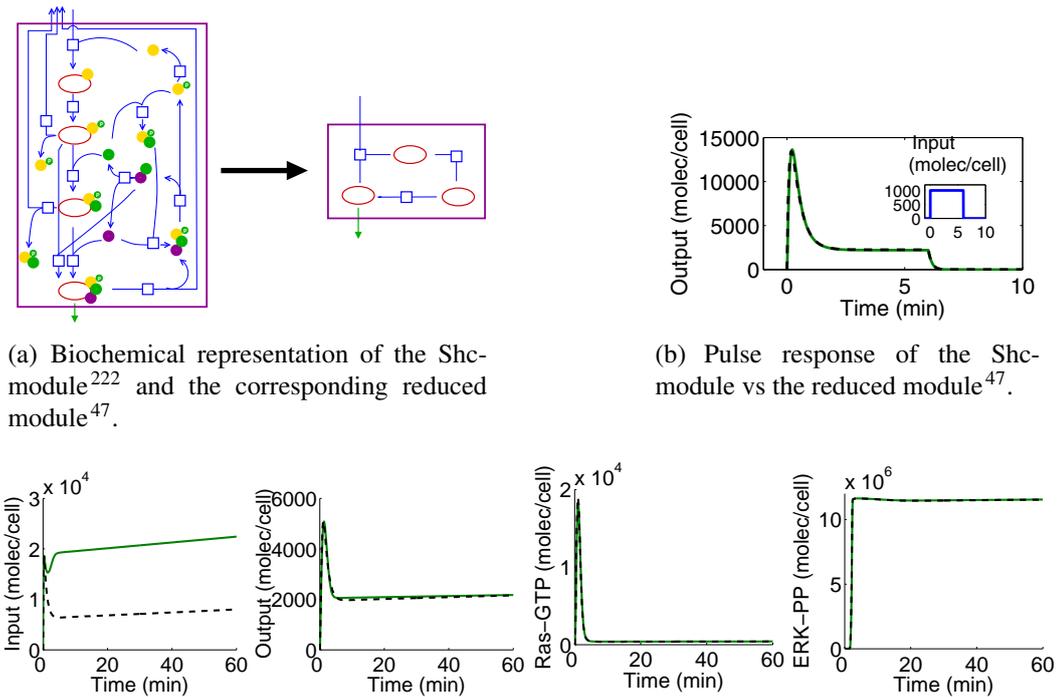
(a) Biochemical representation of the Shc-module[222] and the corresponding reduced module[47].

(b) Pulse response of the Shc-module vs the reduced module[47].

(c) Comparison of the original model (EGF-induced MAPK[234] without receptor internalization) and a reduced one replacing the complex formation with the simple pattern of Figure 7.19(a). In both cases the system is stimulated by a step in the EGF concentration. The input (EGF-EGFR*) and output (active form of Sos) of the module, as well as 2 downstream signals (Ras-GTP and ERK-PP), are depicted. Note how there is a difference at the value of the input, due to the retroactive effects.

Figure 7.19: Modular reduction of the model of the EGF-induced MAPK cascade. Simulation studies (e.g. Figure 7.19(b)) showed that many modules of the cascade (e.g. the Shc module) behave like a simple pattern Figure 7.19(a). Thus, modules of the model could be replaced by the simple pattern. For example, the module comprising the complex formation could be replaced[47], and the behavior was similar to the original model (Figure 7.19(c)). In all cases, the response of the original model is depicted with black, solid line, and of the reduced in green, dashed lines.

harder it is to understand (and thus the more helpful is a simplification). There are a number of model reduction methods which can be applied to signaling systems[147;221]: elimination of non-observable states[48] (see Section 4.1), substitution of differential equations by algebraic equations using conservation relationships[227], separation of fast from slow processes if different time hierarchies are present[149;290], neglect of states using optimization techniques[172], etc. Some of these methods have been recently reviewed elsewhere[147].

However, in many cases, rigorous methods do not provide enough reduction or are even not applicable. In these cases a more heuristic approach, based on simulation studies, can be of help[47]. We shall illustrate its application briefly using as a case study, again, the modularized model for the EGF-induced MAPK cascade of Schoeberl et al.[234].

If possible, the first step should be to linearize the modules, since a linearization provides insight into the behavior of the different modules (see also Section 5.3.1.1). For example,

the module regarding the EGF reception (Figure 3.9), can be analytically linearized, and a subsequent analysis reveals that this subunit behaves as a differentiating system with a third order lag behavior[47]. For more complex modules, such as the complex formation, testing the response of the system to different inputs allows one to find the linear system that best reproduces its behavior. However, due to the strong role played by the nonlinearities, a linear system cannot completely reproduce the dynamics of this and of most modules.

In such cases, one can try to replace the nonlinear modules by less complex nonlinear models, showing approximately the same input/output (I/O) behavior. This substitution requires a detailed knowledge of the model dynamics and its transfer behavior, which not only helps to find suitable model reductions, but also to get a qualitative and quantitative comprehension of the system. However, only few tools exist to analyze the dynamics of nonlinear models and they are mainly restricted to the analysis of low dimensional systems. Therefore, simulation studies were used here , which can be applied to the analysis of the I/O behavior of essentially all systems: considering the system's output responses to input signals like steps, oscillations or pulses may help to categorize the system's transfer performance and to understand the signal processing.

Surprisingly, simulation studies of the EGF model showed that the I/O behavior of several complex modules was strikingly similar (see Figure 7.19(b)) to that of a simple pattern (Figure 7.19(a)). Thus, the most complex modules could be replaced by the structure of the most simple one, obtaining a very good approximation to the original model[47;221] (Figure 7.19(c)). The output of the module and downstream elements showed a behavior remarkably similar to the original model while the input, however, exhibits a clear difference, due to the retroactive effects.

The modules which can be approximated by this pattern involve different biochemical processes. For example, the Raf-Ras module is mainly a transfer of phosphate between two proteins, while the complex formation modules involve mainly physical interactions among proteins. However, they can be both reproduced by the same pattern, meaning that the dynamics, at least as far as they have been analyzed[47], are very similar. It is tempting to speculate whether this pattern might also be applied to other signal transduction systems; it might be that, due to its characteristics for the transfer of information, is a recurring theme in signal transduction.

In summary, simulation studies applied in a modular fashion might be very helpful to analyze the input/output behavior of a system, and may also be helpful in finding suitable model reductions.

7.3 Conclusions

This chapter explored the application of a modular concept to analyze the dynamics of signaling networks, using differential equations as modeling formalism. The focus on the dynamics gives rise to a plethora of new aspects (oscillations, steady-state characteristic curves, etc.) compared to the Boolean approach of Chapter 6, but, on the other hand, these properties can only be studied on systems of moderate size: the EGF-induced MAPK studied here is arguably a limit case.

We started with a simple concatenation of motifs to set up the MAPK cascade. This example presented 'expectable' emergent properties: e.g., that the characteristic curve of a series of motifs is roughly a product of that of the elements, or that closing the loop with a negative feedback can produce oscillations.

In the next section, a feedback of two simple motifs, inspired in TCR signaling, gave rise to a number of interesting results. First, a small model was found which could fit a set of data. However, the system-theoretical properties were not as they should, and the model had to be extended. This result illustrates the fact that solely fitting of data has a limited meaning. Subtle analysis of the properties of the underlying mathematical systems, however, can provide insights into the design of biological systems, as was in our case the fact that a feedback driven by an autocatalytic step plus degradation produces a robust switch to block undesired signaling. In this context, the results of the thorough analysis of the motifs of Chapter 5 proved to be of great help: by knowing what one can expect from the single motifs, one can concentrate on the emergent properties arising from their combination. For example, the fact that the characteristic curve for all motifs are monotone, gave us the security that the results with regard to the characteristic curve of the simple feedback model of Section 7.1.2 would apply independently of the particular implementation of the motifs. Additionally, we could look into the catalog of motifs to find one possessing a particular property, namely, the ability to show bistability, and easily incorporate it into the model.

Finally, in Section 7.2.1, a model already available of the EGF-induced MAPK[234] was extensively analyzed. Firstly, we focus on some holistic properties of the whole cascade. For example, the insensitiveness to changes in the concentration of the input (ligand): using a modular approach and characteristic curves, one can elegantly show that this key property is due to the saturation of the ERK module. Furthermore, thorough simulation studies revealed that the I/O behavior of many modules was similar. This allowed to replace the complex modules with those which, yet simpler, behave similarly from a dynamic point of view, leading to a reduction of the model as a whole. However, these methods might not be applicable to all models, as there may exist models that cannot be decomposed into retroactive-free modules, and thus the reduced modules can not substitute perfectly the original one if the analyzed units are not completely free of retroactive effects. Furthermore, this approach is

rather heuristic, since it is case-dependent and requires a certain expertise from the analyzer (e.g. for finding out a suitable reduced structure), and its applicability to other systems is an open question.

Nevertheless, it can be expected that, as models grow in size and detail, model reduction will be more demanded and, thus, new methods for model reduction should be developed, tailored to the idiosyncrasy of signal transduction systems.

Summary

Two characteristic properties of biochemical (in particular signaling) networks are their complexity and their inherent modularity. The main idea of this work is to take advantage of the latter to facilitate to unravel the former. To apply such a modular approach the following tasks must be performed:

1. decompose the signaling network of interest into sensible modules,

2. analyze the modules thoroughly, and

3. rewire the modules (eventually substituted by a simplified version) into the network and analyze the network.

In this thesis, several proposals to perform these steps have been presented:

First, in Chapter 3 a novel criterion for the decomposition of networks considering the concept of absence of retroactivity (i.e., unidirectional connection) was defined, and subsequently implemented into an algorithm. This allows to take a model (e.g. in SBML format) 'blindly' (i.e., without *a priori* knowledge about it) and untangle it into subunits so that the number of retroactive connections among the modules are minimized. Furthermore, this information can be automatically used in ProMoT to generate a modularly-structured model. Thus, this approach not only provides insights into the structure of the model under study, but, thanks to ProMoT's inherent modularity, also provides a unique framework for a modular approach: one takes a 'flat' (non-modularized) model, decompose it into sensible subunits, and can create models of the modules for the analysis. The method was demonstrated by decomposing several models of high difficulty, either due to their size (Section 3.4.3) or high level of entanglement (Section 3.4.4).

This method follows a sort of 'top-down' approach: consider a network and try to unravel its modular properties. An inverse, 'bottom-up' perspective could also be revealing: consider the molecular architecture of signaling systems with regard to the modularity. Recently, important modeling efforts have been made that consider protein domains as the building blocks of signaling processes (as is the case from a molecular perspective).A rigorous description leads to a combinatorial explosion of the number of states. However, recent work has demonstrated that this complexity can be significantly reduced without loss of exactitude

if a domain-oriented approach is used. Moreover, applying this method, signal transduction networks show a modular structure where each module corresponds to a molecular domain and is connected in a unidirectional manner. Since the possible events taking place at a single domain are discrete, the number of possible modules is limited. Therefore, one can develop a construction kit of motifs, so that most signal transduction networks can be set up as an aggregation of those elements, as explained in Chapter 4. These results provide thus a perspective to modularity complementary to that introduced in Chapter 3.

Regarding the second point (analysis of motifs), the construction kit defined in Chapter 4 was thoroughly analyzed in Chapter 5 with regard to their multistability, monotony and input/output transfer behavior. First, multistability was addressed, because it is essential not only from a system-theoretical point of view, but also from a biological one, as it is required for essential processes such as differentiation or life/death decision-making. Therefore, the identification of biochemical motifs related to multistationarity can provide important insights into the rationale behind signal processing.

Thereby several theoretical methods (bifurcation analysis, identification of positive feedback loops, Chemical Reaction Network Theory and a new method recently developed), were used and subsequently compared. The analysis shows that, apart from double-step activation motifs including a distributive mechanism, only those motifs involving an autocatalytic reaction can show multistationarity. Furthermore, statistical analyses of an extensive number of bifurcation plots reveals that it might not be robust at the single-motif level, since the range of protein concentrations compatible with multistability is relatively narrow and of the same order of the natural fluctuation of the concentrations of proteins in cells of mammals, yeast, and bacteria.

As a second property, the monotony was addressed. Monotone systems are well behaved in a mathematical sense, what facilitates enormously the analysis of systems set up as combination of them. Remarkably, all motifs but one proofed to be monotone. Furthermore, all systems have a monotone characteristic curve. These results suggest that the main source of complexity in biochemical systems are the entanglement of motifs, rather than the single motifs themselves.

Finally, encouraged by these results, the equivalence between simple linear systems and biochemical units was explored. While the former can only approximate the behavior of the latter under extreme conditions (e.g. very weak input), the combination of a non-linear monotone characteristic curve and a linear dynamical system (a so-called Hammerstein module) proofed to approximate relatively well the biochemical systems and provide insight into their properties and the importance of the different kinetic parameters on them.

The results from Chapter 5 provide thus a first step towards a catalog of fully characterized signaling modules. Furthermore, these results can be seen as a proof of principle for the methods used there.

As far as the third point (rewiring and subsequent analysis) is concerned, the rationale is that, thanks to the knowledge won in the second step, one can concentrate on the emergent properties resulting from the influences among the modules. Therefore, whenever possible, one should simplify the description of the single modules to be able to tackle large signaling networks. Based on a strong simplification (replacement of each motif by a Boolean state), a new analysis framework was developed, and the abilities of the tools CellNetAnalyzer and ProMoT were extended, to set up and analyze models of large signaling networks. Even though this methodology requires a strong simplification of the reality, it allows to gain holistic insights into large networks. To illustrate its applicability, a large, curated model of T-cell signaling was set up and subsequently analyzed[224]. This model is, to the best of our knowledge, the largest of its sort up to date, and comprises as many as 94 compounds and 123 interactions among them.

In Chapter 7, the emergent of new dynamic behavior was addressed with kinetic models. First the effects on the steady state and dynamic behavior of concatenating motifs (and eventually embedding them in a loop) was explored using the well-known MAPK cascade as an example.

Later, a minimalist model was outlined to explain a specific non-trivial dynamic behavior in the T-cell induced MAPK cascade in Section 7.1.2. This simplest model (which comprises the minimal number of elements required to explain a set of data) was not only able to explain the observed behavior, but also to provide insights into the design of the mechanisms responsible for it. However, a large, detailed model product of a bottom-up approach was not able to explain this data. These results illustrate the problems of parameter estimation and model discrimination and advocates for the use of simple models at least as a first approach to a particular question.

Finally, the properties of the model describing the EGF-induced MAPK cascade of Schoeberl et al., which had been decomposed in Chapter 3, are analyzed. A modular perspective elegantly shows that it is the last module (ERK) which is responsible for the insensitiveness of the model to differences in the input, a characteristic of this system, among other insights. These results provide a proof of principle of the application of the modular methodology to an already-defined model.

In summary, the methods presented here contribute to the different steps required for a modular approach. This work does certainly not answer all questions, but it provides an initial framework towards an analysis based upon a modular rationale. Different aspects could be further investigated. An important issue would be to facilitate the work-flow between the different steps. It would be for example useful to have methods to (semi)automatically convert models defined using a kinetic approach into Boolean ones and vice versa. ProMoT seems to be a suitable candidate to integrate the different aspects. The potential advantages of a modular perspective grow with the system under consideration. Here, systems of certain

size have been used as case studies. The examples were chosen as a compromise of complexity, relevance, and available information. In the future, however, larger systems should be tackled and successively integrated, with the ultimative goal to consider a whole cell, a whole organ, and, eventually, the whole human organism.

Bibliography

[1] D. R. Alessi, *et al.* Characterization of a 3-phosphoinositide-dependent protein kinase which phosphorylates and activates protein kinase Balpha. *Curr Biol*, 7:261–269, 1997.

[2] U. Alon, M. G. Surette, N. Barkai, S. Leibler. Robustness in bacterial chemotaxis. *Nature*, 397, January 1999.

[3] G. Altan-Bonnet, R. N. N. Germain. Modeling T cell antigen discrimination based on feedback control of digital ERK responses. *PLoS Biol*, 3(11), October 2005.

[4] G. Altan-Bonnet, R. N. N. Germain. Modeling t cell antigen discrimination based on feedback control of digital erk responses. *PLoS Biol*, 3(11), October 2005.

[5] K. Anderson, C. Kane. Ca/calmodulin-dependent protein kinase iv and calcium signaling. *Biometals*, 11(4):331 – 343, 1998.

[6] K. E. Anderson, J. Coadwell, L. R. Stephens, P. T. Hawkins. Translocation of PDK-1 to the plasma membrane is important in allowing PDK-1 to activate protein kinase B. *Curr Biol*, 8:684–691, 1998.

[7] D. Angeli, J. Ferrell, E. Sontag. Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proc. Natl. Acad. Sci. U.S.A.*, 101(7):1822–1827, 2004.

[8] M. Arnaud, *et al.* Interaction of the tyrosine phosphatase SHP-2 with Gab2 regulates Rho-dependent activation of the c-fos serum response element by interleukin-2. *Biochem J*, 382:545–556, 2004.

[9] R. Arnold, *et al.* Activation of hematopoietic progenitor kinase 1 involves relocation, autophosphorylation, and transphosphorylation by protein kinase D1. *Mol Cell Biol*, 25:2364–2383, 2005.

[10] A. Asthagiri, D. Lauffenburger. Bioengineering models of cell signaling. *Annu. Rev. Biomed. Eng.*, 2:31–53, 2000.

[11] A. Asthagiri, D. Lauffenburger. A computational study of feedback effects on signal dynamics in a mitogen-activated protein kinase (MAPK) pathway model. *Biotechnol. Prog.*, 17(2):227–239, 2001.

[12] A. August, B. Dupont. CD28 of T lymphocytes associates with phosphatidylinositol 3-kinase. *Int Immunol*, 6:769–774, 1994.

[13] P. Bak, K. Sneppen. Punctuated equilibrium and criticality in a simple model of evolution. *Physical Review Letters*, 71:4083–4086, dec 1993.

[14] A.-L. Barabasi, Z. N. Oltvai. Network biology: understanding the cell's functional organization. *Nat Rev Genet*, 5(2):101–113, Feb 2004.

[15] V. Batagelj, A. Mrvar. Pajek - analysis and visualization of large networks. *Lect. Notes Comput. Sci.*, 2265:477–478, 2002.

[16] A. Becskei, B. Sraphin, L. Serrano. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J*, 20(10):2528–2535, May 2001.

[17] K. Behrendt, K. Mauch. Verfahren und analyseeinrichtung zum identifizieren von wirkorten in einem signalnetzwerk. Patentanmeldung, 2003.

[18] S. G. Benjamini E, Coico R. *Immunology-A short course.* Wiley-Liss, 2000.

[19] M. Bentires-Alj, *et al.* A role for the scaffolding adapter GAB2 in breast cancer. *Nat Med*, 12:114–121, 2006.

[20] U. Bhalla, R. Iyengar. Emergent properties of networks of biological signaling pathways. *Science*, 283(5400):381–387, January 1999.

[21] B. Binder, R. Heinrich. Interrelations between dynamical properties and structural characteristics of signal transduction networks. *Genome Informatics Serie*, 15(1):13–23, 2004.

[22] T. G. Bivona, *et al.* Phospholipase cgamma activates ras on the golgi apparatus by means of rasgrp1. *Nature*, 424(6949):694–698, August 2003.

[23] M. L. Blinov, J. R. Faeder, B. Goldstein, W. S. Hlavacek. Bionetgen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics*, 20(17):3289–3291, 2004.

[24] N. Blüthgen, H. Herzel. How robust are switches in intracellular signaling cascades? *J. theor. Biol.*, 225(3):293–300, 2003.

[25] N. Blüthgen, F. J. Bruggeman, S. Legewie, H. Herzel, H. V. Westerhoff, B. N. Kholodenko. Effects of sequestration on signal transduction cascades. *FEBS J*, 273:895–906, 2006.

[26] S. Boettcher, A. G. Percus. Optimization with extremal dynamics. *PHYS.REV.LETT*, 86:5211–5214, 2001.

[27] N. M. Borisov, N. I. Markevich, J. B. Hoek, B. N. Kholodenko. Signaling through Receptors and Scaffolds: Independent Interactions Reduce Combinatorial Complexity. *Biophys J*, 89(2):951–966, Aug 2005.

[28] D. Bray. Protein molecules as computational elements in living cells. *Nature*, 376 (6538):307–312, 1995.

[29] F. A. Brightman, D. A. Fell. Diferential feedback regulation of the MAPK cascade underlies the quantitative differences in EGF and NGF signalling in PC12 cell. *FEBS Letters*, 482(3):169–174, 2000.

[30] L. Buday, S. E. Egan, P. Rodriguez Viciana, D. A. Cantrell, J. Downward. A complex of grb2 adaptor protein, sos exchange factor, and a 36-kda membrane-bound tyrosine phosphoprotein is implicated in ras activation in t cells. *J Biol Chem*, 269(12):9019–9023, March 1994.

[31] D. A. Cantrell. Gtpases and t cell activation. *Immunol Rev*, 192:122–130, April 2003.

[32] C. Chan, A. George, J. Stark. T cell sensitivity and specificity- kinetic proofreading revisited. *Discrete and continuous dynamical systems-Series B*, 3(3):343–360, 2003.

[33] C. Chan, J. Stark, A. J. T. George. Feedback control of T-cell receptor activation. *Proc. R. Soc. London B.*, 271(1542):931–939, 2004.

[34] C. Chan, G. A.J., J. Stark. Cooperative enhancement of specificity in a lattice of T cell receptors. *PNAS*, 98(10):5758–5763, 2001.

[35] T. O. Chan, S. E. Rittenhouse, P. N. Tsichlis. AKT/PKB and other D3 phosphoinositide-regulated kinases: Kinase activation by phosphoinositide-dependent phosphorylation. *Annu Rev Biochem*, 68:965–961014, 1999.

[36] F. Chang, *et al.* Involvement of PI3K/Akt pathway in cell cycle progression, apoptosis, and neoplastic transformation: a target for cancer chemotherapy. *Leukemia*, 17:590–603, 2003.

[37] L. Chang, M. Karin. Mammalian MAP kinase signalling cascades. *Nature*, 410(6824): 37–40, 2001.

[38] F. Z. Chaoui, F. Giri, Y. Rochdi, M. Haloua, A. Naitali. System identification based on hammerstein model. *Intern. J. of Control*, 78(6):430–442, 2005.

[39] M. Chaves, R. Albert, E. D. Sontag. Robustness and fragility of boolean models for genetic regulatory networks. *J Theor Biol*, 235(3):431–449, Aug 2005.

[40] J. Chen, B. Yuan. Detecting functional modules in the yeast protein-protein interaction network. *Bioinformatics*, 22(18):2283–2290, Sep 2006.

[41] K. Choudhuri, A. Kearney, T. R. Bakker, P. A. van der Merwe. Immunology: how do T cells recognize antigen? *Curr Biol*, 15(10):382–385, May 2005.

[42] K. Choudhuri, D. Wiseman, M. H. Brown, K. Gould, P. A. van der Merwe. T-cell receptor triggering is critically dependent on the dimensions of its peptide-MHC ligand. *Nature*, 436(7050):578–582, Jul 2005.

[43] F. Ciardiello, G. Tortora. Anti-epidermal growth factor receptor drugs in cancer therapy. *Expert Opin. Investig. Drugs*, 11(6):755–68, 2002.

[44] O. Cinquin, J. Demongeot. Positive and negative feedback:striking a balance between necessary antagonists. *J. theor. Biol.*, 216:229 241, 2002.

[45] C. Conradi, J. Saez-Rodriguez, E. Gilles, J. Raisch. Using chemical reaction network theory to discard a kinetic mechanism hypothesis. *IEE Proc. Systems Biology*, 152 (4):243–248, 2005.

[46] C. Conradi, D. Flockerzi, J. Raisch. Multistationarity in the activation of a MAPK: parametrizing the relevant region in parameter space. *submitted*, 2006.

[47] H. Conzelmann, J. Saez-Rodriguez, T. Sauter, E. Bullinger, F. Allgöwer, E. D. Gilles. Reduction of mathematical models of signal transduction networks: Simulation-based approach applied to EGF receptor signaling. *Systems Biology*, 1(1):159–169, 2004.

[48] H. Conzelmann, J. Saez-Rodriguez, T. Sauter, B. Kholodenko, E. D. Gilles. A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks. *BMC Bioinformatics*, 7:34, 2006.

[49] D. Coombs, A. M. Kalergis, S. G. Nathenson, C. Wofsy, B. Goldstein. Activated TCR s remain marked for internalization after dissociation from pmhc. *Nat Immunol*, 3 (10):926–31, 2002.

[50] M. J. Czar, J. Debnath, E. M. Schaeffer, C. M. Lewis, P. L. Schwartzberg. Biochemical and genetic analyses of the tec kinases itk and rlk/txk. *Biochem Soc Trans*, 29(Pt 6): 863–867, November 2001.

[51] M. A. Daniels, *et al.* Thymic selection threshold defined by compartmentalization of ras/mapk signalling. *Nature*, Nov 2006.

[52] M. S. Dasika, A. Burgard, C. D. Maranas. A computational framework for the topological analysis and targeted disruption of signal transduction networks. *Biophys J*, 91(1):382–398, Jul 2006.

[53] D. Davidson, M. Bakinowski, M. L. Thomas, V. Horejsi, A. Veillette. Phosphorylation-dependent regulation of T-cell activation by pag/cbp, a lipid raft-associated transmembrane adaptor. *Mol Cell Biol*, 23(6):2017–2028, March 2003.

[54] R. J. Davis. Signal transduction by the jnk group of map kinases. *Cell*, 103(2):239–252, October 2000.

[55] S. J. Davis, P. A. van der Merwe. The kinetic-segregation model: TCR triggering and beyond. *Nat Immunol*, 7(8):803–809, Aug 2006.

[56] H. de Jong. Modeling and simulation of genetic regulatory systems: A literature review. *J. of Comp. Biol.*, 9-1:67–103, 2002.

[57] J. A. Deane, D. A. Fruman. Phosphoinositide 3-kinase: Diverse roles in immune cell activation. *Annu Rev Immunol*, 22:563–598, 2004.

[58] B. Dérijard, *et al.* Independent human MAP-kinase signal transduction pathways defined by MEK and MKK isoforms. *Science*, 267:682–685, 1995.

[59] P. P. Di Fiore. Signal transduction: life on mars, cellularly speaking. *Nature*, 424 (6949):624–625, August 2003.

[60] J. Downward. The ins and outs of signalling. *Nature*, 411(6839):759–762, 2001.

[61] L. Duan, A. L. Reddi, A. Ghosh, M. Dimri, H. Band. The Cbl family and other ubiquitin ligases: destructive forces in control of antigen receptor signaling. *Immunity*, 21(1):7–17, Jul 2004.

[62] J. Duch, A. Arenas. Community detection in complex networks using extremal optimization. *Physical Review E*, 72:027104, 2005.

[63] M. Ederer, T. Sauter, E. Bullinger, E. D. Gilles, F. Allgöwer. An approach for dividing models of biological reaction networks into functional units. *Simulation*, 79(12):703–716, 2003.

[64] T. Eissing, F. Allgoewer, E. Bullinger. Robustness properties of apoptosis models with respect to parameter variations and intrinsic noise. *IEE Proc. Sys. Biol.:*, 152: 221–228, 2005.

[65] T. Eissing, H. Conzelmann, E. D. Gilles, F. Allgower, E. Bullinger, P. Scheurich. Bistability analyses of a caspase activation model for receptor-induced apoptosis. *J Biol Chem*, 279(35):36892–36897, Aug 2004.

[66] M. Elowitz, A. Levine, E. D. Siggia, P. Swain. Stochastic gene expression in a single cell. *Science*, 297:1183–1186, 2002.

[67] G. Enciso, H. Smith, E. Sontag. Non-monotone systems decomposable into monotone systems with negative feedback. *J. of Differential Equations*, 224:205–227, 2006.

[68] J. R. Faeder, M. L. Blinov, B. Goldstein, W. S. Hlavacek. Combinatorial complexity and dynamical restriction of network flows in signal transduction. *Systems Biology*, 2 (1):4–15, 2005.

[69] I. Famili, B. O. Palsson. The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools. *Biophysical Journal*, 85:16–26, 2003.

[70] D. Fang, Y. C. Liu. Proteolysis-independent regulation of PI3K by Cbl-b-mediated ubiquitination in T cells. *Nat Immunol*, 2:870–875, 2001.

[71] G. Fanger, N. Johnson, G. Johnson. Mek kinases are regulated by egf and selectively interact with rac/cdc42. *EMBO J*, 16(16):4961–4972, 1997.

[72] M. Feinberg. The existence and uniqueness of steady states for a class of chemical reaction networks. *Archive for Rational Mechanics and Analysis*, 132(4):311–370, 1995.

[73] M. Feinberg. Multiple steady states for chemical reaction networks of deficiency one. *Archive for Rational Mechanics and Analysis*, 132(4):371–406, 1995.

[74] M. Feinberg, P. R. Ellison. The chemical reaction network toolbox. http//www.chbmeng.ohio-state.edu/ feinberg/crnt.

[75] J. E. J. Ferrell. Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. *Trends Biochem. Sci.*, 21(12):460–466, 1996.

[76] J. E. J. Ferrell. How responses get more switch-like as you move down a protein kinase cascade. *Trends Biochem. Sci.*, 22(8):288–289, 1997.

[77] J. E. J. Ferrell. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr. Opin. Cell. Biol.*, 14(2):140–148, 2002.

[78] J. E. J. Ferrell, W. Xiong. Bistability in cell signaling: How to make continuous processes discontinuous, and reversible processes irreversible. *Chaos*, 11(1), March 2001.

[79] S. Feske, J. Giltnane, R. Dolmetsch, L. M. Staudt, A. Rao. Gene regulation mediated by calcium signals in T lymphocytes. *Nat Immunol*, 2:316–324, 2001.

[80] D. Filipp, M. Julius. Lipid rafts: resolution of the "fyn problem"? *Mol Immunol*, 41 (6-7):645–656, July 2004.

[81] D. Filipp, B. L. Leung, J. Zhang, A. Veillette, M. Julius. Enrichment of lck in lipid rafts regulates colocalized fyn activation and the initiation of proximal signals through TCR alpha beta. *J Immunol*, 172(7):4266–4274, April 2004.

[82] O. Föllinger, M. K. Frank Dörrscheidt. *Regelungstechnik : Einführung in die Methoden und ihre Anwendung*. Hüthig, 2005.

[83] R. A. Franklin, A. Tordai, H. Patel, A. M. Gardner, G. L. Johnson, E. W. Gelfand. Ligation of the t cell receptor complex results in activation of the ras/raf-1/mek/mapk cascade in human t lymphocytes. *J Clin Invest*, 93(5):2134–2140, May 1994.

[84] M. Frdin, S. Gammeltoft. Role and regulation of 90 kda ribosomal s6 kinase (rsk) in signal transduction. *Mol Cell Endocrinol*, 151(1-2):65–77, May 1999.

[85] A. Funahashi, N. Tanimura, M. Morohashi, H. Kitano. Celldesigner: a process diagram editor for gene-regulatory and biochemical networks. *Biosilico*, 1:159–162, 2003.

[86] O. Gaide, *et al.* Carma1 is a critical lipid raft-associated regulator of TCR -induced nf-kappa b activation. *Nat Immunol*, 3(9):836–843, September 2002.

[87] S. Gayer. Automatic decomposition of signaling network using the concept of retroactivity. Master's thesis, University of Stuttgart, 2006.

[88] S. Gayer. Analysis of dynamic properties of signal transduction modules. Student thesis, University of Stuttgart, 2006.

[89] E. Genot, D. A. Cantrell. Ras regulation and function in lymphocytes. *Curr Opin Immunol*, 12:289–294, 2000.

[90] M. Ghiotto-Ragueneau, M. Battifora, A. Truneh, M. D. Waterfield, D. Olive. Comparison of CD28-B7.1 and B7.2 functional interaction in resting human T cells: Phosphatidylinositol 3-kinase association to CD28 and cytokine production. *Eur J Immunol*, 26:34–41, 1996.

[91] S. Gibson, *et al.* Efficient CD28 signalling leads to increases in the kinase activities of the TEC family tyrosine kinase EMT/ITK/TSK and the SRC family tyrosine kinase LCK. *Biochem J*, 330 (Pt 3):1123–1128, 1998.

[92] D. Gil, A. G. Schrum, B. Alarcon, E. Palmer. T cell receptor engagement by peptide-MHC ligands induces a conformational change in the CD3 complex of thymocytes. *J Exp Med*, 201(4):517–522, Feb 2005.

[93] E. D. Gilles. Network theory for chemical processes. *Chem. Eng. Technol.*, 21(2): 121–132, 1998.

[94] E. D. Gilles, U. Knöpp. *Skriptum zur Vorlesung Regelungstechnik I.* Universität Stuttgart, Institut für Systemdynamik und Regelungstechnik, 1991.

[95] M. Ginkel, A. Kremling, T. Nutsch, R. Rehner, E. D. Gilles. Modular modeling of cellular systems with ProMoT/Diva. *Bioinformatics*, 19(9):1169–1176, 2003.

[96] A. Goldbeter, D. Koshland. An amplified sensitivity arising from covalent modification in biological systems. *Proc. Nac. Ac. Sci. U.S.A.*, 78(11):6840–6844, 1981.

[97] Y. Gong, X. Zhao. Shc-dependent pathway is redundant but dominant in MAPK cascade activation by EGF receptors: a modeling inference. *FEBS Letters*, 554(3): 467–472, 2003.

[98] L. A. Gravestein, D. Amsen, M. Boes, C. R. Calvo, A. M. Kruisbeek, J. Borst. The TNF receptor family member CD27 signals to Jun N-terminal kinase via Traf-2. *Eur J Immunol*, 28:2208–2216, 1998.

[99] Z. Guan, S. Y. Buckman, A. P. Pentland, D. J. Templeton, A. R. Morrison. Induction of cyclooxygenase-2 by the activated mekk1 right-arrowsek1/mkk4 right-arrowp38 mitogen-activated protein kinase pathway. *J. Biol. Chem.*, 273(21):12901–12908, May 1998.

[100] R. Guimera, L. A. Nunes Amaral. Functional cartography of complex metabolic networks. *Nature*, 433(7028):895–900, Feb 2005.

[101] M. Hanada, J. Feng, B. A. Hemmings. Structure, regulation and function of PKB/AKT–a major therapeutic target. *Biochim Biophys Acta*, 1697:3–16, 2004.

[102] H. Harlin, E. Podack, M. Boothby, M.-L. Alegre. TCR-independent CD30 signaling selectively induces IL-13 production via a TNF receptor-associated factor/p38 mitogen-activated protein kinase-dependent mechanism. *J Immunol*, 169:2451–2459, 2002.

[103] L. Hartwell, J. Hopfield, S. Leibler, A. Murray. From molecular to modular cell biology. *Nature*, 402(6761-supp):C47–C52, 1999.

[104] M. S. Hayden, S. Ghosh. Signaling to NF-kappaB. *Genes Dev.*, 18(18):2195–2224, 2004.

[105] S. P. Hehner, T. G. Hofmann, O. Dienz, W. Droge, M. L. Schmitz. Tyrosine-phosphorylated vav1 as a point of integration for T-cell receptor- and cd28-mediated activation of jnk, p38, and interleukin-2 transcription. *J Biol Chem*, 275(24):18160–18171, June 2000.

[106] R. Heinrich, B. Neel, T. Rapoport. Mathematical models of protein kinase signal transduction. *Molecular Cell*, 9(5):957–970, May 2002.

[107] R. Heinrich, S. Schuster. *The regulation of Cellular Systems*. Chapman Hall, 1996.

[108] Y. Herishanu, *et al.* T-cell ZAP-70 overexpression in chronic lymphocytic leukemia (CLL) correlates with CLL cell ZAP-70 levels, clinical stage and disease progression. *Leukemia*, 19:1289–1291, 2005.

[109] C. S. Hill, J. Wynne, R. Treisman. The Rho family GTPases RhoA, Rac1, and CDC42Hs regulate transcriptional activation by SRF. *Cell*, 81:1159–1170, 1995.

[110] H. J. Hinton, D. R. Alessi, D. A. Cantrell. The serine kinase phosphoinositide-dependent kinase 1 (PDK1) regulates T cell development. *Nat Immunol*, 5:539–545, 2004.

[111] W. S. Hlavacek, J. R. Faeder, M. L. Blinov, A. S. Perelson, B. Goldstein. The complexity of complexes in signal transduction. *Biotechnol. Bioeng.*, 84(7):783–794, 2004.

[112] K. A. Hogquist, S. C. Jameson, W. R. Heath, J. L. Howard, M. J. Bevan, F. R. Carbone. T cell receptor antagonist peptides induce positive selection. *Cell*, 76(1):17–27, Jan 1994.

[113] A. D. Holdorf, *et al.* Proline residues in CD28 and the Src homology (SH)3 domain of Lck are required for T cell costimulation. *J Exp Med*, 190:375–384, 1999.

[114] V. Horejsi, W. Zhang, B. Schraven. Transmembrane adaptor proteins: organizers of immunoreceptor signalling. *Nat Rev Immunol*, 4(8):603–616, August 2004.

[115] H. Housden, *et al.* Investigation of the kinetics and order of tyrosine phosphorylation in the T-cell receptor zeta chain by the protein tyrosine kinase lck. *Eur J Biochem.*, 270(11):2369–2376, 2003.

[116] M. C. Hu, W. R. Qiu, X. Wang, C. F. Meyer, T. H. Tan. Human hpk1, a novel human hematopoietic progenitor kinase that activates the jnk/sapk kinase cascade. *Genes Dev*, 10(18):2251–2264, September 1996.

[117] J. Huang, *et al.* CD28 plays a critical role in the segregation of PKC theta within the immunologic synapse. *Proc Natl Acad Sci U S A*, 99:9369–9373, 2002.

[118] Y. Huang, R. L. Wange. T cell receptor signaling: beyond complex complexes. *J Biol Chem*, 279(28):28827–28830, Jul 2004.

[119] M. Hucka, A. Finney, H. Sauro, H. Bolouri, J. Doyle, H. Kitano. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.

[120] K. Hughes, S. Edin, A. Antonsson, T. Grundstrm. Calmodulin-dependent kinase ii mediates t cell receptor/cd3- and phorbol ester-induced activation of ikappab kinase. *J Biol Chem*, 276(38):36008–36013, September 2001.

[121] T.-M. i, Y. Huang, M. I. Simon, J. Doyle. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc. Natl. Acad. Sci. U.S.A.*, 97:4649–4653, 2000.

[122] A. Isidori. *Nonlinear Control Systems*, chapter Global Decomposition of Control Systems, pages 77–105. Springer, 3rd edition, 1995.

[123] K. A. Janes, D. A. Lauffenburger. A biological approach to computational models of proteomic networks. *Curr Opin Chem Biol*, 10:73–80, 2006.

[124] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvi, A. Barabasi. The large-scale organization of metabolic networks. *Nature*, 407:651–654, 2000.

[125] H. Jeong, S. Mason, A. L. Barabasi, Z. Oltvai. Lethality and centrality in protein networks. *Nature*, 411:41–42, 2001.

[126] D. Jones, M. Sanjuan, J. Stone, I. Merida. Expression of a catalytically inactive form of diacylglycerol kinase alpha induces sustained signaling through RasGRP. *FASEB J.*, 16(6):595–597, 2002.

[127] M. Kaern, T. C. Elston, W. J. Blake, J. J. Collins. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet*, 6(6):451–464, Jun 2005.

[128] S. Kaga, S. Ragg, K. A. Rogers, A. Ochi. Activation of p21-cdc42/rac-activated kinases by CD28 signaling: p21-activated kinase (pak) and mek kinase 1 (mekk1) may

mediate the interplay between CD3 and CD28 signals. *J Immunol*, 160(9):4182–4189, May 1998.

[129] D. Kahn, H. Westerhoff. Control theory of regulatory cascades. *J. theor. Biol.*, 153: 255–285, 1991.

[130] A. Kashishian, M. Howard, C. Loh, W. M. Gallatin, M. F. Hoekstra, Y. Lai. AKAP79 inhibits calcineurin through a site distinct from the immunophilin-binding region. *J Biol Chem*, 273:27412–27419, 1998.

[131] N. Kashtan, U. Alon. Spontaneous evolution of modularity and network motifs. *Proc Natl Acad Sci U S A*, 102(39):13773–13778, Sep 2005.

[132] M. Kaufman, F. Andris, O. Leo. A model for antigen-induced t cell unresponsiveness based on autophosphorylative protein tyrosine kinase activity. *Int Immunol*, 8(4):613–24, Apr 1996.

[133] M. Kaufman, F. Andris, O. Leo. A logical analysis of t cell activation and anergy. *Proc Natl Acad Sci U S A*, 96(7):3894–3899, Mar 1999.

[134] B. N. Kholodenko. Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur. J. Biochem.*, 267(6):1583–1588, 2000.

[135] B. N. Kholodenko, O. Demin, G. Moehren, J. Hoek. Quantification of short term signaling by the epidermal growth factor receptor. *J. Biol. Chem.*, 274(42):30169–30181, 1999.

[136] B. Kholodenko, H. J.B., H. Westerhoff, G. C. Brown. Quantification of information transfer via cellular signal transduction pathways. *FEBS Letters*, 414:430–434, 1997.

[137] B. N. Kholodenko. Cell-signalling dynamics in time and space. *Nature Reviews Molecular Cell Biology*, 7(3):165–176, February 2006.

[138] A. Khoshnan, D. Bae, C. A. Tindell, A. E. Nel. The physical association of protein kinase C theta with a lipid raft-associated inhibitor of kappa B factor kinase (IKK) complex plays a role in the activation of the NF-kappa B cascade by TCR and CD28. *J Immunol*, 165:6933–6940, 2000.

[139] P. M. Kim, B. Tidor. Limitations of quantitative gene regulation models: a case study. *Genome Res*, 13(11):2391–2395, Nov 2003.

[140] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[141] H. Kitano. Systems biology: A brief overview. *Science*, 295(5560):1662–1664, 2002.

[142] S. Klamt, J. Stelling. Two approaches for metabolic pathway analysis? *Trends in Biotechnology*, 21(2):64–69, 2003.

[143] S. Klamt, J. Saez-Rodriguez, J. Lindquist, L. Simeoni, E. D. Gilles. A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics*, 7:56, 2006.

[144] S. Klamt, J. Saez-Rodriguez, E. D. Gilles. Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Systems Biology, 1:2*, 2007.

[145] E. Klipp, W. Liebermeister, C. Wierling. Inferring dynamic properties of biochemical reaction networks from structural knowledge. *Genome Inform*, 15(1):125–137, 2004.

[146] G. Krauss. *Biochemistry of Signal Transduction and Regulation*. Wiley-VCH, 2003.

[147] A. Kremling, J. Saez-Rodriguez. Systems biology - an engineering perspective. *J. Biotechnology*, 129:329, 2007.

[148] A. Kremling, K. Kahreis, J. Lengeler, E. D. Gilles. The organization of metabolic networks: a signal-oriented approach to cellular models. *Metab. Eng.*, 2(3):190–200, 2000.

[149] A. Kremling, S. Fischer, T. Sauter, K. Bettenbrock, E. Gilles. Time hierarchies in the escherichia coli carbohydrate uptake and metabolism. *Biosystems*, 73(1):57–71, 2004.

[150] V. Lafont, E. Astoul, A. Laurence, J. Liautard, D. Cantrell. The t cell antigen receptor activates phosphatidylinositol 3-kinase-regulated serine kinases protein kinase b and ribosomal s6 kinase 1. *FEBS Letters*, 486(1):38–42, December 2000.

[151] D. A. Lauffenburger. Cell signaling pathways as control modules: complexity for simplicity? *Proc. Natl. Acad. Sci. U.S.A.*, 97(10):5031–5033, 2000.

[152] N. Le Novere, T. S. Shimizu. Stochsim: modelling of stochastic biomolecular processes. *Bioinformatics*, 17(6):575–6, June 2001.

[153] N. Le Novre, *et al.* BioModels Database: A free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems, 2006.

[154] K.-H. Lee, *et al.* The Immunological Synapse Balances T Cell Receptor Signaling and Degradation. *Science*, 302(5559):1218–1222, 2003.

[155] K. Y. Lee, F. D'Acquisto, M. S. Hayden, J. H. Shim, S. Ghosh. Pdk1 nucleates t cell receptor-induced signaling complex for nf-kappab activation. *Science*, 308(5718): 114–118, April 2005.

[156] J. W. Lengeler. Metabolic networks: a signal-oriented approach to cellular models. *Biol. Chem.*, 381(9-10):911–920, 2000.

[157] S. E. Levin, A. Weiss. Twisting tails exposed: the evidence for TCR conformational change. *J Exp Med*, 201(4):489–492, Feb 2005.

[158] A. Levitzki. Targeting signal transduction for disease therapy. *Curr. Opin. Cell. Biol.*, 8(2):239–244, 1996.

[159] J. Liang, J. M. Slingerland. Multiple roles of the pi3kpkb akt pathway in cell cycle progression by jiyong liang and joyce m. slingerland: Cell cycle article abstract. *Cell Cycle*, 2(4):339–345, July 2003.

[160] J. A. Lindquist, L. Simeoni, B. Schraven. Transmembrane adapters: attractants for cytoplasmic effectors. *Immunol Rev*, 191:165–182, February 2003.

[161] J. Liou, *et al.* Hpk1 is activated by lymphocyte antigen receptors and negatively regulates ap-1. *Immunity*, 12(4):399–408, April 2000.

[162] P. V. LoGrasso, J. Hawkins, L. J. Frank, D. Wisniewski, A. Marcy. Mechanism of activation for Zap-70 catalytic activity. *Proc Natl Acad Sci U S A*, 93(22):12165–12170, Oct 1996.

[163] L. Lok, R. Brent. Automatic generation of cellular reaction networks with Moleculizer 1.0. *Nat Biotechnol*, 23(1):131–136, Jan 2005.

[164] P. C. Lucas, *et al.* Bcl10 and malt1, independent targets of chromosomal translocation in malt lymphoma, cooperate in a novel nf-kappa b signaling pathway. *J Biol Chem*, 276(22):19012–19019, June 2001.

[165] A. Ma'ayan, *et al.* Formation of regulatory patterns during signal propagation in a Mammalian cellular network. *Science*, 309(5737):1078–1083, Aug 2005.

[166] F. Macian. NFAT proteins: key regulators of T-cell development and function. *Nat Rev Immunol*, 5:472–484, 2005.

[167] M. Mangold, S. Motz, E. D. Gilles. A network theory for the structured modelling of chemical processes. *Chem. Eng. Sci.*, 57:4099–4116, 2002.

[168] N. I. Markevich, J. B. Hoek, B. N. Kholodenko. Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *J. Cell Biol.*, 164 (3):353–359, 2004.

[169] C. J. Marshall. Specificity of receptor tyrosine kinase signaling: Transient versus sustained extracellular signal-regulated kinase activation. *Cell*, 80(2):179–185, 1995.

[170] Mathworks. Matlab. http://www.mathworks.com/, 2006.

[171] S. Matsuda, F. Shibasaki, K. Takehana, H. Mori, E. Nishida, S. Koyasu. Two distinct action mechanisms of immunophilin-ligand complexes for the blockade of T-cell activation. *EMBO Rep*, 1:428–434, 2000.

[172] S. Maurya, M.R and. Bornheimer, V. Venkatasubramanian, S. Subramaniam. Reduced-order modelling of biochemical networks: application to the GTpase-cycle signalling module. *IEE Proc. Systems Biology*, 152(4):229–242, 2005.

[173] T. W. McKeithan. Kinetic proofreading in t-cell receptor signal transduction. *Proc Natl Acad Sci U S A*, 92(11):5042–5046, May 1995.

[174] L. Mendoza, D. Thieffry, E. R. Alvarez-Buylla. Genetic control of flower morphogenesis in arabidopsis thaliana: a logical analysis. *Bioinformatics*, 15(7-8):593–606, 1999.

[175] F. Michel, *et al.* CD28 utilizes Vav-1 to enhance TCR-proximal signaling and NF-AT activation. *J Immunol*, 165:3820–3829, 2000.

[176] T. Millat, E. Bullinger, J. Rohwer, O. Wolkenhauer. Approximations and their consequences for dynamic modelling of signal transduction pathways. *Math Biosci*, Aug 2006.

[177] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298:824–828, 2002.

[178] R. Milo, *et al.* Superfamilies of evolved and designed networks. *Science*, 303(5663): 1538–1542, Mar 2004.

[179] A. Minden, A. Lin, F. X. Claret, A. Abo, M. Karin. Selective activation of the JNK signaling cascade and c-Jun transcriptional activity by the small GTPases Rac and Cdc42Hs. *Cell*, 81:1147–1157, 1995.

[180] A. Mora, D. Komander, D. M. F. van Aalten, D. R. Alessi. PDK1, the master regulator of AGC kinase signal transduction. *Semin Cell Dev Biol*, 15:161–170, 2004.

[181] L. Murphy, S. Smith, R. Chen, J. Fingar, D.C. amd Blenis. Molecular interpretation of erk signal duration by immediate early gene products. *Nature Cell Biology*, 4: 556–567, 2002.

[182] J. R. S. Newman, *et al.* Single-cell proteomic analysis of s. cerevisiae reveals the architecture of biological noise. *Nature*, 441(7095):840–846, Jun 2006.

[183] M. E. J. Newman. Modularity and community structure in networks. *Proc Natl Acad Sci U S A*, 103(23):8577–8582, Jun 2006.

[184] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74:036104, 2006.

[185] M. E. J. Newman, M. Girvan. Finding and evaluating community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys*, 69(2 Pt 2):026113, Feb 2004.

[186] D. Y. Noh, *et al.* Expression of phospholipase C-gamma 1 and its transcriptional regulators in breast cancer tissues. *Anticancer Res*, 18:2643–2648, 1998.

[187] K. Oda, Y. Matsuoka, A. Funahashi, H. Kitano. A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol Syst Biol*, 1:2005.0010–2005.0010, 2005.

[188] K. Okkenhaug, A. Bilancio, J. L. Emery, B. Vanhaesebroeck. Phosphoinositide 3-kinase in T cell activation and survival. *Biochem Soc Trans*, 32:332–335, 2004.

[189] K. Okkenhaug, B. Vanhaesebroeck. PI3K in lymphocyte development, differentiation and activation. *Nat Rev Immunol*, 3:317–330, 2003.

[190] E. Ortega. How do multichain immune recognition receptors signal? a structural hypothesis. *Mol Immunol*, 32(13):941–945, Sep 1995.

[191] F. Ortega, J. L. Garcs, F. Mas, B. N. Kholodenko, M. Cascante. Bistability from double phosphorylation in signal transduction. kinetic and structural requirements. *FEBS J*, 273(17):3915–3926, Sep 2006.

[192] E. M. Ozbudak, M. Thattai, H. N. Lim, B. I. Shraiman, A. V. Oudenaarden. Multistability in the lactose utilization network of escherichia coli. *Nature*, 427(6976): 737–740, Feb 2004.

[193] E. H. Palacios, A. Weiss. Function of the src-family kinases, lck and fyn, in T-cell development and activation. *Oncogene*, 23(48):7990–8000, October 2004.

[194] F. Pan, A. R. Means, J. O. Liu. Calmodulin-dependent protein kinase IV regulates nuclear export of Cabin1 during T-cell activation. *EMBO J*, 24:2104–2113, 2005.

[195] J. A. Papin, B. O. Palsson. The jak-stat signaling network in the human b-cell: an extreme signaling pathway analysis. *Biophys J*, 87(1):37–46, Jul 2004.

[196] J. A. Papin, B. O. Palsson. Topological analysis of mass-balanced signaling networks: a framework to obtain network properties including crosstalk. *J Theor Biol*, 227(2): 283–297, Mar 2004.

[197] J. A. Papin, J. L. Reed, B. O. Palsson. Hierarchical thinking in network biology: the unbiased modularization of biochemical networks. *Trends Biochem Sci*, 29(12): 641–647, Dec 2004.

[198] J. A. Papin, T. Hunter, B. O. Palsson, S. Subramaniam. Reconstruction of cellular signalling networks and analysis of their properties. *Nat Rev Mol Cell Biol*, 6(2): 99–111, Feb 2005.

[199] T. Pawson, P. Nash. Assembly of cell regulatory systems through protein interaction domains. *Science*, 300(5618):445–452, 2003.

[200] J. M. Pedraza, A. van Oudenaarden. Noise propagation in gene networks. *Science*, 307(5717):1965–1969, Mar 2005.

[201] N. D. Price, J. L. Reed, B. Palsson. Genome-scale models of microbial cells: Evaluating the consequences of constraints. *Nat Rev Microbiol*, 2:886–897, 2004.

[202] N. Pullen, *et al.* Phosphorylation and activation of p70s6k by PDK1. *Science*, 279: 707–710, 1998.

[203] S. Qi, J. Groves, A. Chakraborty. Synaptic pattern formation during cellular recognition. *PNAS*, 98(12):6548–6553, 2001.

[204] D. Qian, A. Weiss. T cell antigen receptor signal transduction. *Curr Opin Cell Biol*, 9(2):205–212, Apr 1997.

[205] X. Qu, K. Kawauchi-Kamata, S. M. S. Miah, T. Hatani, H. Yamamura, K. Sada. Tyrosine phosphorylation of adaptor protein 3BP2 induces T cell receptor-mediated activation of transcription factor. *Biochemistry*, 44:3891–3898, 2005.

[206] L. E. Rameh, L. C. Cantley. The role of phosphoinositide 3-kinase lipid products in cell function. *J Biol Chem*, 274(13):8347–8350, March 1999.

[207] N. Rao, M. L. Lupher, S. Ota, K. A. Reedquist, B. J. Druker, H. Band. The linker phosphorylation site tyr292 mediates the negative regulatory effect of cbl on zap-70 in t cells. *J Immunol*, 164(9):4616–4626, May 2000.

[208] J. M. Raser, E. K. O'Shea. Control of stochasticity in eukaryotic gene expression. *Science*, 304(5678):1811–1814, Jun 2004.

[209] J. M. Raser, E. K. O'Shea. Noise in gene expression: origins, consequences, and control. *Science*, 309(5743):2010–2013, Sep 2005.

[210] A. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, A.-L. Barabasi. Hierarchical organization of modularity in metabolic networks. *Science*, 297:1551–1555, 2002.

[211] I. Rechenberg. *Synergistics - From micorscopic to macroscopic order*, chapter The evolution strategy - a mathematical model of Darwinian evolution, pages 122–132. Springer, Berlin-Heidelberg, 1984.

[212] D. Reichmann, O. Rahat, S. Albeck, R. Meged, O. Dym, G. Schreiber. The modular architecture of protein-protein binding interfaces. *Proc Natl Acad Sci U S A*, 102(1): 57–62, Jan 2005.

[213] B. L. Rellahan, *et al.* A dynamic constitutive and inducible binding of c-cbl by plc[gamma]1 sh3 and sh2 domains (negatively) regulates antigen receptor-induced plc[gamma]1 activation in lymphocytes. *Experimental Cell Research*, 289(1):184–194, September 2003.

[214] M. Reth, T. Brummer. Feedback regulation of lymphocyte signalling. *Nat Rev Immunol*, 4(4):269–77, Apr 2004.

[215] A. Rettinger. Parameteroptimierung mittels evolutionsstrategien. Master's thesis, University of Stuttgart, 1005.

[216] A. Rives, T. Galitski. Modular organization of cellular networks. *Proc. Natl. Acad. Sci. U.S.A.*, 100:1128–1133, 2003.

[217] C. Rodriguez-Caso, M. A. Medina, R. V. Sol. Topology, tinkering and evolution of the human transcription factor network. *FEBS J*, 272(24):6423–6434, Dec 2005.

[218] J. P. Roose, M. Mollenauer, V. A. Gupta, J. Stone, A. Weiss. A diacylglycerol-protein kinase c-rasgrp1 pathway directs ras activation upon antigen receptor stimulation of t cells. *Mol. Cell. Biol.*, 25(11):4426–4441, June 2005.

[219] S. Ryeom, R. J. Greenwald, A. H. Sharpe, F. McKeon. The threshold pattern of calcineurin-dependent gene expression is altered by loss of the endogenous inhibitor calcipressin. *Nat Immunol*, 4:874–881, 2003.

[220] J. Saez-Rodriguez, A. Kremling, H. Conzelmann, K. Bettenbrock, E. D. Gilles. Modular analysis of signal transduction networks. *IEEE Contr. Syst. Mag.*, 24(4):35–52, 2004.

[221] J. Saez-Rodriguez, H. Conzelmann, T. Sauter, B. Kholodenko, E. D. Gilles. Domain-oriented and modular approaches to the reduction of mathematical models of signaling networks. In U. Kummer, J. Pahle, I. Surovtsova, J. Zobeley, editors, *4th Workshop on Computation of Biochemical Pathways and Genetic Networks*, pages 13–20. Logos-Verlag Berlin, September 2005.

[222] J. Saez-Rodriguez, A. Kremling, E. D. Gilles. Dissecting the puzzle of life: Modularization of signal transduction networks. *Comput. Chem. Eng.*, 29(3):619–629, 2005.

[223] J. Saez-Rodriguez, S. Mirschel, R. Hemenway, S. Klamt, E. D. Gilles, M. Ginkel. Visual set-up of logical models of signaling and regulatory networks with ProMoT. *BMC Bioinformatics*, 7:506, 2006.

[224] J. Saez-Rodriguez, *et al.* A comprehensive logical model predicts key events in the T-cell signaling network. *submitted*, 2006.

[225] J. M. Salvador, P. R. Mittelstadt, G. I. Belova, A. J. Fornace, J. D. Ashwell. The autoimmune suppressor gadd45 inhibits the t cell alternative p38 activation pathway. *Nature Immunology*, 6(4):396–402, February 2005.

[226] S. Sasagawa, Y.-i. Ozaki, K. Fujita, S. Kuroda. Prediction and validation of the distinct dynamics of transient and sustained ERK activation. *Nat Cell Biol*, 7(4):365–373, Apr 2005.

[227] H. M. Sauro, B. Ingalls. Conservation analysis in biochemical networks: computational issues for software writers. *Biophys Chem*, 109(1):1–15, Apr 2004.

[228] E. M. Schaeffer, *et al.* Requirement for Tec kinases Rlk and Itk in T cell receptor signaling and immunity. *Science*, 284:638–641, 1999.

[229] H. Schaeffer, M. Weber. Mitogen-activated protein kinases: Specific messages from ubiquitous messengers. *Mol. and Cell. Biol.*, 19(4):2435–2444, April 1999.

[230] C. H. Schilling, D. Letscher, B. O. Palsson. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol*, 203(3):229–248, Apr 2000.

[231] J. Schlessinger. Cell signaling by receptor tyrosine kinases. *Cell*, 103(2):211–225, 2000.

[232] H. Schmidt, M. Jirstrand. Systems biology toolbox for matlab: a computational platform for research in systems biology. *Bioinformatics*, 22(4):514–515, February 2006.

[233] B. Schoeberl. *Mathematical Modeling of Signal Transduction Pathways in Mammalian cells at the Example of the EGF induced MAP Kinase Cascade and TNF Receptor Crosstalk*. PhD thesis, University of Stuttgart, 2003.

[234] B. Schoeberl, C. Eichler-Jonsson, E. Gilles, G. Müller. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat. Biotechnol.*, 20(4):370–375, 2002.

[235] I. H. Segel. *Enzyme kinetics. Benavior and analysis of rapid equilibrium and steady-State enzyme systems*. John Wiley and Sons, 1993.

[236] L. A. Segel. On the validity of the steady state assumption of enzyme kinetics. *Bull. Math. Biol.*, 50(6):579–93, 1988.

[237] A. Sensse. *Convex and toric geometry to analyze complex dynamics in chemical reaction systems*. PhD thesis, PhD Thesis, University of Magdeburg, 2005.

[238] X. Shan, *et al.* Deficiency of pten in jurkat t cells causes constitutive localization of itk to the plasma membrane and hyperresponsiveness to CD3 stimulation. *Mol. Cell. Biol.*, 20(18):6945–6957, September 2000.

[239] S. Shen-Orr, R. Milo, S. Mangan, U. Alon. Network motifs in the transcriptional regulation network of escherichia coli. *Nature Genetic*, 31:64–68, 2002.

[240] A. Sigal, *et al.* Variability and memory of protein levels in human cells. *Nature*, Nov 2006.

[241] E. Sontag. Some new directions in control theory inspired by systems biology. *IEE Proc. Systems Biology*, 1:9–18, 2004.

[242] E. Sontag. Molecular systems biology and control. *European Journal of Control*, 11: 396–435, 2005.

[243] E. D. Sontag. *Mathematical control theory*. Springer, 2 edition, 1998.

[244] C. Soule. Graphic requirements for multistationarity. *ComplexUs*, 1:123–133, 2003.

[245] I. Stefanova, B. Hemmer, M. Vergelli, R. Martin, W. Biddison, R. Germain. TCR ligand discrimination is enforced by competing erk positive and shp-1 negative feedback pathways. *Nature Immunology*, 4:248–254, 2003.

[246] I. Stefanova, J. R. Dorfman, R. N. Germain. Self-recognition promotes the foreign antigen sensitivity of naive t lymphocytes. *Nature*, 420(6914):429–434, November 2002.

[247] J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, E. D. Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420:190–193, 2002.

[248] J. Stelling, A. Kremling, M. Ginkel, K. Bettenbrock, E. Gilles. Towards a virtual biological laboratory. in: Foundations of systems biology,, 2002.

[249] J. Stelling. *System analysis of robustness in cellular networks*. PhD thesis, University of Stuttgart, 2004.

[250] J. Stelling. Mathematical models in microbial systems biology. *Curr Opin Microbiol*, 7(5):513–518, Oct 2004.

[251] J. Stelling, U. Sauer, Z. Szallasi, F. J. r. Doyle, J. Doyle. Robustness of cellular functions. *Cell*, 118(6):675–685, Sep 2004.

[252] A. M. Stock, V. L. Robinson, P. N. Goudreau. Two-component signal transduction. *Annu Rev Biochem*, 69:183–215, 2000.

[253] B. Su, E. Jacinto, M. Hibi, T. Kallunki, M. Karin, Y. Ben-Neriah. JNK is involved in signal integration during costimulation of T lymphocytes. *Cell*, 77:727–736, 1994.

[254] K. Sugie, M. S. Jeon, H. M. Grey. Activation of nave cd4 t cells by anti-CD3 reveals an important role for fyn in lck-mediated signaling. *Proc Natl Acad Sci U S A*, 101 (41):14859–14864, October 2004.

[255] H. Teramoto, O. A. Coso, H. Miyata, T. Igishi, T. Miki, J. S. Gutkind. Signaling from the small GTP-binding proteins Rac1 and Cdc42 to the c-Jun N-terminal kinase/stress-activated protein kinase pathway. A role for mixed lineage kinase 3/protein-tyrosine kinase 1, a novel member of the mixed lineage kinase family. *J Biol Chem*, 271: 27225–27228, 1996.

[256] R. Thomas, R. D'Ari. *Biological Feedback*. Boca Raton, CRC Press, 1990.

[257] R. Thomas, M. Kaufman. Multistationarity, the basis of cell differentiation and memory. i. structural conditions of multistationarity and other nontrivial behavior. *Chaos*, 11(1):170–179, Mar 2001.

[258] M. Thome. Carma1, bcl-10 and malt1 in lymphocyte development and activation. *Nat Rev Immunol*, 4(5):348–359, May 2004.

[259] L. A. Tibbles, *et al.* MLK-3 activates the SAPK/JNK and p38/RK pathways via SEK1 and MKK3/6. *EMBO J*, 15:7026–7035, 1996.

[260] M. Togni, *et al.* The role of adaptor proteins in lymphocyte activation. *Mol Immunol*, 41(6-7):615–630, July 2004.

[261] M. K. Topham. Signaling roles of diacylglycerol kinases. *J Cell Biochem*, 97:474–484, 2006.

[262] K. M. Torgersen, *et al.* Release from tonic inhibition of t cell activation through transient displacement of c-terminal src kinase (csk) from lipid rafts. *J Biol Chem*, 276(31):29313–29318, August 2001.

[263] A. Trautmann, C. Randriamampita. Initiation of TCR signalling revisited. *Trends Immunol*, 24(8):425–428, Aug 2003.

[264] V. L. J. Tybulewicz. Vav-family proteins in T-cell signalling. *Curr Opin Immunol*, 17:267–274, 2005.

[265] A. G. Uren, *et al.* Identification of paracaspases and metacaspases: Two ancient families of caspase-like proteins, one of which plays a key role in malt lymphoma. *Molecular Cell*, 6(4):961–967, October 2000.

[266] S. Valitutti, A. Lanzavecchia. Serial triggering of TCR s: a basis for the sensitivity and specificity of antigen recognition. *Immunol. today*, 18(6):299–304, 1997.

[267] M. Villalba, *et al.* Translocation of pkc[theta] in t cells is mediated by a nonconventional, pi3-k- and vav-dependent pathway, but does not absolutely require phospholipase c. *J Cell Biol*, 157(2):253–263, April 2002.

[268] M. von Willebrand, G. Baier, C. Couture, P. Burn, T. Mustelin. Activation of phosphatidylinositol-3-kinase in Jurkat T cells depends on the presence of the p56lck tyrosine kinase. *Eur J Immunol*, 24:234–238, 1994.

[269] R. Weil, A. Israel. Deciphering the pathway from the TCR to NF-kappaB. *Cell Death Differ*, 13:826–833, 2006.

[270] A. Wells. EGF receptor. *Int. J. Biochem. Cell Biol.*, 31(6):637–43, 1999.

[271] G. Werlen, B. Hausmann, D. Naeher, E. Palmer. Signaling Life and Death in the Thymus: Timing Is Everything. *Science*, 299(5614):1859–1863, 2003.

[272] Wikipedia. Guanosine triphosphate — wikipedia, the free encyclopedia, 2006. [Online; accessed 18-October-2006].

[273] Wikipedia. Nullcline — wikipedia, the free encyclopedia, 2006. [Online; accessed 18-December-2006].

[274] H. S. Wiley. Trafficking of the erbb receptors and its influence on signaling. *Exp Cell Res*, 284(1):78–88, Mar 2003.

[275] H. Wiley, S. Shvartsman, D. Lauffenburger. Computational modeling of the EGF-receptor system: a paradigm for systems biology. *Trends Cell. Biol.*, 13(1):43–50, 2003.

[276] D. Wolf, A. Arkin. Motifs, modules and games in bacteria. *Curr Opin Microbiol.*, 6 (2):125–134, 2003.

[277] W. Xiong, J. E. J. Ferrell. A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature*, 426(6965):460–465, 2003.

[278] X.-N. Xu, G. R. Screaton. MHC/peptide tetramer-based studies of T cell function. *J. Immun. methods*, 1(1):21–28, 2002.

[279] S. Yamasaki, *et al.* Docking protein Gab2 is phosphorylated by ZAP-70 and negatively regulates T cell receptor signaling by recruitment of inhibitory molecules. *J Biol Chem*, 276:45175–45183, 2001.

[280] S. Yamasaki, *et al.* Gads/Grb2-mediated association with LAT is critical for the inhibitory function of Gab2 in T cells. *Mol Cell Biol*, 23:2515–2529, 2003.

[281] M. Yan, *et al.* Activation of stress-activated protein kinase by MEKK1 phosphorylation of its activator SEK1. *Nature*, 372:798–800, 1994.

[282] E. Yang, J. Zha, J. Jockel, L. H. Boise, C. B. Thompson, S. J. Korsmeyer. Bad, a heterodimeric partner for Bcl-XL and Bcl-2, displaces Bax and promotes cell death. *Cell*, 80:285–291, 1995.

[283] Y. Yarden. The EGFR family and its ligands in human cancer:signalling mechanisms and therapeutic opportunities. *Eur. J. Cancer.*, 37:S3–S8, 2001.

[284] S. Zakaria, *et al.* Differential regulation of TCR-mediated gene transcription by Vav family members. *J Exp Med*, 199:429–434, 2004.

[285] R. Zamoyska, A. Basson, A. Filby, G. Legname, M. Lovatt, B. Seddon. The influence of the src-family kinases, Lck and Fyn, on T cell differentiation, survival and activation. *Immunol Rev*, 191:107–118, 2003.

[286] I. Zevedei-Oancea, S. Schuster. A theoretical framework for detecting signal transfer routes in signalling networks. *Computers & Chemical Engineering*, 29:597–617, 2005.

[287] J. Zha, H. Harada, K. Osipov, J. Jockel, G. Waksman, S. J. Korsmeyer. BH3 domain of BAD is required for heterodimerization with BCL-XL and pro-apoptotic activity. *J Biol Chem*, 272:24101–24104, 1997.

[288] J. Zhang, *et al.* Cutting edge: regulation of t cell activation threshold by CD28 costimulation through targeting cbl-b for ubiquitination. *J Immunol*, 169(5):2236–2240, September 2002.

[289] P. A. Zipfel, W. Zhang, M. Quiroz, A. M. Pendergast. Requirement for abl kinases in t cell receptor signaling. *Current Biology*, 14(14):1222–1231, July 2004.

[290] J. Zobeley, D. Lebiedz, J. Kammerer, A. Ishmurzin, U. Kummer. A new time-dependent complexity reduction method for biochemical systems. *Lecture Notes in Computer Science*, 3380:90, 2005.

Appendix
A.1 Textual Definition of the Motifs

A.1.1 Mass-action-law-based description

For the sake of completeness, a detailed description of the motifs defined in Chapter 4 and investigated in Chapter 5 is give here. First, the description of the motifs used in Section 5.1 will be presented. The enzymatic reactions in the motifs are described using mass-action-law kinetics. The resulting reactions are presented in the following tables.

C2s	C2n	C2p
$A + E_1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} AE_1 \overset{k_3}{\rightharpoonup} A^* + E_1$ $A^* + E_2 \underset{k_5}{\overset{k_4}{\rightleftharpoons}} A^*E_2 \overset{k_6}{\rightharpoonup} A + E_2$	$A + E_1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} AE_1 \overset{k_3}{\rightharpoonup} A^* + E_1$ $A^* + E_2 \underset{k_5}{\overset{k_4}{\rightleftharpoons}} A^*E_2 \overset{k_6}{\rightharpoonup} A + E_2$ $A^* + A^* \underset{k_8}{\overset{k_7}{\rightleftharpoons}} A^*A^* \overset{k_9}{\rightharpoonup} A + A^*$	$A + E_1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} AE_1 \overset{k_3}{\rightharpoonup} A^* + E_1$ $A^* + E_2 \underset{k_5}{\overset{k_4}{\rightleftharpoons}} A^*E_2 \overset{k_6}{\rightharpoonup} A + E_2$ $A + A^* \underset{k_8}{\overset{k_7}{\rightleftharpoons}} AA^* \overset{k_9}{\rightharpoonup} A^* + A^*$
B1	**C2B1**	**C2B2**
$A + B \underset{k_2}{\overset{k_1}{\rightleftharpoons}} AB$	$A + E_1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} AE_1 \overset{k_3}{\rightharpoonup} A_p + E_1$ $A_p + E_2 \underset{k_5}{\overset{k_4}{\rightleftharpoons}} A_pE_2 \overset{k_6}{\rightharpoonup} A + E_2$ $A_p + B \underset{k_8}{\overset{k_7}{\rightleftharpoons}} A_pB$	$A + E_1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} AE_1 \overset{k_3}{\rightharpoonup} A_p + E_1$ $A_p + E_2 \underset{k_5}{\overset{k_4}{\rightleftharpoons}} A_pE_2 \overset{k_6}{\rightharpoonup} A + E_2$ $A_p + B \underset{k_8}{\overset{k_7}{\rightleftharpoons}} A_pB$ $A_p + C \underset{k_{10}}{\overset{k_9}{\rightleftharpoons}} A_pC$
C3dd	**C3di**	**C3dp**
$A + E_1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} AE_1 \overset{k_3}{\rightharpoonup} A^* + E_1$ $A^* + E_2 \underset{k_5}{\overset{k_4}{\rightleftharpoons}} A^*E_2 \overset{k_6}{\rightharpoonup} A + E_2$ $A^* + E_1 \underset{k_8}{\overset{k_7}{\rightleftharpoons}} AE_1 \overset{k_9}{\rightharpoonup} A^{**} + E_1$ $A^{**} + E_2 \underset{k_{11}}{\overset{k_{10}}{\rightleftharpoons}} A^{**}E_2 \overset{k_{12}}{\rightharpoonup} A^* + E_2$	$A + E_1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} AE_1 \overset{k_3}{\rightharpoonup} A^* + E_1$ $A^* + E_2 \underset{k_5}{\overset{k_4}{\rightleftharpoons}} A^*E_2 \overset{k_6}{\rightharpoonup} A + E_2$ $A^* + E_1 \underset{k_8}{\overset{k_7}{\rightleftharpoons}} AE_1 \overset{k_9}{\rightharpoonup} A^{**} + E_1$ $A^{**} + E_3 \underset{k_{11}}{\overset{k_{10}}{\rightleftharpoons}} A^{**}E_3 \overset{k_{12}}{\rightharpoonup} A^* + E_3$	$A + E_1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} AE_1 \overset{k_3}{\rightharpoonup} A^* + E_1$ $A^{**} + E_2 \underset{k_5}{\overset{k_4}{\rightleftharpoons}} A^{**}E_2 \overset{k_6}{\rightharpoonup} A + E_2$ $A^* + E_1 \underset{k_8}{\overset{k_7}{\rightleftharpoons}} AE_4 \overset{k_9}{\rightharpoonup} A^{**} + E_1$
C3ii	**C3sr**	**C3sr(cont)**
$A + E_1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} AE_1 \overset{k_3}{\rightharpoonup} A^* + E_1$ $A^* + E_2 \underset{k_5}{\overset{k_4}{\rightleftharpoons}} A^*E_2 \overset{k_6}{\rightharpoonup} A + E_2$ $A^* + E_4 \underset{k_8}{\overset{k_7}{\rightleftharpoons}} AE_4 \overset{k_9}{\rightharpoonup} A^{**} + E_4$ $A^{**} + E_3 \underset{k_{11}}{\overset{k_{10}}{\rightleftharpoons}} A^{**}E_3 \overset{k_{12}}{\rightharpoonup} A^* + E_3$	$A + E_1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} AE_1 \overset{k_3}{\rightharpoonup} A^* + E_1$ $A^* + E_3 \underset{k_5}{\overset{k_4}{\rightleftharpoons}} A^*E_3 \overset{k_6}{\rightharpoonup} A + E_3$ $A^* + A^{**} \underset{k_8}{\overset{k_7}{\rightleftharpoons}} A^*A^{**} \overset{k_9}{\rightharpoonup} A^{**} + A^{**}$	$A^* + A^* \underset{k_8}{\overset{k_7}{\rightleftharpoons}} A^*A^* \overset{k_9}{\rightharpoonup} A^{**} + A^*$ $A^{**} + E_2 \underset{k_{11}}{\overset{k_{10}}{\rightleftharpoons}} A^{**}E_2 \overset{k_{12}}{\rightharpoonup} A^* + E_2$
C2C2	**C3C2**	**C3C2(cont)**
$A + E_1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} AE_1 \overset{k_3}{\rightharpoonup} A^* + E_1$ $A^* + B \underset{k_5}{\overset{k_4}{\rightleftharpoons}} A^*B \overset{k_6}{\rightharpoonup} A + B*$ $B^* + E_2 \underset{k_8}{\overset{k_7}{\rightleftharpoons}} B^*E_2 \overset{k_9}{\rightharpoonup} B + E_2$	$A + E_1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} AE_1 \overset{k_3}{\rightharpoonup} A * + E_1$ $A^{**} + B \underset{k_5}{\overset{k_4}{\rightleftharpoons}} A^{**}B \overset{k_6}{\rightharpoonup} A + B*$	$B^* + E_2 \underset{k_8}{\overset{k_7}{\rightleftharpoons}} B^*E_2 \overset{k_9}{\rightharpoonup} B + E_2$ $A^* + E_2 \underset{k_{11}}{\overset{k_{10}}{\rightleftharpoons}} A^*E_1 \overset{k_{12}}{\rightharpoonup} A^{**} + E_1$

A.1.2 Quasi-steady-state description

In Section 5.2, the description is simplified using the quasi-steady-state assumption, which leads to description of the reactions in the form of Michaelis-Menten kinetics[235] if there is no competition for enzymes; otherwise (e.g. for C3dd), a different expression arises[88]. Additionally, the conservation relationship for the total amount of concentration is considered. In Section 5.3, in addition, the equations are normalized with respect to the total concentration and one time constant, reducing thus the number of parameters. The simplifications were explained in Sections 5.2.2 and 5.3; here the resulting equations of the rest are summarized. For a detailed explanation the reader is referred to Gayer[88]. Note that the parameter names do not correspond to those of Section A.1.1.

C2s	C2n
$\dot{A^*} = \dfrac{E_1 \cdot (1 - A^*)}{K_{m1} + (1 - A^*)} - \dfrac{E_2 \cdot k_{-1} \cdot A^*}{K_{m-1} + A^*}$	$\dot{A^*} = \dfrac{E_1 \cdot (1 - A^*)}{K_{m1} + (1 - A^*)} - \dfrac{(E_2 + A^*) \cdot k_{-1} \cdot A^*}{K_{m-1} + A^*}$

C2p	B1 ‡‡
$\dot{A^*} = \dfrac{(E_1 + A^*) \cdot (1 - A^*)}{K_{m1} + (1 - A^*)} - \dfrac{E_2 \cdot k_{-1} \cdot A^*}{K_{m-1} + A^*}$	$\dot{AB} = (1 - AB) \cdot (1 - AB) - k_2 \cdot AB$

C2B1

$$\dot{A^*} = \frac{E_1 \cdot (1 - A^*)}{K_{m1} + (1 - A^*)} - \frac{E_2 \cdot k_{-1} \cdot A^*}{K_{m-1} + A^*} - k_3(1 - AB) \cdot A^* + k_{3r} \cdot AB$$
$$\dot{AB} = k_3 (A^* - AB)(1 - AB) - k_{3r} \cdot AB$$

C2B2

$$\dot{A^*} = \frac{E_1 \cdot (1 - A^*)}{K_{m1} + (1 - A^*)} - \frac{E_2 \cdot k_{-1} \cdot A^*}{K_{m-1} + A^*} - k_3(1 - AB) \cdot A^* + k_{3r} \cdot AB + k_4(1 - AC)A^* - k_{4r} \cdot AC$$
$$\dot{AB} = k_3 (A^* - AB - AC)(1 - AB) - k_{3r} \cdot AB, \quad \dot{AB} = k_4 (A^* - AB - AC)(1 - AB) - k_{4r} \cdot AC$$

C3dd

$$\dot{A} = -\frac{E_1 \cdot A}{K_{m1} + A + \frac{K_{m1}}{K_{m2}}(1 - A - A^{**})} + \frac{E_2 \cdot k_{-1} \cdot (1 - A - A^{**})}{K_{m-1} + (1 - A - A^{**}) + \frac{K_{m-1}}{K_{m-2}}A}$$
$$\dot{A^{**}} = -\frac{E_2 \cdot k_{-2} \cdot A^{**}}{K_{m-2} + A^{**} + \frac{K_{m-2}}{K_{m-1}}(1 - A - A^{**})} + \frac{k_2 E_1 \cdot (1 - A - A^{**})}{K_{m2} + (1 - A - A^{**}) + \frac{K_{m2}}{K_{m1}}A}$$

C3di

$$\dot{A} = -\frac{E_1 \cdot A}{K_{m1} + A + \frac{K_{m1}}{K_{m2}}(1 - A - A^{**})} + \frac{E_2 \cdot k_{-1} \cdot (1 - A - A^{**})}{K_{m-1} + (1 - A - A^{**})}$$
$$\dot{A^{**}} = -\frac{E_2 \cdot k_{-2} \cdot A^{**}}{K_{m-2} + A^{**}} + \frac{k_2 E_3 \cdot (1 - A - A^{**})}{K_{m2} + (1 - A - A^{**}) + \frac{K_{m2}}{K_{m1}}A}$$

C3dp

$$\dot{A} = -\frac{E_1 \cdot A}{K_{m1} + A + \frac{K_{m1}}{K_{m2}}(1 - A - A^{**})} + \frac{E_2 \cdot k_{-1} \cdot (A^{**})}{K_{m-1} + (A^{**})}$$
$$\dot{A^{**}} = -\frac{E_2 \cdot k_{-1} \cdot A^{**}}{K_{m-1} + A^{**}} + \frac{k_2 E_1 \cdot (1 - A - A^{**})}{K_{m2} + (1 - A - A^{**}) + \frac{K_{m2}}{K_{m1}}A)}$$

C3ii

$$\dot{A} = -\frac{E_1 \cdot A}{K_{m1} + A} + \frac{E_4 \cdot k_{-1} \cdot (1 - A - A^{**})}{K_{m-1} + (1 - A - A^{**})}, \qquad \dot{A^{**}} = -\frac{E_2 \cdot k_{-2} \cdot A^{**}}{K_{m-2} + A^{**}} + \frac{k_2 E_3 (1 - A - A^{**})}{K_{m2} + (1 - A - A^{**})}$$

C3sr

$$\dot{A} = -\frac{E_1 \cdot A}{K_{m1} + A} + \frac{E_3 \cdot k_{-1} \cdot (1 - A - A^{**})}{K_{m-1} + (1 - A - A^{**})}, \qquad \dot{A^{**}} = -\frac{E_2 \cdot k_{-2} \cdot A^{**}}{K_{m-2} + A^{**}} + \frac{(1 - A) \cdot k_2(1 - A - A^{**})}{K_{m2} + (1 - A - A^{**})}$$

C2C2

$$\dot{A^*} = \frac{E_1 \cdot (1 - A^*)}{K_{m1} + (1 - A^*)} - k_3 \cdot A^* B, \qquad \dot{B^*} = \frac{E_2 \cdot (1 - B^*)}{K_{m2} + (1 - B^*)} + k_4 \cdot A^* B$$

C3C2

$$\dot{A^*} = \frac{E_1 \cdot (1 - A^*)}{K_{m1} + (1 - A^*)} - k_3 \cdot A^* B, \qquad \dot{B^*} = \frac{E_2 \cdot (1 - B^*)}{K_{m2} + (1 - B^*)} + k_4 \cdot A^* B$$
$$\dot{A^{**}} = -\frac{E_2 \cdot k_{-2} \cdot A^{**}}{K_{m-2} + A^{**}} + \frac{(1 - A) \cdot k_2 E_1 \cdot (1 - A - A^{**})}{K_{m2} + (1 - A - A^{**})}$$

A.2 Histograms for the Distribution of Δ_{A_T}

This histogram plot the distribution of the parameters Δ_{A_T} , $\Delta_{E_{1T}}$, and $\overline{\Delta}_{E_{1T}}$ as defined in Equations 5.1, 5.2, and 5.3, respectively, for the motifs showing multistationarity. See Section 5.1 for more details.

| (a) C2p | (b) C3dd | (c) C3di | (d) C3sr |
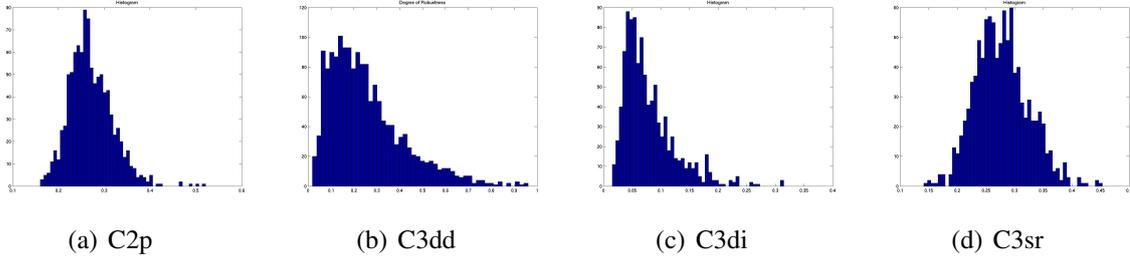
Figure A.1: Distribution of Δ_{A_T} for C2p, C3dd, C3di, and C3sr.

| (a) C2p | (b) C3dd | (c) C3di | (d) C3sr |
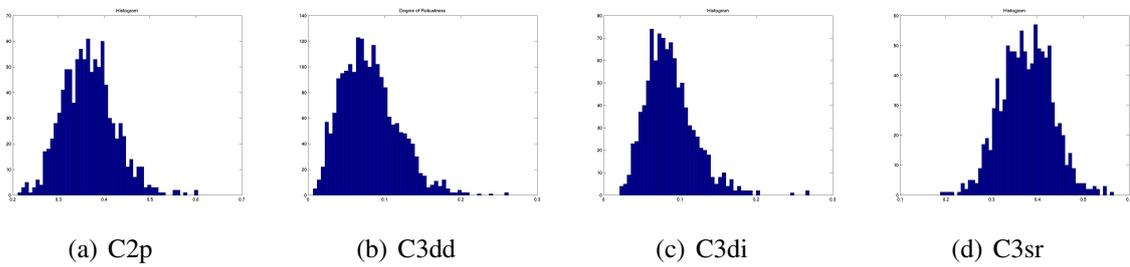
Figure A.2: Distribution of $\Delta_{E_{2T}}$ for C2p, C3dd, C3di, and C3sr.

| (a) C2p | (b) C3dd | (c) C3di |
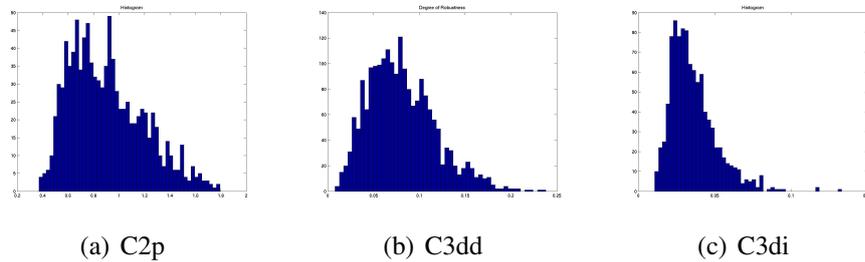
Figure A.3: Distribution of $\Delta_{E_{1T}}$ for C2b and C3a, and C3di.

A.3 Proof of Monotone Characteristic Curve of the Motif C3dp

Let us exemplify the fact that the non-monotone motifs of Chapter 5 have a monotone characteristic curve with the motif C3dp (see Figure 4.5(e)). Note that a similar analysis was performed in Section 7.1.2. Its differential equations (see Section A.1.2) are of the form of Equation 3.2, $\dot{\vec{c}} = f(\vec{c}, \vec{u}, \vec{p})$. For the sake of compactness, we shall rename A^* and A^{**} c_1 and c_2, respectively. Since c depends on u, in the steady state it holds

$$0 = f(\vec{c}_{ss}, u) = f(\vec{c}_{ss}(u), u). \tag{A.30}$$

Deriving with respect to u and applying the chain rule, since $\vec{c}_{ss}$ depends on u, we obtain

$$0 = \frac{\partial f(\vec{c}^{ss}(u), u)}{\partial c} \frac{\partial c}{\partial u} + \frac{\partial f(\vec{c}^{ss}(u), u)}{\partial u} = J \cdot c_u + f_u \rightarrow c_u = -J^{-1} f_u \tag{A.31}$$

with $J = \frac{\partial f}{\partial c}$, $c_u = \frac{\partial c}{\partial u}$, and $f_u = \frac{\partial f}{\partial u}$. Since

$$J^{-1} = \frac{-1}{J_{22}J_{12} - J_{21}J_{11}} \begin{pmatrix} J_{22} & -J_{12} \\ -J_{21} & J_{11} \end{pmatrix} = \frac{-1}{(-)\cdot(-) - (+)\cdot(-)} \begin{pmatrix} - & -(-) \\ -(+) & - \end{pmatrix} = \begin{pmatrix} - & + \\ - & - \end{pmatrix},$$

it follows from Equation A.31

$$c_u = \begin{pmatrix} \frac{\partial c_1^{ss}}{\partial u} \\ \frac{\partial c_2^{ss}}{\partial u} \end{pmatrix} = -\begin{pmatrix} - & + \\ - & - \end{pmatrix} \begin{pmatrix} + \\ 0 \end{pmatrix} = \begin{pmatrix} + \\ + \end{pmatrix}, \tag{A.32}$$

Which means that both characteristic curves $c_1^{ss}(u)$ and $c_1^{ss}(u)$ are monotone. Similar arguments can be applied to C2C2 and C3C2.

A.4 ProMoT Code of the Domain-oriented Library

Here, the architecture behind the ProMoT library for a domain-oriented modeling presented in Section 4.3 will be succinctly described.

New storages, based on the previous multi-storages of the ProMoT libraries were developed. The simplest unit is a module describing a domain with 2 possible states (see Fig. A.4), which we shall use to illustrate the implementation. There is one differential equation for the balance of the percentage of the domain in one of the states (e.g. degree of occupation), `st_2` ; the other state `st_1` is computed algebraically (`1-st_2`). This allows a minimal realization in terms of differential equations.

```
                       (define-module
                        :class "storage-mult-2st"
                        :super-classes ("multistage-storage")
                        :properties (("abstract"    :value "no"))  :icon "icons/1bs-2st"
                        :terminals
                         (("cot" :is-a "term-concentration-in":geometry-side "top" :geometry-position "0.5")
                          ("cf1"  :is-a "term-storage-flux"   :geometry-side "right" :geometry-position "0.25"
                                      :variables        (("c"             :is-eq-to "parent.c_1")))
                          ("cf2"  :is-a "term-storage-flux"   :geometry-side "right" :geometry-position "0.75"
                                      :variables    (("c"               :is-eq-to "parent.c_2"))))
                        :variables
                         (("c_1"  :is-a "var-state-concentration-out"  :report "all")
                          ("c_2"  :is-a "var-state-concentration-out"  :report "all")
                          ("st_2"  :is-a "var-state-fraction" )

                        :equations
                         (("bal_st2" :relation ":diff<t> st_2 == cf2.r/cot.c")
                          ("out1"     :relation "c_1 == (1-st_2)*cot.c")
                          ("out2"     :relation "c_2 == st_2*cot.c")))
```

Figure A.4: Implementation in ProMoT of a domain with 2 possible states. The figure represent an scheme of the module, and the accompanying text the corresponding mdl code.

The module computes the standard concentrations as a product of the state and the total concentration (`cot.c`, see Figure A.4), and sends them via the terminals to the environment. Analogously, the modules receives 'normal' reaction rates, which are transformed into the variables required for the balance of `st_2`. Thus, these modules are compatible with the rest of elements of the library. The same concept applies to domains with 3, 4, etc. states.

Different domains can be combined to give rise to a molecule with different binding sites (see Figure A.5). The key point is that all the domains have a common total concentration.

A more complex situation arises when the motifs are *not* independent. In that case, an exact, full reduction is not possible[48]. However, still a certain reduction can be achieved and the resulting modules are hierarchically structured[48;221], what facilitates an implementation in ProMoT. We shall illustrate it here using the most common case, where one domain D_1 of a molecule M binds an adaptor A, influencing the phosphorylation at another domain D_2. This case corresponds to Figure 4.1(d), where the third domain can be decoupled -and thus ignore for this analysis- since it is independent. The microstates are thus governed by the reactions

$$M_{00} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} M_{A0} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} M_{AP} \tag{A.33}$$

$$M_{00} \underset{k_{-2}}{\overset{k_2^*}{\rightleftharpoons}} M_{0P} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} M_{AP} \tag{A.34}$$

We shall consider, for the sake of simplicity, first order kinetics and an enzyme E responsible for the phosphorylation of D_2. The influence of D_1 on D_2 is encoded by the different activity for the phosphorylation (k_2 and ($k_2^* = k_2(1 + \alpha)$). There are 4 microstates (M_{00}, M_{A0}, M_{0P}, and M_{AP}) and 1 conservation relationship. Thus, there are 3 independent ODEs. Defining $M_{AT} = M_{A0} + M_{AP}$ and

```
(define-module
  :class "storage-mult-2bs-23st"
  :super-classes ("storage-mult-2bs")
  :properties (("abstract"   :value "no"))
  :icon "icons/2bs-23st"  :geometry-width "258" :geometry-height "306"

  :terminals
  (("cf11" :is-eq-to "bs1.cf1"  :geometry-side "right" :geometry-position "0.18")
   ("cf12" :is-eq-to "bs1.cf2"  :geometry-side "right" :geometry-position "0.32")
   ("cf21" :is-eq-to "bs2.cf1"  :geometry-side "right" :geometry-position "0.62")
   ("cf22" :is-eq-to "bs2.cf2"  :geometry-side "right" :geometry-position "0.75")
   ("cf23" :is-eq-to "bs2.cf3"  :geometry-side "right" :geometry-position "0.88")
   ("cot"  :is-eq-to "adapter_conc.in"
                                :geometry-side "top"     :geometry-position "0.47"))

  :modules
  (("adapter_conc" :is-a "adapter-conc"  :geometry-x "59"  :geometry-y "59")
   ("bs1"          :is-a "storage-mult-2st" :geometry-x "104" :geometry-y "99")
   ("bs2"          :is-a "storage-mult-3st" :geometry-x "108" :geometry-y "221"))

  :links
  (("adapter_concout_bs1cot" :terminals("adapter_conc.out""bs1.cot""bs2.cot"))))
```

Figure A.5: Implementation in ProMoT of a molecule with 2 domains, with 2 and 3 possible states, respectively. The figure represent an scheme of the module, and the accompanying text the corresponding mdl code.

$M_{PT} = M_{0P} + M_{AP}$, and considering the total amount of M $M_T = M_{00} + M_{0P} + M_{A0} + M_{AP}$ constant, one obtains

$$\frac{dM_{AT}}{dt}(M_{AT}) = \frac{dM_{A0}}{dt} + \frac{M_{AP}}{dt} = k_1 \cdot A \cdot M_T - k_{-1} \cdot A_{AT} \tag{A.35}$$

$$\frac{dM_{PT}}{dt}(M_{PT}, M_{AP}, M_{AT}) = \frac{M_{A0}}{dt} + \frac{M_{AP}}{dt} = k_2 \cdot E \cdot (M_T - M_{PT}) - k_{-2} \cdot A_{AT} + \alpha \cdot (M_{AT} - M_{AP}) \tag{A.36}$$

For α=0, $\dot{M}_{PT}(M_{PT}) = k_2 \cdot E \cdot (M_T - M_{PT}) - k_{-2} \cdot A_{AT}$ and one can thus model the domains separately. If $\alpha \neq 0$, however, an additional state A_{BP} must be considered. This is included in the module describing the reactions of the dependent domain D_2, see Figure A.6.
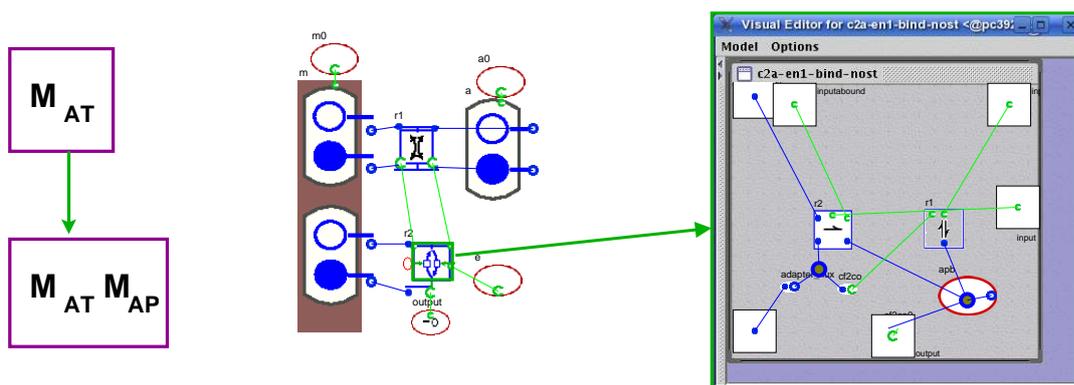
Figure A.6: Implementation in ProMoT of a case of two non-independent domains. The left figure schematically shows the structure of the ODE system, and the right side the actual implementation. The domain D_1 is modeled as a normal B1 module (see Sections 4.2 and 4.3), and D_2 as a modified version of C2s with an additional state.

A.5 Experimental Validation of the Predictions of the Logical Model of T-cell Activation

The in vitro experiments performed to test the simulation results of th logical model describing the TCR-induced Signaling Network (Section 6.3) are depicted here[224]. They were performed by the colleagues at the Institute of Immunology, University of Magdeburg.
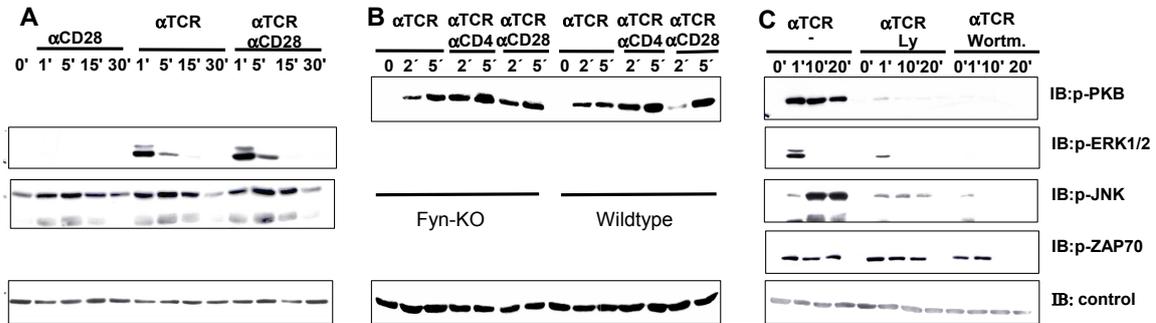
Figure A.7: In vitro analysis of model predictions. A, Activation of ERK and JNK upon CD28, TCR (CD3) or TCR+CD28 stimulation in mouse splenic T cells. **B**, Activation of ERK and PKB upon TCR, TCR+CD4 and TCR+CD28 stimulation in Fyn-deficient and heterozygous splenic mouse T cells. **C**, Inhibition of PI3K with both Ly 294002 and Wortmannin blocks the phosphorylation of PKB, ERK and JNK but not ZAP-70 in human T cells. As a control, total concentration of ZAP70 (A) or β-actin (B,C) was used.
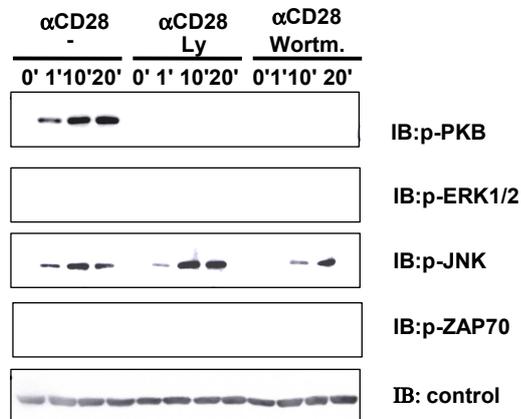
Figure A.8: In vitro analysis of PI3K inhibitors on CD28 signaling. Inhibition of PI3K with both Ly 294002 and Wortmannin blocks the phosphorylation of PKB, but not of JNK in human T-cells upon CD28 activation. β-actin was included as the loading control.
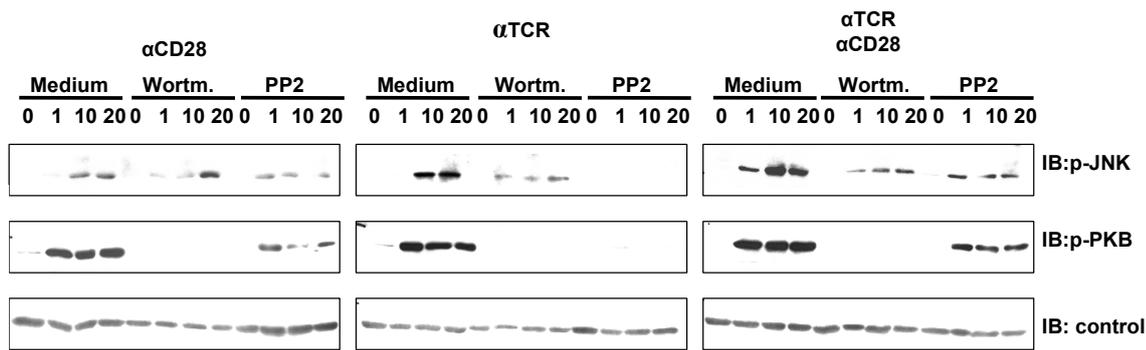
Figure A.9: In vitro analysis of Src kinase inhibition. Inhibition of Src-Kinases (Lck and Fyn) with PP2 blocks TCR-induced but not CD28-induced PKB and JNK activation in human T-cells, therefore we concluded that CD28 signaling is not strictly Src kinase dependent. The effect was compared with PI3K inhibition via Wortmannin (ccf. Figure A.8. 2C and Figure A.8), which blocks the phosphorylation of PKB but not of JNK. β-actin was included as the loading control.

Materials and Methods Immunoblotting. Human or mouse T-cells were purified using an AutoMACS magnetic isolation system according to the manufacturer's instructions (Miltenyi). Mouse T-cells were stimulated with 10 μg/ml of biotinylated CD3ε (a subunit of the TCR) antibody (145-2C11, BD Biosciences), 10 μg/ml of biotinylated CD28 antibody (37.51, BD Biosciences), CD3 plus CD28 mAbs or with CD3 plus 10 μg/ml of biotinylated CD4 (GK1.5, BD Biosciences) followed by crosslinking with 25 μg/ml of streptavidin (Dianova) at 37°C for the indicated periods of time. Human T-cells were stimulated with CD3ε mAb MEM92 (IgM, kindly provided by Dr. V. Horejsi, Prague, Czech Republic) or with CD3 plus CD28 mAbs (248.23.2). Cells were lysed in buffer containing 1% NP-40, 1% laurylmaltoside (N-dodecyl β-D-maltoside), 50 mM Tris pH 7.5, 140 mM NaCl, 10mM EDTA, 10 mM NaF, 1 mM PMSF, 1 mM Na3VO4. Proteins were separated by SDS/PAGE, transferred onto membranes, and blotted with the following antibodies: anti-phosphotyrosine (4G10), anti-ERK1/2 (pT202/pT204), anti-JNK (pT183/pY185), anti-phospho-Akt (S473)(all from Cell Signaling), anti-ZAP70 (pTyr 319, Cell Signaling), anti-ZAP70 (cloneZ24820, Transduction Laboratories) or against β-Actin (Sigma). Where PI3K and src-kinase inhibitors were used, T-cells were treated with 100nM Wortmannin (Calbiochem) or 10μM PP2 (Calbiochem) for 30 min at 37°C prior to stimulation.

A.6 Detailed Description of the Logical Model of T-cell Activation

The following tables provide a comprenhensive documentation of the logical model of T-cell activation described and analyzed in Section 6.

Table A.1: List of compounds in the logical T-cell model. Model name corresponds to the name in Fig. 6.5 and the Table A.2. Common abbreviations are those usually used in the literature, while name is the whole name. Type classifies the molecules, if applies, as follows: K =Kinase, T=Transcription Factor, P = Phosphatase, A =Adaptor Protein, R = Receptor, G= GTP-ase. In the case where two pools of a molecule were considered, a'reservoir' was included which was required for both pools. This allows to perform a simultaneous knock-out of both pools.

Nr	Model Name	Common abbreviation	Name	Type
1	abl	Abl	Abelson protein tyrosine kinase	K
2	akap79	Akap79	A kinase (PRKA) anchor protein 79	A
3	ap1	Ap1	v-jun sarcoma virus 17 oncogene homology	T
4	bad	Bad	BCL2-antagonist of cell death	
5	Bcat	?-catenin	Catenin (cadherin-associated protein) beta 1	
6	bcl10	BCL10	B-cell CLL/lymphoma 10	
7	bclxl	BCL2L1	BCL2-like 1	
8	ca	CA, Ca2+	Calcium	
9	cabin1	Cabin1, CAIN	Calcineurin-binding protein	
10	calcin	Calcineurin	Calcineurin	P
11	calpr1	Csp1	Calcipressin 1	
12	cam	CaM	Calmodulin	
13	camk2	Camk2,CamKII	calcium/calmodulin-dependent protein kinase (CaM kinase) II	K
14	camk4	CamK4, CamKIV	calcium/calmodulin-dependent protein kinase IV	K
15	card11a	CARD11 + CARMA1 + BCL10	Complex of CARD11, CARMA1, and BCL10, which is their fully active form (see Fig. 1)	
16	card11	CARMA1, CARD11	caspase recruitment domain family, member 11	
17	cblb	Cbl-b	Cas-Br-M (murine) ecotropic retroviral transforming sequence b	
18	ccblp1	c-Cbl	Casitase B-lineage lymphoma proto-oncogene (ZAP70-dependent pool)	A
19	ccblp2	c-Cbl	Casitase B-lineage lymphoma proto-oncogene (Fyn-dependent pool)	A
20	ccblr	c-Cbl	Casitase B-lineage lymphoma proto-oncogene (Reservoir, see legend)	A
21	cd28	CD28	CD28 antigen (Tp44)	R
22	cd45	CD45, PTPRC	protein tyrosine phosphatase, receptor type, C	P
23	cd4	CD4, p55	CD4 antigen (p55)	R
24	cdc42	CDC42, BB1	cell division cycle 42 (GTP binding protein, 25kDa)	G
25	cre	Cre	cAMP responsive element binding protein 1	
26	creb	CREB, CREB1	cAMP responsive element	T
27	csk	Csk	c-src tyrosine kinase	K

Nr	Model Name	Common abbreviation	Name	Type
28	cyc1	Cyc1	cytochrome c-1	
29	dag	DAG	Diacylglycerol	
30	dgk	DGK	diacylglycerol kinase	K
31	erk	ERK, MAPK1	Extracellular signal-regulated kinases, mitogen-activated protein kinase 1	K
32	fkhr	FOXO1, fkhr	forkhead box O1A	T
33	fos	Fos, FRA	FOS-like antigen	
34	fyn	Fyn	FYN oncogene related to SRC, FGR, YES	K
35	gab2	Gab2	GRB2-associated binding protein 2	A
36	gadd45	GADD45A	growth arrest and DNA-damage-inducible, alpha	
37	gads	GRAP-2, GADS, P38	GRB2-related adaptor protein 2	A
38	gap	GAP	GTPase-activating protein	
39	grb2	Grb2	growth factor receptor-bound protein 2	
40	sos	Sos1	Son of sevenless homolog 1	
41	gsk3	GSK3	glycogen synthase kinase 3	K
42	hpk1	HPK1, MAP4K1	Hematopoietic progenitor kinase 1, mitogen-activated protein kinase kinase kinase kinase 1	K
43	ikb	IκB	Inhibitor protein IκB	
44	ikkab	IκKA + IκKB	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase alpha and kinase beta	
45	ikkg	IκKG	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase gamma	
46	ip3	IP3K	inositol 1,4,5-trisphosphate	K
47	itk	ITK	IL2-inducible T-cell kinase	K
48	jnk	JNK, SAPK, MAPK8	mitogen-activated protein kinase 8	K
49	jun	JUN, ap1	v-jun sarcoma virus 17 oncogene homolog	T
50	lat	LAT	linker for activation of T cells	A
51	lckp1	Lck	lymphocyte-specific protein tyrosine kinase (CD4-dependent pool)	K
52	lckp2	Lck	lymphocyte-specific protein tyrosine kinase (CD28/TCR-dependent pool)	K
53	lckr	Lck	lymphocyte-specific protein tyrosine kinase (Reservoir, see legend)	K
54	malt1	MALT1	mucosa associated lymphoid tissue lymphoma translocation gene 1	
55	mek	MEK, MAP2K, MAPKK	mitogen-activated protein kinase kinase	K
56	mekk1	MEKK1, MAP3K1, MAPKKK1	mitogen-activated protein kinase kinase kinase	K
57	mkk4	MKK4, MAP2K4, MAPKK4	mitogen-activated protein kinase kinase 4	K
58	mlk3	MLK3, MAP3K11	mitogen-activated protein kinase kinase kinase 11	K
59	nfat	NFATC1, NFATC2, NFATC3	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1	T

Nr	Model Name	Common abbreviation	Name	Type
60	nfkb	NfκB	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1, 2	T
61	p21c	P21Cip1, Cdkn1a, p21	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	
62	p27k	CDKN1B, p27Kip1	cyclin-dependent kinase inhibitor 1B (p27, Kip1)	
63	p38	P38, MAPK14	mitogen-activated protein kinase 14	K
64	p70s	RPS6KB1, p70s6k	ribosomal protein S6 kinase, 70kDa, polypeptide 1	K
65	pag	PAG1	phosphoprotein associated with glycosphingolipid microdomains 1	A
66	pdk1	PDPK1, PDK1	3-phosphoinositide dependent protein kinase-1	K
67	pi3k	PI3K, PIK3CA	phosphoinositide-3-kinase, catalytic, alpha polypeptide	K
68	pip3	PIP3	Phosphatidylinositol (3,4,5)-trisphosphate	
69	pkb	PKB, AKT1	v-akt murine thymoma viral oncogene homolog 1	K
70	pkcth	Prkcq, PKC-γ	protein kinase C, theta	K
71	plcga	Active form of PLCγ1, PLC-gamma-1	phospholipase C, gamma 1	
72	plcgb	PLCG1, PLC-γ-1 bound to LAT (but not active, see plcga)	phospholipase C, gamma 1	
73	pten	PTEN	phosphatase and tensin homolog (mutated in multiple advanced cancers 1)	P
74	rac1p1	RAC1, p21-Rac1, MIG5	ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1) (Vav-1 dependent pool)	G
75	rac1p2	RAC1, p21-Rac1, MIG5	ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1) (Vav-3 dependent pool)	G
76	rac1r	RAC1, p21-Rac1, MIG5	ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1) (Reservoir, see footnote)	G
77	raf	RAF1	v-raf-1 murine leukemia viral oncogene homolog 1	K
78	ras	Ras	Ras	G
79	rasgrp	RasGRP1	Ras guanyl releasing protein 1	
80	rlk	TXK, RLK, TKL	TXK tyrosine kinase	K
81	rsk	RPS6KA1, RSK	ribosomal protein S6 kinase, 90kDa, polypeptide 1	K
82	Sh3bp2	3BP2, SH3BP2	SH3-domain binding protein 2	A
83	ship1	INPP5D, SHIP1	inositol polyphosphate-5-phosphatase, 145kDa	P
84	shp1	Shp1, PTPN6	protein tyrosine phosphatase, non-receptor type 6	P
85	shp2	SHP2, PTPN11	protein tyrosine phosphatase, non-receptor type 11 (Noonan syndrome 1)	P
86	slp76	LCP2, SLP76	lymphocyte cytosolic protein 2 (SH2 domain containing leukocyte protein of 76kDa)	A
87	sre	Sre	Serum responsive element	
88	tcrb	TCR	T-cell Receptor (bound to ligand)	R

Nr	Model Name	Common abbreviation	Name	Type
89	tcrlig	PMHC, Ab	Ligand of T cell receptor (peptide-MHC complex or antibody)	R
90	tcrp	TCR	T-cell Receptor phosphorylated	R
91	vav1	VAV1	Vav 1 oncogene	
92	vav3	VAV3	Vav 3 oncogene	
93	X	-	Non-identified kinase involved in CD28-mediated signaling (see main text)	K
94	zap70	ZAP70	zeta-chain (TCR) associated protein kinase 70kDa	K

Table A.2: List of reactions (hyperarcs) of the logical T-cell signaling model. Exclamation mark ('!') denotes a logical NOT and dots within the equations indicate AND operations. The names of the substances in the explanations are those used in the model and Fig. 6.5; the biological names are displayed in the Table A.1. In the case where two pools of a molecule were considered (e.g. lckp1 and lckp2), a'reservoir' (lckr) was included which was required for both pools. This allows to perform a simultaneous knock-out of both pools acting on the reservoir (see Figure 6.5).

Nr	Reaction	τ	Documentation
1	$\rightarrow$ cd28	1	Binding of ligand or antibody to cd28 is an input of the model.
2	$\rightarrow$ cd4	1	Binding of ligand or antibody to cd4 is an input of the model.
3	$\rightarrow$ tcrlig	1	Binding of ligand or antibody to the tcr is an input of the model.
4	!bad $\rightarrow$ bclxl	1	bad inhibits bclxl[282;287].
5	!cabin1 · !calpr1 · !akap79 · cam $\rightarrow$ calcin	1	cam binds to and activates calcineurin (calcin), while cabin1, calpr1, akap79 inhibit calcin[130;171;219].
6	!camk4 $\rightarrow$ cabin1	1	camk4 regulates via phosphorylation nuclear export of Cabin1[194].
7	card11a · pkcth $\rightarrow$ ikkg	1	The complex card11+bcl10 +malt1 is required for ikkg activation[104;258;269]. Phosphorylation, probably via pkcth[138], is also required.
8	!ccblp1 · tcrlig $\rightarrow$ tcrb	1	Binding of ligand activates the tcr, while active ccbl ubiquinates it, thus leading to tcr degradation[118].
9	!ccblp1 · tcrp · abl $\rightarrow$ zap70	1	abl phosphorylates and thus activates zap70[289] once it is bound to the tcr. Active ccbl can degrade zap70.
10	!cd28 $\rightarrow$ cblb	2	cd28 induces cblb ubiquitination and degradation[288] after the early events thus, τ=2.
11	!dgk · plcga $\rightarrow$ dag	1	The active form of plc γ 1 (plcga) splits pip2 into diacylglycerol (dag) and ip3 (see hyperarc 83)[118]. Active dgks degrade dag into phosphatic acid[261].
12	!erk · lckp1 $\rightarrow$ shp1	2	lck phosphorylates shp1 leading to its activation which allows it to dephosphorylate and thus deactivate lck. erk phosphorylates lck at p59, protecting it from shp1's effect[4;246]. Since shp1 activation comes some time after lck activation, it takes place at τ=2.
13	!gab2 · zap70 gads $\rightarrow$ slp76	1	slp76 associates with lat via gads[114;260]. gab2 competes for binding, and thus inhibits binding of slp76 to gads[279;280].
14	!gsk3 $\rightarrow$ bcat	1	Gsk3 inhibits bcat[159].
15	!gsk3 $\rightarrow$ cyc1	1	Gsk3 inhibits cyc1[159].
16	!ikb $\rightarrow$ nfkb	1	nfkb is retained in the cytoplasm by tight binding to the inhibitory protein ikb[118].
17	!ikkab $\rightarrow$ ikb	1	ikb is phosphorylated by ikkab, leading to its ubiquination and subsequent degradation[118;146].
18	ikkg · camk2 $\rightarrow$ ikkab	1	Both the regulatory molecule ikkg and phosphorylation probably (but not only) via camk2 are required for the activation of the kinase subunits ikkalpha and beta (ikkab)[104;258;269].
19	!pkb $\rightarrow$ bad	1	pkb inhibits bad[101].
20	!pkb $\rightarrow$ fkhr	1	pkb inhibits fkhr[101].
21	!pkb $\rightarrow$ gsk3	1	pkb inhibits gsk3[101;159].

Nr	Reaction	τ	Documentation
22	!pkb → p21c	1	pkb inhibits p21c[101;159].
23	!pkb → p27k	1	pkb inhibits p27k[101;159].
24	!gadd45 · zap70 → p38	1	gadd45 inhibits the zap70 mediated activation of p38[225].
25	!shp1 · cd45 · cd4 · !csk · lckr → lckp1	1	Full activation of the cd4-bound pool (there is also a tcr-dependent pool, see hyperarc 62/64 and legend) of lck requires dephosphorylation of the negative regulatory site (by cd45, and in absence of csk, which phosphorylates it) and autophosphorylation of the positive regulatory site, which cd4-bound lck can perform upon cd4 crosslinking[193].
26	!tcrb → pag	1	upon ligand binding to the tcr, pag is dephosphorylated by an unidentified phosphatase (probably cd45)[53].
27	ap1 →	1	the transcription factor ap1 is an output of the model.
28	bcat →	1	bcat is an output of the model.
29	bclxl →	1	bclx is an output of the model.
30	ca → cam	1	calcium binds to calmodulin and this complex to calcineurin[79].
31	calcin → nfat	1	calcineurin dephosphorylates nfat leading to nuclear translocation and activation of nfat[118;146;166].
32	cam → camk4	1	camk2 activation is dependent on calmodulin (cam)[5].
33	ccblr · fyn → ccblp2	2	Upon Fyn phosphorylation, ccbl can inhibit plcg[213]. This is one out of 2 mechanisms ccbl is involved in, and we call it ccblp2 (pool 2, see legend). Since ccbl mediated inhibition is slower than the early events, τ=2.
34	ccblr · zap70 → ccblp1	2	ccbl binds to activated (and thus phosphorylated) zap70, leading to the ubiquination and subsequent degradation of zap70 and tcr[207]. This is one out of 2 mechanisms ccbl is involved in, and we call it ccblp1 (pool 1, see legend). Since ccbl mediated degradation has to be slower than the early events, τ=2.
35	x → vav1	1	CD28 stimulation leads to Vav1 activation[105;175], a process mediated by a yet unidentified kinase (see hyperarc 48).
36	cdc42 → mekk1	1	The GTP bound cdc42 (and rac1, see hyperarc 87) is able to bind mekk1[71]; CD28 activates mekk1 in a cdc42 mediated manner[128].
37	cre →	1	cre is an output of the model.
38	creb → cre	1	The creb protein is a transcription factor that binds to cre ativating the related genes[146].
39	cyc1 →	1	cyc1 is an output of the model, and is involved in cell cycle regulation[146].
40	dag → rasgrp	1	dag causes the cytoplasmic rasgrp1 to move to the golgi, where it can act on Golgi associated-ras[22;59]. Even though pkcth phosphorylates rasgrp at t184,[218] we did not include connection pkcth → rasgrp1 since this effect is not specific to pkcth, but general to other pkcs (less well-characterized in T cells and therefore not included in the model); inclusion of this effect would make this step strictly dependent on pkcth, which is not the case.
41	dag · vav1 · pdk1 → pkcth	1	Activation of pkcth requires binding to dag, phosphorylation by pdk1[155], and vav1[267].

Nr	Reaction	τ	Documentation
42	erk $\rightarrow$ fos	1	erk phosphorylates fos[118].
43	erk $\rightarrow$ rsk	1	erk activates rsk via phosphorylation[84].
44	fkhr $\rightarrow$	1	The transcription factor fkhr is an output of the model.
45	fos $\cdot$ jun $\rightarrow$ ap1	1	Binding of jun with fos leads to the formation of ap1[118;146].
46	fyn $\rightarrow$ abl	1	abl kinases are activated following tcr stimulation via a Src kinase (lck or fyn, see hyperarc 59)[289].
47	fyn $\rightarrow$ pag	2	fyn phosphorylates pag[53], leading to the binding of csk. This process takes place 3-5 min after tcr activation, and thus it belongs to the time scale τ=2[262].
48	cd28 $\rightarrow$ x	1	Vav1 activation requires cd28 activation[175] and is mediated by an non-identified kinase x (see hyperarcs 35 and 63).
49	gab2 $\rightarrow$ shp2	1	Gab2 recruits shp2[8].
50	gads $\cdot$ lat $\cdot$ zap70 $\rightarrow$ gab2	2	zap70 phosphorylates gab2 upon binding to lat and gads[279;280]. This process must take place after the early events to allow signal propagation, thus τ=2.
51	grb2 $\cdot$ lat $\cdot$ zap70 $\rightarrow$ gab2	2	zap70 phosphorylates gab2 upon binding to lat and grb2[279;280]. This process must take place after the early events to allow signal propagation, thus τ=2.
52	hpk1 $\rightarrow$ mekk1	1	hpk1 binds and phosphorylates mekk1[116].
53	hpk1 $\rightarrow$ mlk3	1	hpk1 binds and phosphorylates mlk3[259].
54	ip3 $\rightarrow$ ca	1	Binding of ip3 to the ip3 receptor in the endoplasmatic reticulum leads to the release of calcium[35].
55	jnk $\rightarrow$ jun	1	jnk phosphorylates jun[146].
56	lat $\rightarrow$ grb2	1	grb2 (which in turn binds sos) can bind to phosphorylated lat[160][114].
57	lat $\rightarrow$ hpk1	1	hpk1 binds to lat and is recruited to the lipid raftss[161].
58	lat $\rightarrow$ plcgb	1	plcgamma binds to lat[114;260].
59	lckp1 $\rightarrow$ abl	1	abl kinases are activated following tcr stimulation via a Src kinase (lck or fyn, see hyperarc 46)[289].
60	lckp1 $\rightarrow$ rlk	1	lck phosphorylates rlk leading to its activation[238].
61	lckp1 $\cdot$ cd45 $\rightarrow$ fyn	1	lck activates fyn[81], a process where the dephosphorylation of the negative regulatory site of fyn by cd45 is also required.
62	lckp2 $\cdot$!cblb $\rightarrow$ pi3k	1	pi3k is dependent on the Src kinase lck for activation[57]. Additionally, cblb promotes pi3k ubiquination[70].
63	x $\cdot$!cblb $\rightarrow$ pi3k	1	pi3k is also activated upon CD28[12;90] via an non-determined kinase x (see hyperarc 48). Even though Lck has been proposed to be involved in this process[91;113;117;268], our experiments show that, at least for primary human T-cells, PI3K activation is not strictly Src kinase dependent (see Fig. A.8). A reasonable candidate would be a Tec kinase, but since it is not experimentally verified, we keep an undetermined x.
64	lckr $\cdot$ tcrb $\rightarrow$ lckp2	1	The activation of pi3k is determined by a second pool of lck (lckp2) (see legend) which can be activated by tcr activation[189].
65	malt1 $\cdot$ card11 $\cdot$ bcl10 $\rightarrow$ card11a	1	The binding of malt1 to card11 and bcl10 forms the active card11 complex[86;164;258;265].
66	mek $\rightarrow$ erk	1	mek phosphorylates erk leading to erk activation[118;146].
67	mekk1 $\rightarrow$ jnk	1	mekk1 activates jnk[54].

Nr	Reaction	τ	Documentation
68	mekk1 → mkk4	1	mekk1 is able to phosphorylate MKK4 leading to its activation[281].
69	mekk1 → p38	1	mekk1 leads to p38 activation[99].
70	mkk4 → jnk	1	MKK 4 activates jnk[54;58].
71	mlk3 → mkk4	1	mlk3 phosphorylates mkk4[259].
72	nfkb →	1	nfkb is an output of the model.
73	p21c →	1	p21cip is an output of the model controlling the cell cycle.
74	p27k →	1	p27kip is an output of the model controlling the cell cycle.
75	p38 →	1	p38 is an output of the model.
76	p70s →	1	p70s is an output of the model.
77	pag → csk	1	Phosphorylation of pag allows csk to bind it and then act on lck[114;160].
78	pdk1 → p70s	1	pkd1 phosphorylates p70s leading to its activation[110;202].
79	pdk1 → pkb	1	pdk1 phosphorylates pkb leading to its activation[1;6;150].
80	pi3k · !ship1 · !pten → pip3	1	pi3k leads to the production of pip3, while ship1 and pten inhibit this process[188;206].
81	pip3 → pdk1	1	pip3 is required for pdk1 activation[180].
82	pip3 · zap70 · slp76 → itk	1	When phosphorylated, slp76 can bind to itk; additional binding to pip3 and phosphorylation via zap70 activates itk[50;118;260].
83	plcga → ip3	1	Active plcga splits pip2 into ip3 and diacylglycerol (dag,see hyperarc 11)[118;260].
84	plcgb · !ccblp2 · slp76 · zap70 · vav1 · itk → plcga	1	Once bound to phosphorylated lat, plcgb is activated by the combined action of vav and itk (or rlk, see hyperarc 85)[50]. Additionally, binding to slp76 (phosphorylated by zap70) is required to establish and stabilize the complex. Activated ccbl degrades plcga[213].
85	plcgb · !ccblp2 · zap70 · vav1 · slp76 · rlk → plcga	1	Once bound to phosphorylated lat, plcgb is activated by the combined action of vav and rlk (or itk, see hyperarc 84)[50]. Additionally, binding to slp76 (phosphorylated by zap70) is required to establish and stabilize the complex. Activated ccbl degrades plcga[213].
86	rac1p1 → mlk3	1	Rac1p1 activates mlk3[255].
87	rac1p2 → mekk1	1	GTP-bound Rac1p2 is able to bind mekk1[71], and active mekk1 leads to JNK activation[128;179].
88	rac1p2 → sre	1	Vav3-dependent Rac1 is able to activate Sre via SRF[109].
89	rac1r · vav1 → rac1p1	1	Downregulation of Vav1 but not Vav3 affects IL-2 production in T cells[284] via the rac1-mediated jnk pathway. Since rac1 mediates this process, we defined a vav1-dependent pool of rac1 (see hyperarc 90 and legend).
90	rac1r · vav3 → rac1p2	1	Downregulation of Vav3 but not Vav1 affects Sre activity[284]. Since rac1 mediates this process, we defined a vav3-dependent pool of rac1 (see hyperarc 89 and legend)
91	raf → mek	1	Raf phosphorylates mek leading to mek activation[83].
92	ras → raf	1	Ras mediates raf localization to the membrane, and consequently, raf is activated[146].
93	rsk → creb	1	Rsk phosphorylates creb increasing its activity[84].

Nr	Reaction	τ	Documentation
94	sh3bp2 → vav3	1	sh3bp2 binds vav3 via an sh2 domain, leading to its activation[284].
95	sos · !gap · rasgrp → ras	1	Bound to lat via grb2, sos catalyzes the exchange of GTP for GDP in the cellular-membrane-located ras, while rasgrp1 catalyzes the exchange of GTP for GDP in golgi-located ras[59]. In turn gap catalyzes the conversion GTP to GDP and thus deactivates ras[89].
96	sre →	1	Sre is an output of the model.
97	tcrb → dgk	1	dgks get activated after tcr activation in yet an unclear manner, we therefore make it dependent on activation of the tcr. Since dag must be produced in the early events, we assign it a $\tau=2$[126].
98	tcrb · fyn → tcrp	1	Upon ligand binding to the tcr, active fyn can phosphorylate the tcr[80].
99	tcrb · lckp1 → tcrp	1	The co-localization of tcr with cd4 mediated by peptide-MHC or antibody crosslinking results in an increased local concentration of lck around the tcr leading to phosphorylation of ITAMs[81].
100	tcrb · lckr → fyn	1	A fraction of fyn is bound to the tcr, and tcr crosslinking leads to fyn auotophosphorylation and activation[80]. Since lck is required in the development for having capable fyn[285], lckr (existence of lck in the cell) is required as well.
101	zap70 → lat	1	zap70 phosphorylates lat at different sites[118].
102	zap70 · lat → sh3bp2	1	sh3bp2 binds to phosphorylated lat upon phosphorylation by zap70[205].
103	zap70 · sh3bp2 → vav1	1	zap70 phosphorylates vav1[284] which together with binding of vav1 to sh3bp2[205], leads to vav1 activation.
104	→ card11	1	Regulation of card11 is not clear, thus we set an external input to it. Default value is 1.
105	→ gadd45	1	Regulation of gadd45 is not clear, thus we set an external input to it. Default value is 1.
106	→ gap	1	GTP activating proteins (gaps) are important regulators of ras activation but their own regulation is not clear[31]. Therefore they are included in the model with an external input.
107	→ lckr	1	Input to the system (presence of Lck in the cell). Default value is 1.
108	cam → camk2	1	cam (calmodulin) activates calmodulin-dependent kinase II (camk2)[120].
109	grb2 → sos	1	sos binds to grb2 and thus get recruited to the membrane via lat[30].
110	lat → gads	1	gads can bind to phosphorylated lat[114;260].
111	cdc42 → sre	1	cdc42 is able to activate Sre via SRF[109].
112	nfat →	1	nfat is an output of the model.
113	shp2 →	1	shp2 is an output of the model.
114	→ cd45	1	Regulation of cd45 is not clear, thus we set an external input to it. Default value is 1.
115	→ pten	1	Regulation of pten is not clear, thus we set an external input. Default value is 0.

Nr	Reaction	τ	Documentation
116	→ bcl10	1	Regulation of bcl10 is not clear, thus we set an external input to it. Default value is 1.
117	→ ccblr	1	Input to the system (presence of ccbl in the cell). Default value is 1.
118	→ cdc42	1	Regulation of cdc42 is not clear, thus we set an external input to it. Default value is 0.
119	→ malt1	1	Regulation of malt1 is not clear, thus we set an external input to it. Default value is 1.
120	→ rac1r	1	Input to the system (presence of rac1 in the cell). Default value is 1.
121	→ ship1	1	Regulation of ship1 is not clear, thus we set an external input. Default value is 0.
122	→ akap79	1	Regulation of akap79 is not clear, thus we set an external input to it. Default value is 0.
123	→ calpr1	1	Regulation of calpr1 is not clear, thus we set an external input to it. Default value is 0.

A.7 Description of the model for the TCR-induced MAPK and fit of experimental data

A.7.1 Minimal model

The model is defined as follows, implemented in the Matlab Systems Biology Toolbox[232]:

```
********** MODEL NAME
FeedbGenExtMult
********** MODEL STATES
d/dt(x1)  = r1                    x1(0)  = 0.000000
d/dt(x2)  = r2                    x2(0)  = 0.000000
d/dt(x3)  = r3                    x3(0)  = 0.000000
d/dt(x4)  = r4                    x4(0)  = 0.000000
********** MODEL PARAMETERS
Switch=1 Input = 1
x10   = 1         k1  = 1  km1 = 1    n1  = 1     k1r = 1  km1r= 1        n1r = 1
x20=1             k2  = 1  km2 = 1    n2  = 1     k2r = 1  km2r= 1        n2r = 1
kmco = 1 nco  = 1    kf1 = 0       kf2 = 0      kf4 = 0
x30 = 1  k3  = 1   km3 = 1        n3  = 1     k3r = 1  km3r= 1      n3r = 1
x40 = 1 k4  = 1   km4 = 1        n4  = 1     k4r = 1  km4r= 1  n4r = 1
********** MODEL VARIABLES
Su1   = (x10-x1 -kf1*x2 )          Su1r   = x1
Su2   = (x20-x2)                   Su2r   = x2
Coup  = power(x1 ,nco)
Su3   = (x30-x3)                   Su3r   = x3
Su4   = (x40-x4)                   Su4r   = x4
********** MODEL REACTIONS
r1 = ((1+Switch*Input)  /(1+kf2*x2 ))*HillF(Su1 ,k1,km1,n1)-HillF(Su1r ,k1r,km1r,n1r)
r2 = Coup      *HillF(Su2 ,k2,km2,n2)-HillF(Su2r ,k2r,km2r,n2r)
r3 = (x1   /(1+kf4*x4 ))*HillF(Su3 ,k3,km3,n3)-HillF(Su3r ,k3r,km3r,n3r)
r4 = x3        *HillF(Su4 ,k4,km4,n4)-HillF(Su4r ,k4r,km4r,n4r)
********** MODEL FUNCTIONS
HillF(Su,ka,km,n) = ka*power(Su,n)/(power(km,n)+power(Su,n))
```

The model was fitted to the experimental data depicted in Figure 7.3. It should be noted that the experimental points in the objective function were not weighted with the inverse of the Standard Error of the Mean but rather, for t=10 and 30 minutes were given a 2 and 5-fold weight, respectively, to impose the model to fit preferentially the long-time data, which is the characteristic for the qualitative behavior. Since the concentration and parameters are normed, the intervals for the parameter estimation were set to $10^5 - 10^{-5}$ for the enzymatic activities, $10^3 - 10^{-3}$ for the Michaelis-Menten constants and $10 - 10^{-1}$ for the Hill coefficients. These ranges are large enough to provide the system sufficient flexibility to fit the data and, at the same time, narrow enough to avoid too different parameter values, which would be impossible *de facto* due to physical limitations such as diffusion coefficients, etc. Narrower limits were tested which provided qualitatively similar results but worse fits of the data.

Different methods, all wrapped around the Matlab Systems Biology Toolbox[232], were successive applied:

1. Simmulated Annealing

2. Simplex

3. Eolutionary algorithm[215;249] based on the ideas of Rechenberg[211] (2 times, each time 10 parents, 30 children, and 5000 generations)

4. Simplex

The following results were obtained, which provide the fits plotted in Figure 7.6(b).

Initial parameter values:

Switch	Input	x10	k1	km1	n1	k1r	km1r	n1r	k2	km2	n2	k2r	km2r	n2r	x20	kmco	nco	kf1	kf2	kf4	x30	k3	km3	n3	k3r	km3r	n3r	x40	k4	km4	n4	k4r	km4r	n4r
1	1e+02	1	1	1	0.4	3	0.4	1	0.7	4	1	3	1e+02	3	1	1	1	2	0	0	8e+02	2	1	0.7	2	4	2	0.3	1	2	0.01	2	0.2	2

Parameters fitted and their low (LB) and high (HB) bounds:

Par	Input	x10	k1	k1r	km1	km1r	n1	n1r	k2	k2r	km2	km2r	n2	n2r	x30	k3	k3r	km3	km3r	n3	n3r	x40	k4	k4r	km4	km4r	n4	n4r
LB	1e-5	1e-5	1e-5	1e-5	0.001	0.001	0.01	0.01	1e-5	1e-5	0.001	0.001	0.01	0.01	1e-5	1e-5	1e-5	0.001	0.001	0.01	0.01	1e-5	1e-5	1e-5	0.001	0.001	0.01	0.01
HB	1e+5	1e+5	1e+5	1e+5	1e+3	1e+3	1e+2	1e+2	1e+5	1e+5	1e+3	1e+3	1e+2	1e+2	1e+5	1e+5	1e+5	1e+3	1e+3	1e+2	1e+2	1e+5	1e+5	1e+5	1e+3	1e+3	1e+2	1e+2

Results:

Fit	Input	x10	kf1	k1	k1r	km1	km1r	n1	n1r	k2	k2r	km2	km2r	n2	n2r	x30	k3	k3r	km3	km3r	n3	n3r	x40	k4	k4r	km4	km4r	n4	n4r
2.46	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0.776	1e+2	1	2	1	3	0.4	0.4	1	1	0.7	3	4	1e+2	3	1	8e+2	2	4	1	1	2	2	0.3	4	2	0.3	0.2	1	2
0.776	1e+2	1	2	1	3	0.4	0.4	1	1	0.7	3	4	1e+2	3	1	8e+2	2	4	1	1	2	2	0.3	4	2	0.3	0.2	1	2
0.776	1e+2	1	2	1	3	0.4	0.4	1	1	0.7	3	4	1e+2	3	1	8e+2	2	4	1	1	2	2	0.3	4	2	0.3	0.2	1	2
0.577	8e+1	1	1e+2	1	1e+3	5	8	1	2	2e+3	3	7e+2	3e+2	2	1	1e+3	2e+04	2e+04	5e+2	1e+3	3	3	0.4	6	1e+1	0.6	0.6	2	2
0.577	8e+1	1	1e+2	1	1e+3	5	8	1	2	2e+3	3	7e+2	3e+2	2	1	1e+3	2e+04	2e+04	5e+2	1e+3	3	3	0.4	6	1e+1	0.6	0.6	2	2
0.556	8e+1	1	1e+2	1	1e+3	5	8	1	2	2e+3	2	7e+2	4e+2	2	1	9e+2	2e+04	2e+04	5e+2	1e+3	4	3	0.5	7	8	0.5	0.3	2	3

A.7.2 Extended model

The model is defined as follows, implemented in the Matlab Systems Biology Toolbox[232]:

```
********** MODEL NAME
FeedbGenExtMultPosFInt
********** MODEL STATES
d/dt(x1) = r1            x1(0)  = 0.000000
d/dt(x2) = r2            x2(0)  = 0.000000
d/dt(x2i)= r2i           x3(0)  = 0.000000
d/dt(x3) = r3            x1(0)  = 0.000000
d/dt(x4) = r4            x5(0)  = 0.000000
********** MODEL PARAMETERS
Switch=1 Input = 1
x10   = 1    k1  = 1    km1 = 1     n1  = 1     k1r = 1    km1r= 1    n1r = 1
x20=1        k2  = 1    km2 = 0.3  n2  = 1      k2r = 60  km2r= 0.1   n2r = 1          kpf=220
kmco = 0     nco  = 1   kf1 = 0                 kf2 = 0    kf4 = 0
k2i  = 1    km2i = 1    n2i  = 1 CoD=1   For=1
x30 = 1  k3  = 1    km3 = 1         n3  = 1      k3r = 1    km3r= 1        n3r = 1
x40 = 1 k4  = 1    km4 = 1          n4  = 1      k4r = 1    km4r= 1   n4r = 1
********** MODEL VARIABLES
Su1   = (x10-x1 -kf1*x2i )                    Su1r  = x1
Su2   = (x20-x2)                              Su2r  = x2
Coup  = (power(x1 ,nco)/(1+kmco*power(x1 ,nco)))
Su2i  = x1
Su3   = (x30-x3)                              Su3r  = x3
Su4   = (x40-x4)                              Su4r  = x4
********** MODEL REACTIONS
r1 = ((1+Switch*Input)  /(1+kf2*x2i ))*HillF(Su1 ,k1,km1,n1)-HillF(Su1r ,k1r,km1r,n1r)
r2 = (40*Coup           +kpf*x2)*HillF(Su2 ,k2,km2,n2)-HillF(Su2r ,k2r,km2r,n2r)
r2i=                    (CoD+x2)  *HillF(Su2i ,k2i,km2i,n2i)-For
r3 = (x1   /(1+kf4*x4 ))*HillF(Su3 ,k3,km3,n3)-HillF(Su3r ,k3r,km3r,n3r)
r4 = x3                *HillF(Su4 ,k4,km4,n4)-HillF(Su4r ,k4r,km4r,n4r)
********** MODEL FUNCTIONS
HillF(Su,ka,km,n) = ka*power(Su,n)/(power(km,n)+power(Su,n))
```

The model was fitted to the experimental data depicted in Figure 7.3. The same considerations of the minimal model (Section A.7.1) with respect to the Standard Error of the Mean, optimization methods applied, and ranges for the parameters apply here. Furthermore, and to drive the system to produce the desired steady state behavior, an additional, fictive point at t=1000 minutes, with the same values for all concentrations as for t=30 minutes (the latest available), was added. The resulting parameters provide the fits plotted in Figure 7.14(b) and and the switch behavior of Figure 7.13(b).

Initial parameter values:

Switch	Input	x10	k1	km1	n1	k1r	km1r	n1r	nco	kf1	kf2	kf4	k2i	km2i	n2i	CoD	For	x30	k3	km3	n3	k3r	km3r	n3r	x40	k4	k4r	km4	km4r	n4	n4r
	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Parameters fitted and their low (LB) and high (HB) bounds:

Par	Input	x10	k1	k1r	km1	km1r	n1	n1r	k2i	km2i	n2i	x30	k3	k3r	km3	km3r	n3	n3r	x40	k4	k4r	km4	km4r	n4	n4r
LB	1e-4	0.01	1e-4	1e-4	0.01	0.01	0.1	0.1	1e-4	0.01	0.1	1e-4	1e-4	1e-4	0.01	0.01	0.1	0.1	1e-4	1e-4	1e-4	0.01	0.01	0.1	0.1
HB	1e+04	100	1e+04	1e+04	100	100	10	10	1e+04	100	10	1e+04	1e+04	1e+04	100	100	10	10	1e+04	1e+04	1e+04	100	100	10	10

Results:

Fit	Input	x10	kf1	k1	k1r	km1	km1r	n1	n1r	k2i	km2i	n2i	x30	k3	k3r	km3	km3r	n3	n3r	x40	k4	k4r	km4	km4r	n4	n4r
2.7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0.829	1e+04	1	2	2	9e+02	0.8	9e+01	1	1	1	1e+01	1	4	9e+03	6e+02	2e+01	2	2	1	3e+03	6e+03	5e+03	1e+02	0.4	0.6	2
0.779	1e+04	1	2	3e+02	4e+01	7e+01	1e+02	2	1	0.8	2e+01	0.8	4	9e+03	1e+03	9e+01	4	1	1	3e+03	6e+03	5e+03	1e+02	0.4	0.8	3
0.756	1e+04	1	2	3e+02	5e+01	6e+01	1e+02	2	1	0.7	1e+02	0.7	5	6e+03	1e+03	3e+01	4	4	0.9	9e+03	6e+03	6e+03	2e+01	0.4	8	3
0.756	1e+04	1	2	3e+02	5e+01	6e+01	1e+02	2	1	0.7	2e+01	0.7	5	6e+03	1e+03	3e+01	4	4	0.9	9e+03	6e+03	6e+03	2e+01	0.4	8	3

List of Symbols and Abbreviations

τ Signaling time of a certain response, as defined by Heinrich et al.[106], page 67

$\tau_{0.9}$ Signaling time of a certain response, as defined by Saez-Rodriguez et al.[222], page 67

θ Signal duration of a certain response, as defined by Heinrich et al.[106], page 67

$\vec{p}$ Vector of parameters of a dynamical system, page 18

$\vec{u}$ Vector of inputs of a dynamical system, page 18

J Jacobian of a ODE system, page 29

N stoichiometric matrix, page 20

N^C product of the stoichiometric matrix by its transpose multiplied by (-1) $N(-N)^T$, page 32

R Retroactivity matrix, page 28

S Signal amplitude of a certain response, as defined by Heinrich et al.[106], page 67

T Time constant of a temporally lagged element of first order (PT_1), page 64

CNA   CellNetAnalyzer[144], page 85

CRNT   Feinberg's Chemical Reaction Network Theory, page 55

EGF   Epidermal Growth Factor, page 7

EGFR   Epidermal Growth Factor Receptor, page 7

I/O   Input/Output, page 61

IG   Interaction graph, page 79

LIH   Logical interaction hypergraph, page 81

MAPK   Mitogen-Activated Protein Kinase, page 8

MIS   Minimal Intervention Set, page 84

molec/cell   Number of molecules per cell, page 119

nM   nano-mol/liter, page 118

ODE   ordinary differential equation, page 18

OT-1   Transgenic mouse genetically modified to express the T-cell Receptor specific for ovalbumin[112], page 10

ProMoT   Process Modeling Tool[95], page 32

SBML   Systems Biology Markup Language, page 40

Note that the abbreviations of the molecules in the T-cell model are listed in Table A.1

Index

Lebenslauf

Dipl. Ing. Julio Sáez Rodríguez

15 Littell Road
02446 Brookline MA USA
Tel: +001-617-432-6929
saez@hms.harvard.edu

Geboren am 06/07/1978 in Oviedo – Spanien
Staatsangehörigkeit: spanisch
Familienstand: ledig

Studium und Ausbildung

- 1992-1996 Gymnasium Calderon de la Barca in Gijón (Spanien)

- 1996 „Selectividad" (entspricht Abitur)

- 1996-2000 Studium der Chemieingenieuertechnik an der Facultad de Química de Oviedo – Spanien. Zusätzlich Besuch von Vorlesungen über Molekulargenetik und Gentechnik.

- 2000-2001 Austausch Jahr im Universität Stuttgart im Rahmen des Erasmusprogramm. Diplomarbeit am Institut für Bioverfahrenstechnik. Thema ' Computer Aided Set-up and Analysis of Dynamic Signalling Networks: The TGF-ß Signal Transduction in Eukaryotic Systems'.

- 2001 „Premio Du Pont" von Universidad de Oviedo (Auszeichnung für das beste Abschlußzeugnis).

- 2002 „Mención Especial" von spanische Ministerium für Bildung (Auszeichnung für besondere gute Abschlußzeugnis).

Arbeit

- Oktober 2001-Januar 2002: Arbeit als Wissenschaftlicher Hilskraft am Institut für Bioverfahrenstechnik.

- Marz 2002-Januar 2007: Arbeit als Wissenschaftlicher Mitarbeiter am Max-Planck-Institut für Dynamik komplexer technischer Systeme in Magdeburg.

- Ab Februar 2007: Arbeit als Wissenschaftlicher Mitarbeiter am Systems Biology Department, Harvard Medical School in Boston, USA.