Dissertation

submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto Carola University of Heidelberg, Germany
for the degree of

Doctor of Natural Sciences

presented by

Dipl. Inf. Yvonne Koch
born in: Ulm, Germany
Oral-examination: 5th of April 2011

Analysis of Gene Expression Data by Systems Biology Methods

Referees:

PD Dr. Rainer König

Prof. Dr. Ursula Kummer

Eidesstattliche Erklärung

Hiermit erkläre ich, Yvonne Koch, an Eides Statt, dass ich die an der Universität Heidelberg vorgelegte Dissertation selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen übernommenen Gedanken sind als solche kenntlich gemacht.

Die Arbeit wurde bisher in gleicher oder ähnlicher Weise keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht.

Heidelberg, den 22. Februar 2011

_____

Unterschrift

Acknowledgements

First of all, I would like to thank Professor Dr. Roland Eils who gave me the opportunity to do my PhD in the exciting field of systems biology at the Department of Theoretical Bioinformatics at the German Cancer Research Center. The support and freedom he granted enabled me to conduct a research project of interests.

I am grateful to PD Dr. Rainer König, who filled in for my first supervisor for his support and helpful discussion.

Further, I thank Professor Dr. Ursula Kummer for her scientific advice as part of my thesis advisory committee (TAC) and for being member of my thesis committee at the Combined Faculties of the Natural Sciences and Mathematics of the University of Heidelberg.

Thanks to Dr. Lars Kaderali, also part of my TAC, for his research advice.

Special thanks go to my supervisor Dr. Benedikt Brors, head of the research group Computational Oncology at the German Cancer Research Center. Work discussions during all stages of my PhD thesis have been helpful and his support enabled to shape my ideas. Moreover, breadth and depth of his knowledge have been truly inspiring to me.

Furthermore, I thank Barbara Roider, a biotechnology student, who has been eager to learn about gene expression analysis and mechanistic modeling during a student project. Many thanks go to Barbara for her collaboration.

Dr. Leo Neumann, Dr. Hannah Schmidt-Glenewinkel and Dr. Peter Bewerunge were great colleagues who have been always open to discuss and help me with the basics of mathematical modeling and statistical analyses in Matlab and R. I am grateful, that I met them at the German Cancer Research Center.

Further, I would like to thank Esteban Czwan and Dr. Joel Beaudouin for carefully reading the manuscript and for their helpful comments.

I want to express my gratitude to my parents for their constant care and encouragement as well as to my friends Julia Korte and Karmen Lau for being incredible friends.

Finally, I thank Torsten Franz for his loving support during the past years.

Abstract

Perturbed regulation of programmed cell death (apoptosis) plays a major role in tumor development. The process of eliminating unwanted and damaged cells by apoptosis is based on complex molecule interactions. Studies of various diseases generally analyse patient samples in order to find markers or patterns in the data associated with the disease. However, for analysing biological processes like apoptosis, the focus of interest is to discover molecular mechanisms within cells and hence consider a different observational level. Simulation and analysis of differential equation models aid in elucidating complex interrelations on a mechanistic level. Nevertheless, interpretation of how molecules mechanisms give rise to a cellular process is complicated, time consuming and the bottleneck of systems biology research.

In order to understand interactions and mechanisms of molecules involved in apoptosis on a systems level in the context of tumor development, this work combines analysis of patient data and analysis of molecular interactions. To achieve this aim, two different approaches were followed.

The first approach analyses gene expression data of different tumor entities for their ability to undergo cell death in silico by simulation of a mathematical model of apoptosis. The second approach identifies hypothetical conditions required for apoptosis. To this end, a novel method for model analysis was developed to reveal multivariate control mechanisms of model behaviour. The method automatically suggests model components and their crucial interrelations for a specific system response

Conditions concerning certain apoptotic molecules and molecule relations could be confirmed. Moreover, interactions and relations were found, which suggest new hypotheses how apoptosis as a system is governed by a few molecules in a specific way.

Zusammenfassung

Die gestörte Regulation des programmierten Zelltods (Apoptose) spielt eine wichtige Rolle bei der Krebsentstehung. Der Prozess, der unerwünschte und beschädigte Zellen eliminiert, basiert auf komplexen Molekülinteraktionen. Studien zur Untersuchung von Erkrankungen konzentrieren sich typischerweise auf die Analyse von Patientenproben, um auf diese Weise Marker oder Muster zu finden, die mit der Krankheit assoziiert werden können. Dagegen werden bei der Analyse biologischer Prozesse wie der Apoptose eine Ebene betrachtet, bei der Zellen und Moleküle im Mittelpunkt stehen mit dem Ziel molekulare Mechanismen aufzudecken. Dabei unterstützen Simulation und Analyse von Differentialgleichungsmodellen die Untersuchungen bei der Aufklärung der komplexen Zusammenhänge. Jedoch ist die Interpretation inwiefern molekulare Mechanismen funktionieren meist kompliziert und mit erheblichem Aufwand verbunden. Darüber hinaus stellt die Interpretation bezüglich molekularer Mechanismen einen Engpass in der systembiologischen Forschung dar.

Um Interaktionen und Mechanismen der beteiligten Moleküle zu verstehen und in den Kontext zu Tumorentwicklung zu stellen, wurde in dieser Arbeit die Analyse von Patientendaten und die Analyse molekularer Interaktionen verknüpft.

Dabei wurden zwei verschiedene Ansätze verfolgt. Zum einen wurden Genexpressionsdaten verschiedener Tumorentitäten analysiert und auf ihre Apoptosefähigkeit durch mathematische Simulation geprüft. Zum anderen wurden hypothetische Bedingungen unter denen Apoptose stattfinden kann durch Analyse mathematischer Modelle ermittelt. Zu diesem Zweck wurde eine neuartige Methode zur Modellanalyse entwickelt, die auf multivariate Kontrollmechanismen des Systemverhaltens hindeutet. Auf diese Weise werden nicht nur wichtige Komponenten des apoptotischen Systems sondern auch ihre mechanistische Beziehung in einem automatisierten Prozess vorgeschlagen.

Die auf diese Weise gefundenen Bedingungen bezüglich bestimmter Moleküle und Relationen zueinander bestätigen bekannte Interaktionen, die für Apoptose wichtig sind. Des Weiteren wurden Interaktionen und Relationen gefunden, die neue Hypothesen vorschlagen, wie der programmierte Zelltod als System von wenigen Molekülen in bestimmter Weise kontrolliert wird.

Table of Contents

List of Figures

List of Tables

1 Introduction

Biological processes are characterized by the dynamic behaviour of interacting molecules. These interactions are performed in an orchestrated fashion to perform a certain task or give rise to a specific cellular function.

One important function is apoptosis, or programmed cell death. The activation of apoptotic cell death is an essential mechanism of cells during the development and maintenance of a healthy organism to eliminate infected or defective cells. Impaired function of cell death plays an important role in development of several diseases. The evasion from apoptosis is considered a hallmark of cancer [42] and assumed to arise from disruption of the apoptosis pathway.

At the molecular level, several stimuli that trigger apoptosis have been identified. For the CD95/Fas-mediated apoptosis pathway, Bentele et al. revealed a threshold mechanism for cell death based on molecule levels and hence a cell's fate to be dependent on relative molecule concentrations [11].

Concentrations of molecules can be measured at the gene expression level by DNA microarrays. Microarray technology has been widely used to analyse the molecular state of tissue samples with regard to its corresponding phenotype. This technology has led to the discovery of markers for various diseases, for prognosis or to assist in choosing the appropriate therapy for a patient [34] [82].

In addition to finding individual genes that distinguish or characterize states or classes of samples, a *globaltest* [40] analyzes groups of genes towards their ability of distinguishing phenotypes and yields a significance measure for the association of the tested group with a phenotype or class. Tested sets can be pathways, i.e. groups of interacting proteins (or rather the genes that code for them) that act in a concerted manner to allow for flux of information or matter. High significance of the test suggests accumulation of changes within the tested pathway associated with the phenotype.

1.1 Scope of thesis

Signaling pathways of apoptosis are studied in this work based on the assumption that molecule concentrations are crucial for the decision whether to execute apoptosis. The dynamic behaviour of molecule interactions is simulated by mathematical models. These

models are constructed from ordinary differential equations (ODEs) and combined with microarray analysis of data sets from normal and tumor tissue in order to find mechanisms at the molecular level that decide on cell fate in association with a certain disease phenotype. Hypotheses for mechanisms at the molecule level for a certain model and its system response are generated towards apoptosis as the cellular function of interest and connected to a phenotypic disease state resulting in dynamic or mechanistic features for phenotypes instead of single genes or groups of genes representing a static "snapshot" of the disease state. To this end, two main approaches were followed.

1.2 Contribution and outline

In this work the level of phenotypic observation in terms of a cancerous or non-cancerous disease state and the level of molecular cell signaling are jointly investigated by focussing on apoptosis as the biological process essential for phenotype development (Fig 1.1).

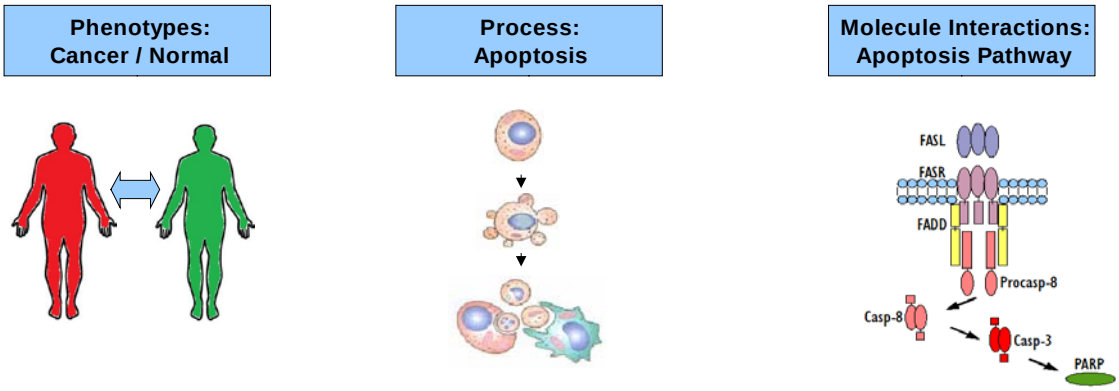

Figure 1.1: Observational levels integrated in this thesis
In order to understand interactions and mechanisms of molecules involved in apoptosis in the context of tumor development, the level of clinical phenotypes (left) is connected via the process apoptosis (middle) to molecular interactions on the pathway level (right).

By connecting phenotypes to mechanistic models of a system of interacting molecules, the aim of this work is to answer how a phenotype, in this work a state of disease, relates to the molecular interaction dynamics that bring about a certain process, here apoptosis and in that way find a discriminative pattern similar to markers, but as dynamic or mechanistic features for cancer.

To this aim gene expression analysis and mathematical modeling were combined.

Two main approaches were followed to bridge disease state and inspection of molecular interaction mechanisms.

The first approach questions if the molecular state measured by gene expression data can explain disruption of apoptosis in cancer by scaling input values of a mathematical model according to the observed gene expression levels of pathway components involved in apoptosis. This approach was performed in collaboration with Barbara Roider during a student project at the German Cancer Research Center.

The second approach elucidates which conditions with respect to molecule abundances could theoretically lead to a certain cell fate according to the analysis of a mathematical model of apoptosis. These conditions can be tested by analyzing how frequently they are met in data from normal tissue compared to tumor tissue.

For the second approach a novel method was developed. It detects relations of pathway component responsible for a defined systems (pathway) response in an automated fashion.

In chapter two of this thesis introduces apoptotic cell signaling by describing, pro- and anti-apoptotic molecules involved in apoptosis and how they are affected in different tumor types. Contemporary experimental as well as computational methods from systems biology are explained in chapter three. Chapter four describes mathematical models of cell signaling pathways and experimental data sets analysed for this thesis. Chapter five presents how mathematical models and data are combined to bridge the mechanistic level of molecule interactions and the phenotype level in order to understand interactions and mechanisms of molecules in the context of tumor development. The results from these approaches are then given in chapter six. This thesis concludes with discussion and outlook of the work in chapter seven.

2 Cell signaling and apoptosis

Coordinated cell action is based on interactions of molecules, for the most part of proteins, which constitute the cellular signaling system. Cell signaling enables communication between and within cells by signal perception, transmission and response of a cell to its environment. Signaling also plays a major role in cancer development. The six identified characteristics entitled as “The hallmarks of cancer”, self-sufficiency in growth signals, insensitivity to growth-inhibitory signals, limitless replicative potential, sustained angiogenesis, metastasis and evasion of programmed cell death (apoptosis) are largely based on cell signaling. This chapter comprises the basics about cell signaling during apoptosis and the importance of intact cell signaling to prevent carcinogenesis as well as examples for signaling pathway disruptions occurring in cancer development.

2.1 Apoptosis - Cell death: An essential process for life

Apoptosis is a type of cell death that has been described for the first time as a process involved in embryonic development by Carl Vogt in 1842. It is an essential process for tissue homeostasis, the development of organs and a multi-cellular organism [24] but has been also found to be implicated in a number of diseases, where either too much (e.g. in neurodegenerative diseases, Parkinson’s, Alzheimer’s, sinal muscular atrophy, AIDS) or too little apoptosis (e.g. in cancer or autoimmune diseases like diabetes type I and encephalomyelitis) takes place [61].

While necrosis is described as an unordered and accidental form of cell death in response to acute injury which leads to uncontrolled disintegration of cellular structures, release of cytosolic content and subsequent inflammatory response, apoptosis occurs without leakage of cell content and thus without inflammation [61]. Instead, apoptosis is characterized by its induction in individual cells and their clean elimination. Morphological characteristics of apoptosis are shrinkage and rounding of the cell, fragmentation of the nucleus, membrane blebbing and formation of membrane enclosed compartments called apoptotic bodies which are ingested by phagocytosis of non-apoptotic cells [50].

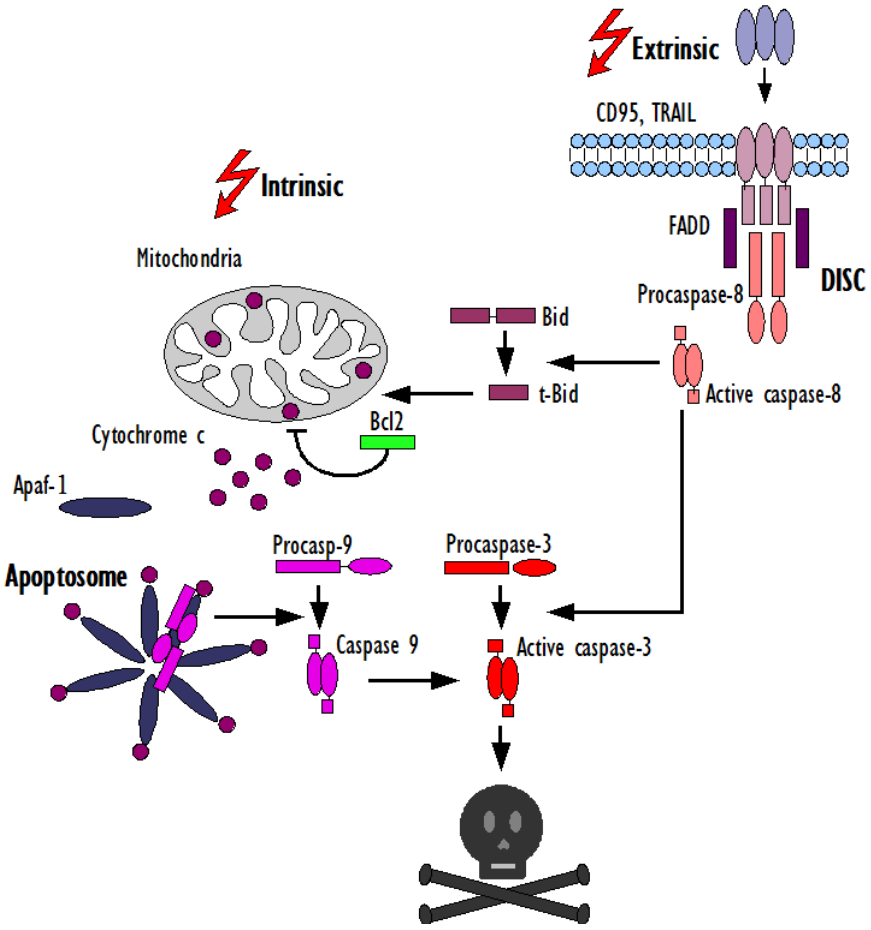

Figure 2.1: Extrinsic and intrinsic apoptosis pathway
For simplicity, the pro-apoptotic effect by Smac sequestering XIAP is not depicted in this figure.

2.1.1 The extrinsic and the intrinsic apoptosis pathway

On the molecular level, apoptosis is induced by cell signaling pathways. Two main pathways are common to mammalian cells: the extrinsic and the intrinsic apoptosis pathway.

The extrinsic pathway is triggered by extra-cellular signals e.g. ligand binding to death receptors situated at the membrane of the target cell, while the intrinsic pathway is caused by cellular stress inducing mitochondrial outer membrane permeabilization (MOMP) followed by release of apoptotic signaling proteins from the mitochondria into the cytoplasm. An overview of apoptotic pathways is illustrated in Fig. 2.1.

Extrinsic apoptosis is mediated by ligand-induced activation of death receptors which are part of the tumor necrosis factor (TNF) receptor super-family including eight members: TNFR1 (DR1), CD95 (Fas/APO-1/DR2), DR3, TRAILR1 (APO-2/DR4), TRAILR2 (DR5), DR6, EDAR and NGFR [59]. These death receptors are characterized by a pro-

tein-protein interaction domain in their intracellular region called death domain (DD) [33] which couples the death receptors to the apoptosis-inducing machinery.

Upon ligand activation, receptors recruit signaling molecules to their death domain. Among these molecules is FADD which, in addition to a DD also contains a death effector domain (DED). The DD of the receptor interacts with the DD of FADD whose DED enables interaction with molecules like initiator caspases or FLICE-like inhibitory proteins ($FLIP_{L/S/R}$/CFLAR) containing a DED [59]. The formation of a complex involving ligand, receptor and FADD able to bind procaspase-8/-10 and other proteins containing a DED is called death inducing signaling complex (DISC) [76] [51] and is found to be part of the CD95, TRAILR1 and TRAILR2 signaling pathway [77].

Binding of inactive initiator caspases (procaspase-8/-10) to the DISC leads to their activation and subsequent cleavage of more pro-death signaling molecules like effector caspases to accomplish cell death [96].

Activation of the intrinsic apoptosis pathway is regulated by members of the Bcl-2 family and requires mitochondrial outer membrane permeabilization (MOMP) followed by the release of cytochrome-C and Smac/Diabolo from the mitochondrial inter-membrane space into the cytosol. Smac (second mitochondria-derived activator of caspases), Apaf-1 (apoptotic-protease activating factor 1) and ATP/dATP then form the apoptosome, which, similarly to the DISC, acts as a complex to activate the initiator caspase-9 [10] able to cleave effector caspases.

Smac further relieves XIAP-mediated inhibition active caspase-3 [86].

While the intrinsic pathway is reported to be triggered as a response to intracellular signals due to cellular stress like DNA damage, hypoxia or cytotoxic insults [25], the extrinsic pathway is responsible for T-cell selection and elimination of virus infected or transformed cells [12].

The two pathways were shown to act independently [89], but also to be coupled through caspase-8 cleaving Bid, a pro-apoptotic Bcl-2 family member [97]. Both pathways converge at caspase activation for execution of cell death.

2.1.2 Caspases

Both, the intrinsic and the extrinsic pathway lead to activation of caspases (cystein-dependent aspartate specific proteases), which are considered as executioners of cell

death. Healthy cells contain several caspases as inactive precursors called caspase zymogens or procaspases [57].

The pro-apoptotic caspases can be divided into the group of initiator caspases involved in the initiation of the apoptotic response including procaspases-2, -8, -9 and –10, and the group of executioner caspases containing procaspases-3, -6, and –7 which cleave downstream many cellular substrates resulting in the characteristic features of cell death.

The executioner caspases possess only short prodomains, whereas the initiator caspases have long prodomains containing death effector domains (DED) in the case of procaspases-8 and –10 or caspase recruitment domains (CARD) as in the case of procaspase-9 and procaspase-2. Via their prodomains, the initiator caspases are recruited to and activated at death inducing signaling complexes either in response to the ligation of cell surface death receptors (extrinsic apoptosis pathways) or in response to signals originating from inside the cell (intrinsic apoptosis pathway).

The extrinsic apoptosis pathway includes recruitment of procaspase-8 by its DEDs to the death inducing signaling complex (DISC). Due to their close proximity to each other at the DISC, several procaspase-8 molecules become activated by autoproteolysis [28].

In contrast, procaspase-9 is activated after binding via CARD-domains to the apoptosome, a holoenzym consisting of seven Apaf-1 molecules formed after interaction with cytochrome C released from the mitochondria.

While the catalytic activity of initiator caspases follows binding and cleavage at initiation complexes (DISC and apoptosome), executioner caspases exist constitutively as homodimers and require proteolytic cleavage of an initiator caspase for activation [75]. The downstream substrates of executioner caspases are kinases, enzymes of the DNA-repair machinery and proteins involved in DNA-replication and splicing [66]. Caspases are catalytically active and even caspase precursors contain a small amount of catalytic activation, but are both kept in check by anti-apoptotic regulators.

2.1.3 Regulators of caspase activity and apoptosis

Cell death in terms of apoptosis is a central process for an organism. Therefore, mechanisms are needed to control the core of the apoptotic machinery represented by caspases. The activity of caspases is governed by a variety of regulators.

FLIP-isoforms

A group of molecules able to modulate caspase-8 activity upstream of the apoptosis pathway are the cellular FLICE- inhibitory proteins (cFLIP). Different variants of cFLIP exist. At the protein level, three isoforms have been detected so far: $cFLIP_L$, $cFLIP_S$ and $cFLIP_R$ [17]. All of them contain two DEDs which resemble the N-terminal part of caspase-8 and enable binding to the DISC [36].

However, the C-terminus of $cFLIP_L$ is longer than the one of $cFLIP_S$ and $cFLIP_R$ and similar to the structure of caspase-8, but lacks caspase activity. Still, the long variant is able to form heterodimers with caspase-8 and can be cleaved to p43cFLIP [17], whereas the two short variants, $cFLIP_S$ and $cFLIP_R$, both compete with caspase-8 for binding of FADD. Both, the long and the short variants $cFLIP_L$ and $cFLIP_S$ can protect cells from apoptosis induced by CD95, TNFR1, TRAILR1, TRAILR2 and DR3 due to competition with caspase-8 for binding sites. Also, the $cFLIP_R$-isoform has been reported as anti-apoptotic component acting similar to $cFLIP_S$ with regard to apoptosis.

The role of c-$FLIP_L$ has been discussed controversially to be anti-apoptotic at high concentrations [20] [56] or pro-apoptotic at low concentrations [20]. A recent study of Fricker et al. experimentally and theoretically revealed that not solely the amount of c-$FLIP_L$, but receptor stimulation influences the pro- or anti-apoptotic effects of c-$FLIP_L$ [36]. It has also been proposed earlier, that relative levels of $cFLIP_L$ and $cFLIP_S$ potentially determine a crucial balance of caspase activation and NFκB-signaling, the latter associated with survival [17].

IAPs (inhibitors of apoptosis)

A family of proteins both preventing activation of caspases as well as inhibiting catalytically active caspases are IAPs (inhibitors of apoptosis). Characteristic for this group of apoptosis inhibiting molecules are BIR (baculoviral IAP repeat) domains which are essential for their anti-apoptotic effect. Eight human IAPs have been identified so far [84], out of which XIAP is the best characterized protein and also the most potent inhibitor of cell death in vitro [18]. XIAP binds to the active-site pocket of effector caspase-3 and -7 via its linker region located between its BIR1 and BIR2 domains. In that way, the binding of substrate to caspase is blocked.

Inhibition of caspase-9 is achieved by a different strategy. Here, XIAP binds via its BIR3 domain to caspase-9 homodimerization surface and hence prevents caspase-9 from forming a catalytically active molecule [69].

Bar (Bifunctional inhibitor of apoptosis)

Another molecule able to interact with caspase-8 is Bar (bifunctional apoptosis regulator). Bar contains a DED domain and is reported to sequester and inhibit active caspase-8 hence it prevents further caspase cleavage and apoptosis [88].

In addition to a DED, Bar includes a zinc-binding RING (really interesting new gene) domain, a TM (transmembrane) domain and a SAM (sterile alpha motive) domain. The SAM domain is required for interactions with Bcl-2 family members and apoptosis suppression [98]. Thus Bar represents a protein able to interfere with components of the extrinsic and intrinsic pathway and is suggested to interrupt the amplification loop of caspase-8 via the Bcl2 family member Bid [81].

2.1.4 The Bcl-2 family

Mitochondrial outer membrane permeabilization (MOMP), the central event of the intrinsic apoptotic pathway, is considered the “point of no return” of apoptosis [22]. It involves cytochrome C and Smac release from mitochondria, components required for apoptosome formation and XIAP inhibition, both promoting caspase activation. Hence, careful control of MOMP is important and it is regulated by proteins of the Bcl-2 (B-cell CLL/Lymphoma 2) family [38] [97]. The Bcl-2 family can be categorized into three subgroups with regard to their structural elements and function during apoptosis: The anti-apoptotic members (BCL-2, BCL-X_L, BCL-w, A1, MCL1), the pro-apoptotic Bax-family (BAX, BAK, BOK) and the pro-apoptotic BH3-only family (BID, BIM, BIK, Bad, BMF, NOXA, PUMA) [75]. Structurally, the categorization into the subgroups is based on the number and type of BH-domains (Bcl-2 homology domains).

Anti-apoptotic members contain BH1-BH4 domains and can form a hydrophobic groove that enables interaction with the BH3 domain of a BH3-only protein. The anti-apoptotic BCL2 members, also known as guardians of mitochondrial integrity, have been shown to be potent inhibitors of cell death.

In contrast, molecules of the BAX-family are called ‘effector molecules of the BCL-2 family’. They are widely believed to act on the mitochondrion and to induce permeabilzation of the outer mitochondrial membrane triggering the release of intermembrane space proteins like cytochrome c [62]. The BAX-family members contain the BH1-BH3 domains. Murine cells that lack BAX and BAK fail to undergo MOMP and show resistance to several apoptotic stimuli [75] which indicates that the presence of BAX and BAK is needed to release cytochrome-c.

However, BH3-only proteins are not able to induce cell death in the absence of BAX and BAK and therefore clearly act upstream of the effector molecules [21].

The proteins BID and BIM of the BH-3-only family are considered direct activators of BAX and BAK via a transient interaction [23]. After cleavage of BID by caspase-8 to tBid [64] [65] and translocation to the mitochondria, it interacts with other BCL2-family members in order to enhance pro-apoptotic signaling [87]. The BH3-only protein Bid therefore is regarded as a mediator of the extrinsic and intrinsic apoptosis pathway. BID was also identified to interact with both BAX and BCL-2 and is associated with a higher cell death rate if abundantly present [94]. BAX was identified to interact with BCL-2 as well. Moreover, the ratio of BAX to Bcl-2 was found to increase cell death rate, with high BAX in relation to BCL-2 accelerating cell death [74].

Due to this and further findings, the balance between pro- and anti-apoptotic BCL-2 proteins have been suggested to determine cell fate [23] [74].

Nevertheless, the exact mechanisms of inhibition and activation of BCL-2 proteins for guarding or disrupting the mitochondrial membrane are not yet fully understood [62] [23].

2.1.5 PARP

Poly(ADP)ribose polymerase-1 is a nuclear protein involved in DNA repair. The N-terminal region of PARP-1 bears a DNA binding domain (DBD) of two zinc fingers that are necessary for recognition of DNA single and double strand breaks.

Binding to DNA strand breaks is followed by conformational change leading to catalytic activity of PARP-1 [14]. Activate PARP-1 catalyses the cleavage of NAD+ (nicotinamidadeninenukleotid) into nicotinamid and ADP-ribose. The latter is then attached to nu-

clear proteins including histones and PARP-1 itself leading to chromatin structure relaxation enabling the access to the damaged site for repair enzymes [**48**].

During apoptosis, it is specifically cleaved by effector caspases resulting in two fragments of 89 and 24 kDa [**19**] and this is considered as an indicator for apoptosis [58]. PARP fragments can therefore be used for detection of apoptosis in various cell types [55].

2.2 Apoptosis and cancer

The complex interplay of pro-and anti-apoptotic molecules of the apoptosis signaling network governs cell fate decisions. The alternative expression for apoptosis, "programmed cell death" indicates an orchestrated fashion of molecules to collaboratively trigger a cell's destruction. In fact, perturbed molecule interactions can crucially affect cell fate leading to either abundant or reduced apoptosis as observed in various diseases. In the case of cancer, too little apoptosis takes place [61].

The strategy of a healthy organism is to terminate cells with irreparable damage. This is considered a very efficient way to protect against tumor development since malignancies arise rarely with regard to the total number of cell divisions occurring during a lifetime [69] [42]. However, tumor cells have acquired the ability to inhibit or evade apoptosis which is also considered as one of the hallmarks of cancer [42]. Yet, in case the inhibition is restored, it has been reported that apoptosis can be triggered again [69]. Hence, the components of the apoptosis pathway represent promising points of efficient drug intervention [1]. For various cancer entities, different anti-apoptotic proteins are overexpressed and pro-apoptotic proteins down-regulated [27]. Table 2.1 gives an overview of pro- and anti-apoptotic components and their role in different types of cancer.

Table 2.1 Apoptotic pathway components perturbed in cancer
Perturbations of pro- and anti-apoptotic pathway components inhibiting cell death are observed in clinical phenotypes of cancer. Malignant cells have developed strategies for survival e.g. by over-expression of anti-apoptotic proteins and inactivation of pro-apoptotic proteins.

Proapoptotic Components	**Perturbation in Cancer**	**Examples**
Bax	Mutation	Hemic malignancies, colon carcinoma
Death-receptors	Deletion/mutation	Head and neck cancer
	Reduced expression	Hepatocellular carcinoma, melanoma, lymphoma
Caspases	Deletion	Lung carcinoma, neuroblastoma
	Reduced expression	Lymphoma
Bad	Hyperphosphorylation	Various malignancies
Apaf-1	Reduced expression	Human leukemic blast cells, bladder cancer
	Inactivation	Malignant melanoma
TRAIL-receptor1/2	Mutation	Metastatic breast cancer
Cytochrome C	Inhibited relase from mitochondria	Various malignancies
Antiapoptotic Components	**Perturbation in Cancer**	**Examples**
Bcl-2	Chromosomal translocation	Follicular lymphoma
	Overexpression	Breast carcinoma, gastrointestinal carcinoma
c-FLIP	Overexpression	Lymphoma, pancreas carcinoma, melanoma
IAPs	Chromosomal translocation	Marginal zone lymphoma
	Overexpression	Melanoma, neuroblastoma, breast and colorectal carcinoma
Kinases	Down-regulation of PI3K-inhibitor PTEN	Glioblastoma, melanoma, prostate carcinoma
	Overexpression	Ovarian carcinoma

3 Methods for analysing biological systems

Research of biological systems has been pursued in various ways. High-throughput methods for measuring omics data fostered data-driven approaches based on statistical methods for hypothesis generation. Statistical methods for data analysis of large scale or even genome-wide data generally identify single components, groups of genes or proteins associated with a certain phenotype. However, to understand cellular functions and how they arise from molecular interplay as a system, fundamentally different computational methods are required to examine structure and dynamics of molecule interactions. This chapter will give an overview of current methods, their underlying ideas, advantages and limitations.

3.1 Gene expression profiling

High-throughput gene expression profiling by microarrays enables data acquisition on a large scale by measuring thousands of genes at once and therefore allows the study of biological systems on a genome-wide level. Genes can be considered as the units of information for long-term storage within a cell which encode for proteins. Transcription is the process by which this information is used to synthesize a gene product, in this study a protein. To this aim, a gene's DNA sequence is first transcribed into mRNA which is then translated into a sequence of amino acids required by a functional protein. Gene expression can be regulated on different levels: By varied transcription, RNA splicing and silencing, translation and post-translational modifications of proteins. However, the levels of mRNA represent at the transcriptional level the activated genes and hence the relative amounts of protein that a cell is supposed to produce.

By means of microarrays, relative abundances of mRNA transcripts can be measured.

For a microarray experiment, RNA extracts taken from the tissue of interest are amplified, fluorescently labelled and hybridized to the microarray, which contains several thousands of spots of complementary DNA probes representing genes. After quantification of the fluorescence signal and normalisation, gene expression data is ready for analysis.

Microarray measurements are widely used to compare samples of different states or classes, for example healthy and tumor tissue, in order to find molecular markers as characteristics for a certain phenotype. In particular, differentially expressed genes as single

components, patterns of several genes or as a functional group are identified in several studies and used in order to classify or further elucidate phenotypes.

3.2 Identification of altered individual molecules

An easy and intuitive way to compare gene expression data of two phenotype classes is calculation of the fold change (*FC*) expression. The fold change is obtained by calculating the mean of a gene expression value for a certain gene g_i (or microarray probe) of all samples in one phenotype class (e.g. tumor tissue) divided by the mean of the second class e.g. control group of healthy individuals:

$$FC_{g_i} = \frac{mean(g_i^{class})}{mean(g_i^{control})}$$

For genes equally expressed in both classes, the *FC* equals 1. In case of down-regulation of gene *i* in the group of cases, the *FC* results in a value between 0 and 1 and in the case of over-expression a value exceeding 1 is obtained [54]. This intuitive measure was used for the first time for detecting genes in Sacharomyces cerevisiae [29].

Various statistical methods exist to detect differentially expressed genes which additionally yield a measure for significance of the finding [49]. However, the calculation of fold changes help in spotting interesting genes associated with a phenotype and gives an intuitive measure of how much a gene differs on average from measurement in the control group.

3.3 *Globaltest* for testing groups of genes

Data mining or statistical methods for identification of individual genes result in a list of components, which represent a characteristic pattern of genes associated with a certain phenotype, but often do not reveal functional dependencies. Databases, ontologies and literature help assign the obtained list of components to biological processes, functional groups or modules. For a given group of functionally related components, group testing methods exist which test for a significant association of the whole set with a phenotype.

The use of a group test could be motivated by a question like: Is there a differential gene expression between tumor and normal samples for the apoptosis pathway?

One kind of group tests is a *globaltest* for groups of genes [40]. This statistical procedure tests a global null hypothesis as an intersection of many particular null hypotheses. The underlying assumption of a single null hypothesis is that there is no differential expression for a specific gene. The assumption corresponding to the global null hypotheses is that all single null hypotheses are true. For a gene set of *k* genes, the *globaltest* is the intersection of all *k* individual null hypotheses. In case the global null hypothesis is rejected, the data provides evidence, that the selected gene set shows a different behaviour between the two observational classes. In the case of the *globaltest* presented by Goeman et al. [40] a significant association between a gene set and phenotype classes is identified by linear modeling. The *globaltest* results in a *p*-value for the whole group instead of a single gene and indicates significance of association as well as the ability to predict a phenotypic class using the group of genes.

3.4 Introduction to systems biology

Present-day systems biology aims at understanding biological systems, for example cells, and how their functions arise from the interplay of their components. It involves examination of structure and dynamics of molecule interactions as a whole, rather than isolated parts of a cell [52]. According to Westerhoff and Palsson, two lines of investigation led to contemporary systems biology. One of them starts from identification of individual molecules and aims at up-scaling to a simultaneous view on all molecules and their interactions. The second line originates from non-equilibrium thermodynamics and focuses on the formal analysis of new functional states arising from molecule interactions aiming at the discovery of general principles rather than being descriptive [95].

Based on these two lines of systems biology, two different approaches exist: Top-down and bottom-up approach. Top-down systems biology is characterized by the use of potentially complete, that is large, genome-wide data sets for identification of network interaction structures or modules employing phenomenological models. In contrast, bottom-up systems biology starts with the parts of the system formulating the behaviour of each component and integrating these formulations usually as mechanistic models to predict system behaviour [15].

For bottom-up systems biology, model design plays an important role. Focus and aim of research determines granularity, scope and formalism used for modeling a certain aspect or part of reality.

3.5 Formalisms for modeling molecular interactions

Modeling of biological processes helps to incorporate existing knowledge about molecule interactions. Knowing the interaction structure of a set of functionally related cellular components, computational or mathematical models can be constructed. The model representation can then be used to simulate the system's behaviour. Different modeling formalisms for describing pathways and their characteristics as a system taking also into account the interactions among the components differ in their ability to represent temporal and stochastic behaviour as well as in their level of granularity [37].

Graph representation of molecular interactions

A simple and intuitive way to represent knowledge about molecules and their interaction structure is a graph consisting of vertices and edges. Vertices correspond to molecules and edges refer to interactions among the molecules. Graphs enable analysis of the interaction structure by graph theoretic approaches. In that way, topological features like connectivity of compounds or network motives can be revealed [26] [93][71]. This kind of modeling approach lacks the temporal aspect of molecule interactions and therefore cannot reflect their dynamic nature. However, a graph representation is often the starting point for further studying a network.

Computational and mathematical models

For a more detailed analysis of pathways, Fisher and Herzinger [35] distinguish between computational and mathematical models. Computational models are suitable for pathway modeling if precise quantitative relationships or parameters are unknown. Examples for computational modeling are Boolean and Bayesian models.

Boolean models assign values of 1 or 0 to each node, which reflects a molecule's state in terms of on/off or active/inactive. A Boolean function or logic rule takes into account the state of the variables at a certain time-point in order to get an evaluation for the next time-

point. In that way Boolean models enable the study of the causal and temporal relationships on a coarse grained modeling level.

Another computational method for modeling based on a graph representation is a Bayesian network [37] [53]. While Boolean networks can only handle discrete values, the Bayesian formalism can also handle continuous values. Here, a node represents a random variable for the conditional probability distribution of each pathway component. Bayesian modeling offers the ability to describe stochastic processes and dealing with uncertainty, incomplete knowledge and noisy observations. Limitations are the static and acyclic nature of Bayesian networks.

Another graph-based modeling approach is a Petri-net. Here, the set of nodes consists of places representing molecule species and transitions representing reactions. Concentrations of molecules correspond to so-called tokens associated with each place. By firing rules, tokens are transferred through the graph from one place to another simulating a flux through the network. Originally, Petri nets were restricted to qualitative simulation with discrete time steps. However, advanced Petri nets are able to mimic Boolean, Bayesian and even systems of ordinary differential equation.

A system of differential equations enables to describe changes in molecule concentrations in a quantitative, time-continuous fashion and therefore it is considered a fine grained modeling formalism. Computationally more demanding are partial differential equations (PDE) and the chemical master equation (CME), which are applied in case either spatial phenomena or stochastic effects of low molecule number fluctuations are of interest.

However, the common method for representation and simulation of molecular interaction network dynamics is a deterministic system of ordinary differential equations (ODEs). These equations describe changes in abundance of molecule species, where parameters of the system have a physical interpretation like concentration, reaction rate or binding affinitiy [5]. Hence, ODEs reflect physical reality of molecule dynamics to a high degree with parameters corresponding to experimentally measurable values.

3.5.1 Mathematical modeling of chemical reaction systems

For modeling a biological process by means of ODEs, the change in concentration of a single chemical species x_i e.g. (a protein) is described as a function depending on time.

During a biological process, several chemical species and reactions are involved. Biochemical reactions can be expressed by reaction equations of the general form [53]:

$$(-s_1)C_1 + \cdots + (-s_p)\,C_p \rightarrow\ (s_{p+1})C_{p+1} + \cdots + (s_{p+q})C_{p+q}$$

Here, the components $C_1,\ldots,C_p$ are educts and $C_{p+1},\ldots,C_{p+q}$ are products. The stoichiometric coefficients of the reaction are represented by s_i, where $s_{1,\ldots,p}$ take negative and $s_{p+1}\ldots s_{p+q}$ positive values.

In case one or more educts interact directly to form products within a single step, this type of interaction is called elementary reaction. For an elementary reaction the notion of some mechanistic significance between reaction participants (e.g. collision) is assumed [79], which is highly improbable for three or more molecules. Therefore, a system of reactions is decomposed into elementary reactions in which only one or two educts generate a product in a single reaction step. Assuming temperature, pressure and density of the reaction mixture to be uniform, we can define the reaction rate *v(t)* or velocity of a reaction according to the law of mass action to be proportional to the product of involved components. Hence the change in abundance per time unit of an elementary reaction can be defined as:

$$v(t) = k \prod_{Reactants} x_i(t)^{-s_i}$$

where k is called rate coefficient or rate constant of the reaction, x_i the concentration of the molecular species C and s_i the stoichiometric coefficient of the reactant i [7]. Hence, the reaction rate *v(t)* is proportional to the product of components involved in the reaction.

The temporal change in concentration of x_i is then obtained by linear combination of reaction velocities of elementary reactions involved in forming x_i multiplied by the stoichiometric coefficients s_{ij}:

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} s_{ij} k_j \prod_{Reactants_j} x_i(t)^{-s_{ij}}$$

This statement represents a differential equation expressing the change of concentration for one component x_i for a matrix of stoichiometric coefficients given by

$$\begin{pmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{m1} & \cdots & s_{mn} \end{pmatrix}$$

where *m* is the total number of chemical entities and *n* the number of reactions.

3.6 Methods for model analysis

With a system of differential equations at hand that describes a biological process, exploration and prediction can begin. Execution of ODE-models yield time dependent concentrations of simulated molecular species which corresponds to a specific experimental setting. By varying parameters, different settings can be simulated, which can correspond to experiments under changed conditions (e.g. changed ligand concentration) up- or down-regulation and increased degradation or generation of molecules. The output of these parameter variations in terms of trajectories for changing component concentrations can then be visually inspected.

However, computational methods for model analysis enable us to investigate a model's input-output behaviour in various ways. Methods for model analysis differ in their ability to determine qualitative or quantitative, local or global and uni- or multivariate characteristics of a system's behaviour. Basic ideas involve the concepts of robustness, stability, uncertainty and sensitivity for either maintaining or perturbing a certain process or state. Results obtained by analysing a system's behaviour can be helpful for the modeling process (e.g. for model reduction and model specification). The focus in this work however, is to discover a model's interaction structure and how it governs the system behaviour related to a biological function. Therefore, basic concepts in this section will be explained

with regard to how these methods aim at identification of conditions which are important for a cellular process.

Qualitative methods

Qualitative analyses aim at revealing single parameters, regions in parameter space or stimulation patterns as system input that lead to changes with regard to steady-states and their stability behaviour [8] [32] [63] [70]. In the case of steady states, the partial derivative with respect to time is not changing i.e. the left hand side of the last equation in section 3.5.1 equals zero, which means that molecule concentrations do not change over time.

Linear ODE-models can be solved analytically for steady state solutions. More complex models require computational methods which help finding important factors affecting steady state behaviour.

Qualitative methods associate certain steady-states of the system to distinct cellular states like death or survival of a cell [8] [32] [9] [63] and assume bistability to be required for decision making of a cell. Stability refers to attraction to a steady state due to parameter variation. In the case of bistability, the system can alternate between two stable steady states similar to a toggle switch [6].

Bifurcation analysis can be applied to identify important parameters which affect the number of steady states and identify which parameter value is associated with a certain state and its corresponding stability behaviour.

Quantitative methods

In order to determine essential parameters to influence signaling strength, signal duration or overall effect of a system's output, quantitative methods for model analysis are used. A widely used quantitative method is sensitivity analysis. It apportions the change in model output to the change of model parameters and in that way identifies most influential factors [67] [83]. Another quantitative method which determines the most important reactions or proteins is control analysis [45], an extension of metabolic control analysis to signaling pathways.

Global and local methods

Both qualitative and quantitative analysis can be local or global. Local sensitivity analysis focuses on a particular point in parameter space and varies a single parameter at a time to detect its effect on the system's output. Local analyses are easy to implement and interpret and therefore common, but are restricted to one specific point in parameter space and further do not account for parameter interactions [30]. Only recently, also global analyses are applied to biochemical models. Global procedures involve random sampling of parameter space to explore the whole range of possible parameter values and its corresponding effects. Since rate constants and molecular concentrations vary in an interactive manner, simultaneous parameter variations as part of global sensitivity analysis should be preferred over local methods for investigation of biochemical models. Examples for local and global sensitivity analysis applied to biochemical network models can be found in [91] and [30].

Methods to identify multi-parameter effects

Both, qualitative and quantitative model analysis methods are usually univariate that is the change in system behaviour or response is analysed with respect to variation of one parameter at a time and identification of single parameters as important factors. However, global methods for sensitivity analysis detecting higher order sensitivities are able to capture possible parameter interactions [30] yet fail to suggest underlying mechanisms.

An example for detecting interactions of parameters is parameter clustering, which identifies factors depending on a set of parameters, however aims at system identification [11]. Control analysis [45] is able to point out important reactions instead of single proteins, but fails to explain underlying mechanisms.

Direct Lyapunov exponent (DLE) analysis can be regarded as an example for capturing multi-parameter effects in the sense of investigating how a system governs or controls a certain system response by suggesting an explicit mechanism. DLE analysis is a multivariate local method for sensitivity analysis able to handle transient signals. It aims at finding conditions on component relations resulting in a certain system response. DLE analysis however, requires visualisation of the phase space, that is, the set of initial chemical species concentrations [4]. For a large model with a high dimensional phase space, visual inspection of DLE values can be tedious.

Not much work has been done in the field of multivariate model analysis of ODE systems. Nevertheless, multivariate computational methods for analysis of coupled ODEs are highly desirable, since they take into account changes of several parameters to give rise to a certain system response and in that way help elucidate intertwined interactions and mechanisms at the molecular level.

Hence, an automated multivariate method for model analysis that takes into account changes of several parameters with regard to a certain system response is desirable. In that way, intervention strategies involving several co-operating components instead of single perturbation points could be identified to cause a certain system outcome.

3.7 Decision trees

Decision trees originate from two different branches of research. Developed in the field of statistics by Breiman et al. [13] as classification and regression trees (CART) in 1984 and in the area of artificial intelligence by Quinlan [78] as a method called ID3 in 1986 followed by later versions named C4.5 and C5.0, the methodologies differ only slightly and share the fundamental concept. The basic idea of a decision tree is to build a classifier which partitions the input space R^r into cuboid regions (see Fig. 3.1A) for predicting an output variable Y. The classifier is graphically represented as an acyclic graph (Fig. 3.1B), called tree which contains nodes and edges and can be translated into a corresponding rule set (Tab. 3.1).

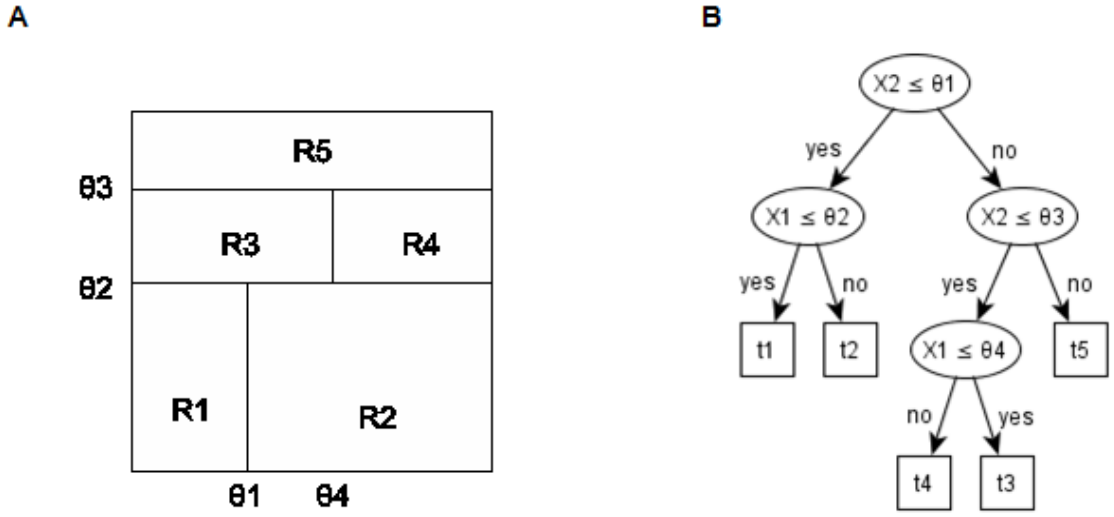

Figure 3.1: Input regions and decision tree representation
A: The partitioning obtained from a decision tree algorithm corresponds to partitioning the input space into regions according to threshold values. Each region is assigned a leaf node. **B:** Graph representation of a decision tree.

Table 3.1: Rule set corresponding to decision tree of Fig. 3.1

Path	Rule
[1]	If X2 ≤ θ1 and X1 ≤ θ2 then t1
[2]	If X2 ≤ θ1 and X1 > θ2 then t2
[3]	If X2 > θ1 and X2 ≤ θ3 and X1 ≤ θ4 then t3
[4]	If X2 > θ1 and X2 ≤ θ3 and X1 > θ4 then t4
[5]	If X2 > θ1 and X2 > θ3 then t5

Two different types of nodes are part of a decision tree: decision nodes and leaf nodes. A decision node represents a criterion used for testing a variable X_i from a set of variables $X_1, X_2,\ldots, X_D$. A test could be a question like: Is the value of variable X_i lower than θ_i? The decision node at the top of the tree containing only outgoing edges is called root. In contrast, a leaf node contains only incoming edges and appears at the bottom of the tree representing a class which is associated with a certain partitioning based on a sequence of tests. By following a particular path from root to bottom where each edge corresponds to a certain value of the variable tested upstream, an item can be classified.

The decision tree methodology according to Breiman *et al.* involves growing a maximal tree, computing an estimate for the misclassification error, pruning the maximal tree to a set of subtrees and choosing the best pruned subtree. The whole procedure is explained in detail in the following section.

3.7.1 Construction of decision trees for classification

The construction of a decision tree requires a labelled data set

$$L = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$$

also called historic data, learning sample or training data, where x_n is one object or vector measured in the input variables $X_1, X_2,\ldots, X_D$ with $n = 1,\ldots,N$. The measurement vector can consist of variables of two general types. Either a variable is numerical in case variable values are real numbers or a variable is categorical in case values of a finite set without any natural ordering can be taken. The object's corresponding class label is expressed by y_n which can take a value $k \in \{1, \ldots, C\}$. In case of a binary classification problem, an

item can belong to either of two classes i.e. $k \in \{1, 2\}$. The tree classifier is then constructed by recursively partitioning of the training data set of measurement vectors into "purer" subsets ideally containing vectors of one class. To this aim, every possible value of each variable is considered and evaluated to choose the criterion for the best split.

The goodness of split is evaluated by an impurity function, the gini index in case of CART [44][13]:

$$i = \sum_{k=1}^{C} p(k|t)(1 - p(k|t))$$

where *p(k|t)* is an estimate of the conditional probability that an observation *x* is in class *k* given that it falls into node t [47], which can be reduced for two classes to

$$i(t) = 2p(1 - p)$$

for *p*, the proportion of the second class. Since the aim of a good split is to partition the data into descendant subsets L_L, L_R that are more homogenous than the parent subset with regard to their assigned class, a split is chosen which yields the largest decrease in impurity i.e. the one that maximizes:

$$\Delta_i(s, t) = i(t) - p_L \cdot i(t_L) - p_R \cdot i(t_R)$$

For p_L and p_R, the proportions of items that go into the subset L_L and L_R respectively resulting from a candidate split *s*. The criterion for splitting is illustrated within the tree representation as a decision node. Recursive splitting of the learning sample can be done until all items are split into subsets containing only representatives of one class or until a certain stop criterion is reached e.g. a minimum number of cases falling into one node.

In order to evaluate a constructed tree T, the estimate of the resubstitution error at node *t* is defined as

$$r(t) = 1 - max_k p(k|t)$$

and for tree T with its set of terminal nodes $\tilde{T} = \{t_1, t_2, \dots, t_L\}$:

$$R(T) = \sum_{l=1}^{L} r(t_l)p(t_l)$$

with $p(t_l)$, the proportion of all items that fall into node t_l. The resubstitution error can be regarded as the error rate on the learning sample. In order to choose a classifier, which does not overfit the data, Breiman *et al.* suggest to grow the full decision tree T_{max} first and then reduce the tree by cost-complexity pruning. This pruning method involves the construction of a set $T_0, T_1, \dots, T_M$ of subtrees of T_{max} where $T_0 = T_{max}$ and T_M is the root tree. Subtree T_{i+1} is then obtained from T_i by removing those branches that account for the lowest increase in error rate per pruned leave. Let L_{T_i}be the number of leaves of T_i. A tree T pruned in a node t by removing the branch T_t which has t as root increases its error by the amount $r(t) - r(T_t)$ and decreases its number of leaves by $L_{T_{i-1}}$. Hence, the ratio $\alpha = \frac{t(t)-r(T_t)}{L_{T_{t-1}}}$ represents the increase in error rate per pruned leave. The tree T_{i+1}results from pruning all nodes in T_i with the lowest value of alpha. T_i and T_{i+1} are characterized each by a certain value α_i and α_{i+1} respectively, with $\alpha_i < \alpha_{i+1}$. Finally, the best tree of the set $T_{max}(\alpha) = \{T_0, T_1 \dots, T_M\}$ can be chosen by cross validation. A method called *n*-fold cross validation refers to the procedure of splitting data *D* into *n* roughly equal parts and use each subset $D_i, i \in n$, for estimating the performance of a classifier that was previously trained on all other data sets $D_j, j \neq i$.

In the case of decision trees, Breiman *et al.* suggested as the final classifier the simplest subtree that is within one standard error of the subtree yielding minimal misclassification cost on the test set. The resulting tree is then regarded as the tree pruned to the best level [13]. Further pruning can be also done in order to obtain a more comprehensible tree and hence can be regarded as "trading accuracy for simplicity" [68].

Decision trees are very popular in the biomedical regime because of their intuitive, easy understandably nature. Some further favourable properties of CART are [90]:

- CART is non-parametric and does not require specification of any functional form
- CART does not require selection of variables in advance, since the algorithm identifies most significant variables and eliminates non-significant
- CART results are invariant to monotone transformations of its variables
- CART can easily handle outliers, isolating them into a separate node

4 Bridging phenotypes and molecular mechanisms

Contemporary experimental and computational methods explained in chapter 3 enable the analysis of biological systems on different levels. With genome-wide data at hand, it is possible to identify single genes and pathways, yet mechanisms at the molecular level stay obscure. For investigation of pathway dynamics and molecular mechanisms, dynamic modeling and model analysis are needed (Fig. 4.1).

In order to understand interactions and mechanisms of molecules at a system level in the context of tumor development with a focus on apoptosis using gene expression data, two main approaches were developed to bridge clinical phenotypes and pathway mechanisms which will be explained in the following sections.

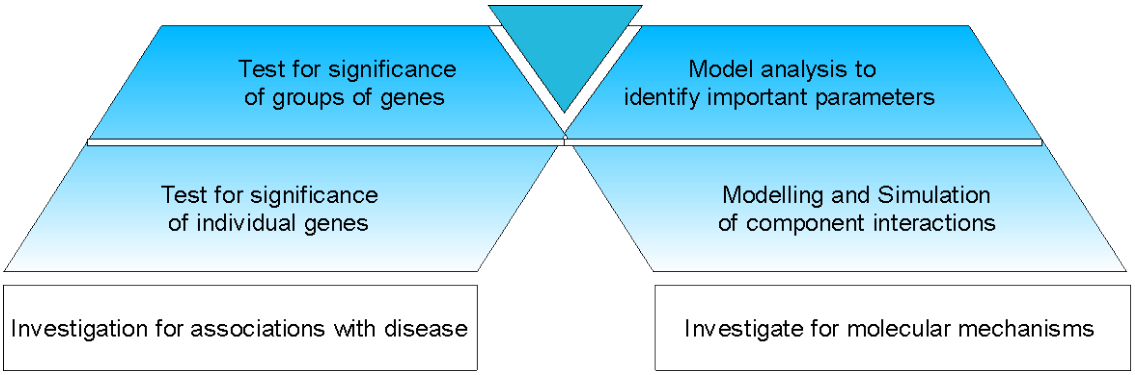

Figure 4.1: Bridging clinical phenotypes and molecular mechanisms
Current computational methods enable investigation of individual or groups of genes associated with a disease or alternatively investigation of molecular mechanisms. How to bridge clinical phenotypes and molecular mechanism by novel computational methods in order to understand cancer on a systems level is explained in this chapter.

4.1 Approach 1: From data to model

The first approach addresses the question: Can gene expression data explain why apoptosis is perturbed in cancer? Given gene expression data sets with a putative disruption in one class of samples (here cancer) with regard to a pathway governing the process of apoptosis, this approach investigates if gene expression data reflect underlying conditions for either death or survival and consequently gives rise to a certain system response.

For this approach, original start values of an ODE-based apoptosis model are scaled according to gene expression data from tumor and normal tissue. The employed model was developed by Albeck et al. [3] who studied apoptosis by observing the concentration of

PARP and its respective time course as an indicator of effector caspase activity. It is has been shown that caspase-3 activity follows a dynamic of the form [80]:

$$c(t) = f - \frac{f}{1 + e^{(t-T_d)/4T_s}}$$

Here, *c(t)* is the amount of substrate cleaved at time *t*, *f* is the fraction cleaved at the end of the reaction, T_d is the delay period between ligand (disposure) and half maximal cleavage ($c(t) = 0.5\, f$) and T_s is the switching time between initial and complete effector substrate cleavage. A short switching time of less than 30 minutes has been previously associated with apoptotic cell death and was calculated based on PARP time courses by Albeck et al. [3]. Based on these findings, we calculated delay time T_d and switching time T_s of PARP trajectories resulting in a system response of component start values scaled by gene expression data of tumor and normal samples.

This first approach is complemented by traditional microarray analysis, both, of single genes and the whole set of pathway components by a *globaltest*. Two different methods for scaling start values were applied: Scaling by fold change and scaling per sample.

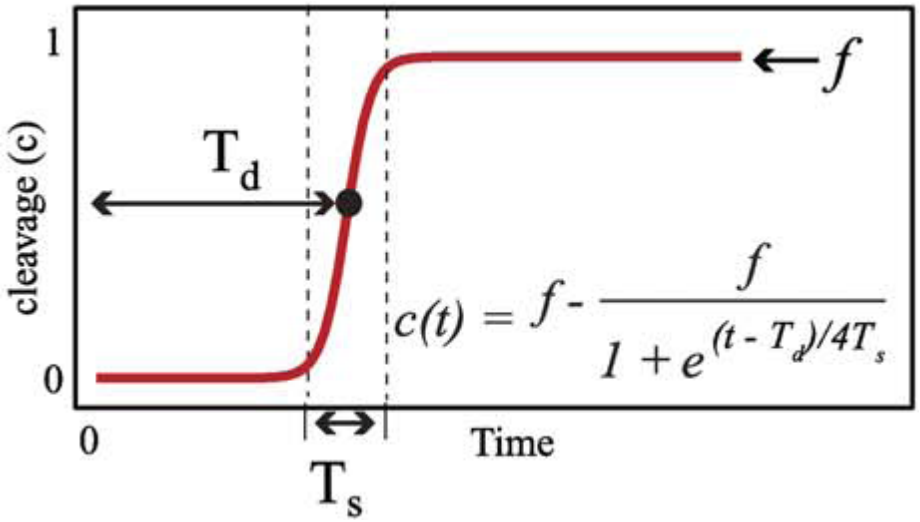

Figure 4.2: Delay time T_d and switching time T_s
Delay time T_d and switching time T_s are characteristics of effector caspase-3 activity according to Albeck et al [3] and are used in this thesis for analysis of the system output. T_d and T_s were calculated based on PARP trajectories where a short switching time has been previously associated with apoptotic cell death [3]. Figure was taken from Albeck *et al.* [3].

4.1.1 Method 1: Scaling by fold change

The motivation for this scaling method was modeling the average difference between normal and tumor tissue represented by the fold change of the gene expression values. The relative nature of gene expression measurements by microarray data enables the use of the data as scaling factors for original start values. Although the fold change is calculated for each gene separately, an implicit change of molecule relations is assumed to be introduced by the scaling. Since most of the original start values of Albeck et al. [3] were estimated by using HeLa cells, these original values are considered to represent the class "tumor", and the fold change-scaled values are regarded to represent the class "normal". In case several probes were corresponding to one model component, the median of the probe values was calculated to represent the gene's expression value. In case, a gene expression value was not measured for a model component, the original start value was taken as the initial concentration. As a system response, a trajectory of model component PARP was obtained from simulation of scaled start values and compared to the trajectories arising from the original start values.

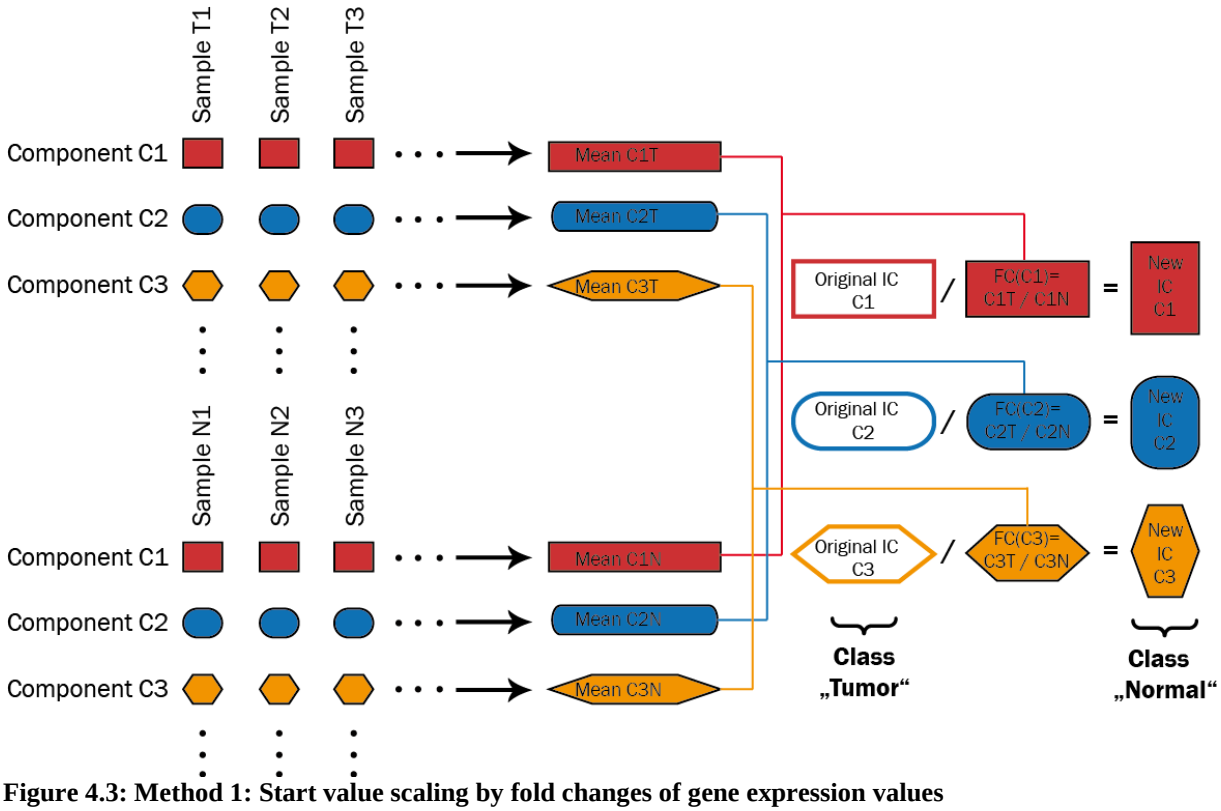

Figure 4.3: Method 1: Start value scaling by fold changes of gene expression values
For simulation of the average difference between normal and tumor tissue, the fold change (FC) of gene expression values for each model component (C1, C2, C3 …) is calculated. The resulting value (FC(C1), FC(C2), FC(C3)) is then used for scaling the original initial concentrations (IC) representing the class "tumor" (T) in order to obtain the initial concentration representing the class "normal" (N).

4.1.2 Method 2: Scaling per sample

In order to avoid averaging of samples of one class and get a more differentiated system response, a second scaling method was applied. Here, the molecule relations are scaled per sample by keeping one model component to its original start value and scale all others in relation to the fixed value. The start value to be unchanged was set for all calculations to be the PARP component. Therefore this scaling corresponds to normalisation of each sample to the start value of the model component PARP.

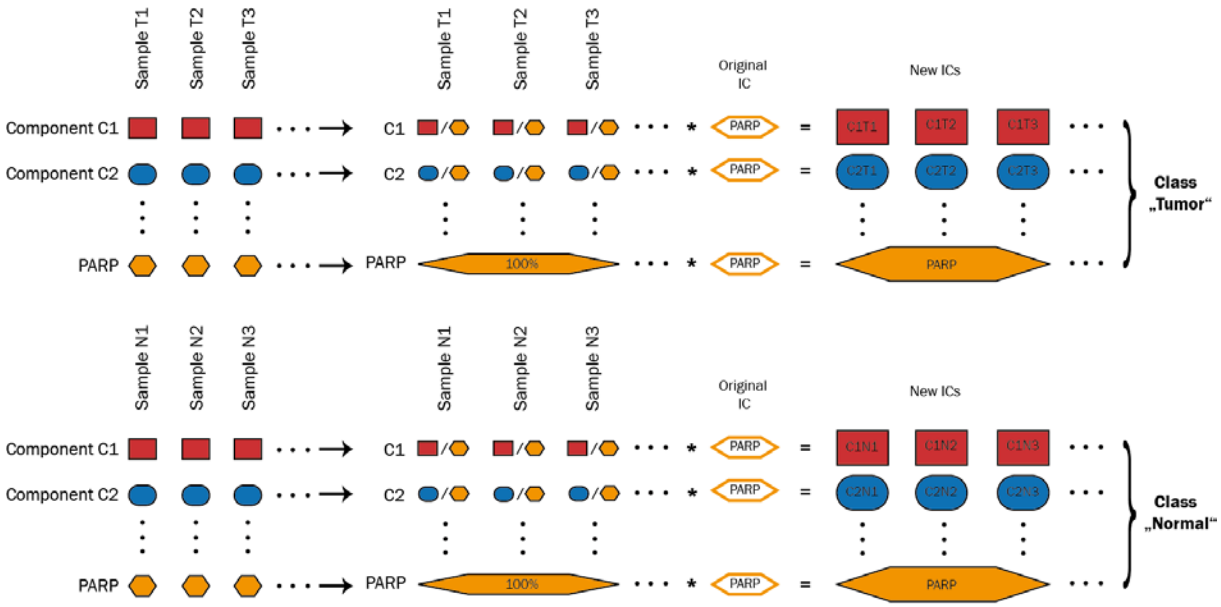

Figure 4.4 Method 2: Scaling of start values by individual samples
For scaling original start values using individual samples of the classes tumor (T1,T2,T3…) and normal (N1, N2, N3…), gene expression values of each sample were scaled relative to PARP by setting gene expression of PARP to 100%. In that way scaling factors for model components (C1, C2, C3,…) were obtained and used to calculate start values for every model component by multiplication with the original initial concentration (IC) of model component PARP. This procedure enabled apoptosis reflecting component relations of individual samples.

4.1.3 Relations of pathway components

Based on the assumption balanced pro- and anti-apoptotic relations of pathway components and threshold mechanisms to be crucial for signal transduction [11], a matrix of probe set ratios was constructed for each gene expression data set. To this aim, the ratio of each probe intensity to every other probe intensity corresponding to the model compo-

nents of Albeck et al. [3] within one data set was calculated (pobe X / probe Y). The result of this ratio was encoded as

0 in case (probe X / probe Y) < 1

1 in case (probe X / probe Y) = 0

2 in case (probe X / probe Y) > 1

In that way, the set of probes reflecting gene expression as features for a class where mapped to a new feature space of probe relations. The relative frequencies of {0, 1, 2} for tumor and normal samples respectively were then determined and the frequencies of discretized ratios that deferred the most for the two phenotype classes were identified.

Intuitively, anti-apoptotic components are expected to be more frequently higher expressed than pro-apoptotic components in tumor samples, since cancerous cells are assumed to be resistant to apoptosis.

4.1.4 Complementary traditional analysis of gene expression data

Fold changes were calculated to reveal single genes to be up- or down-regulated and further assist to categorize differences between tumor to normal data to either have changes rather affecting the intrinsic or the extrinsic parts of the apoptosis pathway.

Further, a *globaltest* on the selected probe sets corresponding to the model components of Albeck et al. [3] was performed in order to test for significance of this gene set for discriminative gene expression pattern of tumor versus normal data. A *p*-value below 0.05 for the tested pathway components will be considered in this study as indicator for significant association to the phenotype. The *p*-value is calculated based on a permutations method [39] and can be computed by means of the R package *globaltest*.

4.2 Approach 2: From model to data

The second approach tests whether computationally identified conditions are reflected in experimental data, which can potentially lead to apoptosis. This approach involves the development of a novel multivariate method for model analysis identifying strategies towards reaching a certain system response or cell fate.

The developed method was applied to two small models, one of them reflecting EGF-receptor internalization and the other one representing core pathway interactions for cas-

pase-3 activation. Since these models have been previously analyzed by different model analysis methods, these systems were used for evaluation and comparison of methods. Further, two apoptosis models with biological relevance were analyzed. For these systems, predicted conditions were checked in gene expression data as well as in literature with regard to the molecule levels of predicted pathway components to activate or inhibit apoptosis.

4.2.1 A novel method to multivariate model analysis

Given an ODE-model of a biological process, methods for model analysis aim at identifying system inherent properties and parameters that govern a certain system response. Since a system's behaviour is influenced by the intertwined interaction of several components, particularly in the case of cellular decision making, a novel multivariate method for model analysis that identifies conditions on model parameters controlling the system response was developed. The method simulates an ODE-based model representing a biological process for randomly generated parameter sets and further applies a decision tree algorithm for finding patterns in the parameter sets that lead to a certain model outcome. The workflow of the method is sketched in Fig 4.4 and comprises five steps:

1. Defining system response classes

The first step of the method is defining response classes of the model outcome. Since the method is applied to systems of signaling pathways, we are interested in a binary outcome represented by the classes 0 and 1 for 'pathway off' or 'pathway on' respectively. According to the shape of the component output trajectory of interest, mathematical criteria are defined to classify a system response as either class 0 or 1. Example trajectories representing activation classes of a pathway for EGF-receptor internalization (see chaper 5) are shown in Fig. 4.6. In this case the defined criterion for activation of the component of interest is to reach at least 80% of another component amount.

However, it is important to note, that classification criteria are adapted specifically to the analysed model and scientific question.

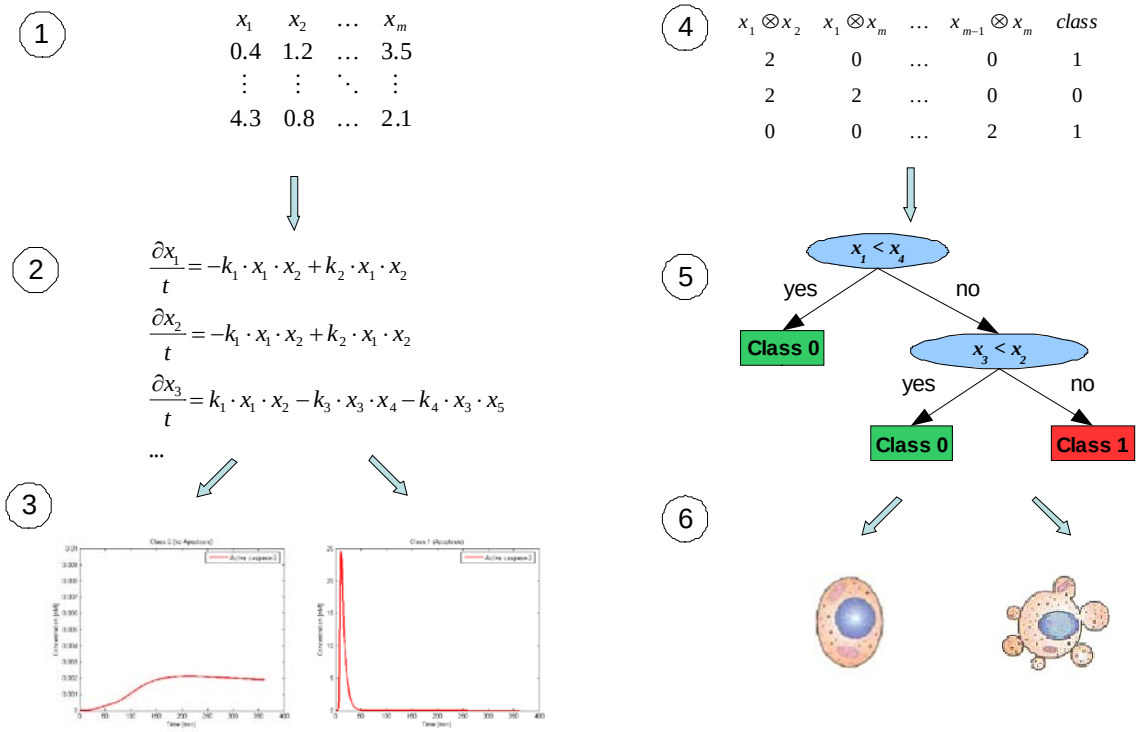

Figure 4.5: A novel method for multivariate model analysis.
(1): Random initial values for model components are generated and (2) used as input to a system of differential equations. (3) The system response as a result from each simulation is classified to a predefined criterion in the classes "pathway on" and "pathway off" represented by 1 and 0, respectively. (4) A matrix of start value relation is constructed and used for construction of a decision tree (5). The activation classes can be associated to a cellular phenotype (6).

2. Generation of random initial values

The underlying assumption for the approach is that molecule abundances influence pathway activation. Hence, only parameters corresponding to initial values for model components that represent molecular species are considered. A matrix of random initial values drawn from a probability distribution within considerable ranges is generated, in which each column corresponds to a specific component having a non-zero start value at simulation time t_0 and each row vector relates to one parameter set of initial values used as input to the ODE-system for one simulation run. Model components that require initial values are generally un-cleaved and un-complexed molecular species and typically represent proteins corresponding to gene products. As a probability distribution, an equal distribution was chosen. In general, range for sampling took of four orders of magnitude centered around original start values.

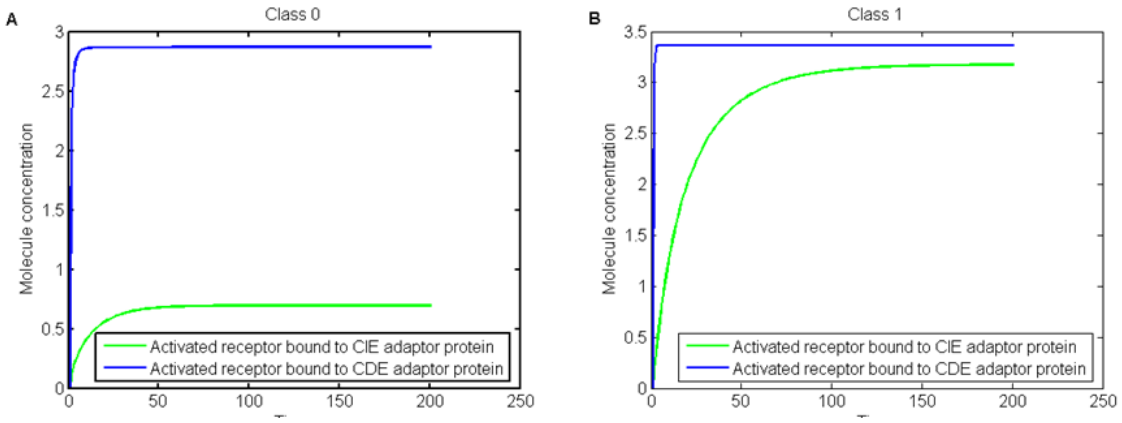

Figure 4.6: Pathway activation classes to categorize a system response

3. Model simulation and classification of generated parameter sets

Each row of the generated matrix in the previous step is used as input to the model. For simulation, the Matlab programming environment with its ODE-solver *ode15s* was used. Resulting output trajectories of interest are then classified according to the predefined criteria in step 1.

4. Construction of a matrix of components relations

The hypothesis, that threshold mechanisms of molecule concentrations are crucial for a certain system response motivates the construction of a matrix containing pair-wise start value relations based on the matrix of initial values. To this end, for each parameter set, the relation of all unordered pairs of non-zero molecule species X and Y was encoded as '0' in case X_0 < Y_0, as '1' in case X_0 = Y_0 and '2' in case X_0 > Y_0. In that way the numerical values of initial conditions are transformed into categorical attributes of a parameter set. Since equal values for randomly generated numbers are highly unlikely, this procedure comes down to binarization of start value relations used as features for the decision tree algorithm.

5. Application of a decision tree algorithm to component relations

The matrix of component relations together with the class information obtained from step 3 are then used as training data in a decision tree algorithm that yields a tree representation of molecule relations leading to a certain system response. For construction of a decision tree as explained in section 3.7, the algorithm developed by Breiman et al [13] as

part of Matlab programming environment was used. The performance of the tree was calculated by using a resubstitution method and a 10-fold cross validation [13] expressing the cost of the tree by means of the misclassification error.

An equal number of representatives of both system response classes was selected before applying the decision tree algorithm to the parameter sets in order to build an unbiased classifier. The method yields an intuitive tree representation of conditions leading to a certain system response and was applied to the models explained in chapter five.

5 Mathematical models and experimental data

In this chapter, data sets and mathematical models employed for this study are described. Within the first approach, gene expression data sets from different cancer entities representing clinical phenotypes with a putative disruption of the apoptosis pathway were analyzed and compared to data from normal samples to evaluate their ability to undergo apoptosis during model simulation. For this approach, the currently most comprehensive model of apoptosis developed by Albeck et al. [3] was simulated with its original start values scaled according to the data as explained in section 4.1.1 and section 4.1.2.

For the second approach, aiming at the discovery of conditions leading to a specific system response, four ODE based models of cell signaling were analyzed: A model for EGFR internalization, a caspase activation model, an apoptosis model for CD95/Fas-induced signaling and the comprehensive apoptosis model by Albeck et al. [3]. Conditions predicted based on the apoptosis models were checked in literature, in case of the latter model, also gene expression data was analyzed with regard to predicted conditions.

5.1 Data sets and gene expression data analysis

Eight gene expression data sets from six different cancer entities (Table 5.1) all of them measured on one of the Affymetrix platforms (hgu133a, hgu95 or hgu95av2) were analyzed. Each of the data sets was normalized by vsn [46] and probes corresponding to apoptosis model components of Albeck et al. [3] were selected.

Table 5.1: Gene expression data sets analyzed in for this thesis

Data Set	Tumor Type	Samples (T/N)	Platform	Reference
OralTongue	Oral tongue squamous cell carcinoma	31/26	HG-U95Av2	Estilo *et al.*,2009
Breast	Breast Cancer	42/43	HG-U133A	GSE15852
Lung	Lung adenocarcinoma	27/27	HG-U133A	Su *et al.*2007
ProstateSingh	Prostate cancer	52/50	HG-U95Av2	Singh *et al.*,2002
ProstateErnst	Prostate cancer	18/9	HG-U95A	Ernst *et al.*,2002
CML	Chronic myeloic leukemia	11/10 (CP)/(BC)	HG-U133A	Zheng *et al.*,2006
HeadAndNeck	Head and neck squamous cell carcinoma	22/22	HG-U95Av2	Kuriakose *et al.*,2004
Microdissected	Prostate cancer	7/8	HG-U133A	Dept. Gröne, dkfz (unpublished)

5.2 Mathematical models

This study focussed on analysing data and models of apoptotic pathway components. Two biologically relevant apoptosis models where considered. However, two toy models were analysed as well to compare the developed method for model analysis to alternative analysis methods.

5.2.1 Comprehensive apoptosis model used for both approaches

The study of Albeck et al. [3] aims at comprehensively modeling the whole complexity of interactions involved in apoptosis including the extrinsic pathway triggered by TRAIL ligand and the mitochondria dependent intrinsic pathway. Their extrinsic apoptosis reaction model (EARM v1.0) aims to be complete with respect to biochemical mechanisms and contains 58 equations of which 18 species have non-zero start-values. Fig. 5.1 depicts the molecule interactions of EARM v1.0. The authors also specified characteristic fea-

tures based on the PARP component to reflect effector caspase activity and relate those characteristics to apoptosis. The characteristic feature as indicator for apoptosis is the switching time as explained in section 4.1. The model of Albeck et al. [3] and its parameters specified in the original publication was used in this work for both approaches as presented in chapter 4. Detailed information about model components is given in the appendix.

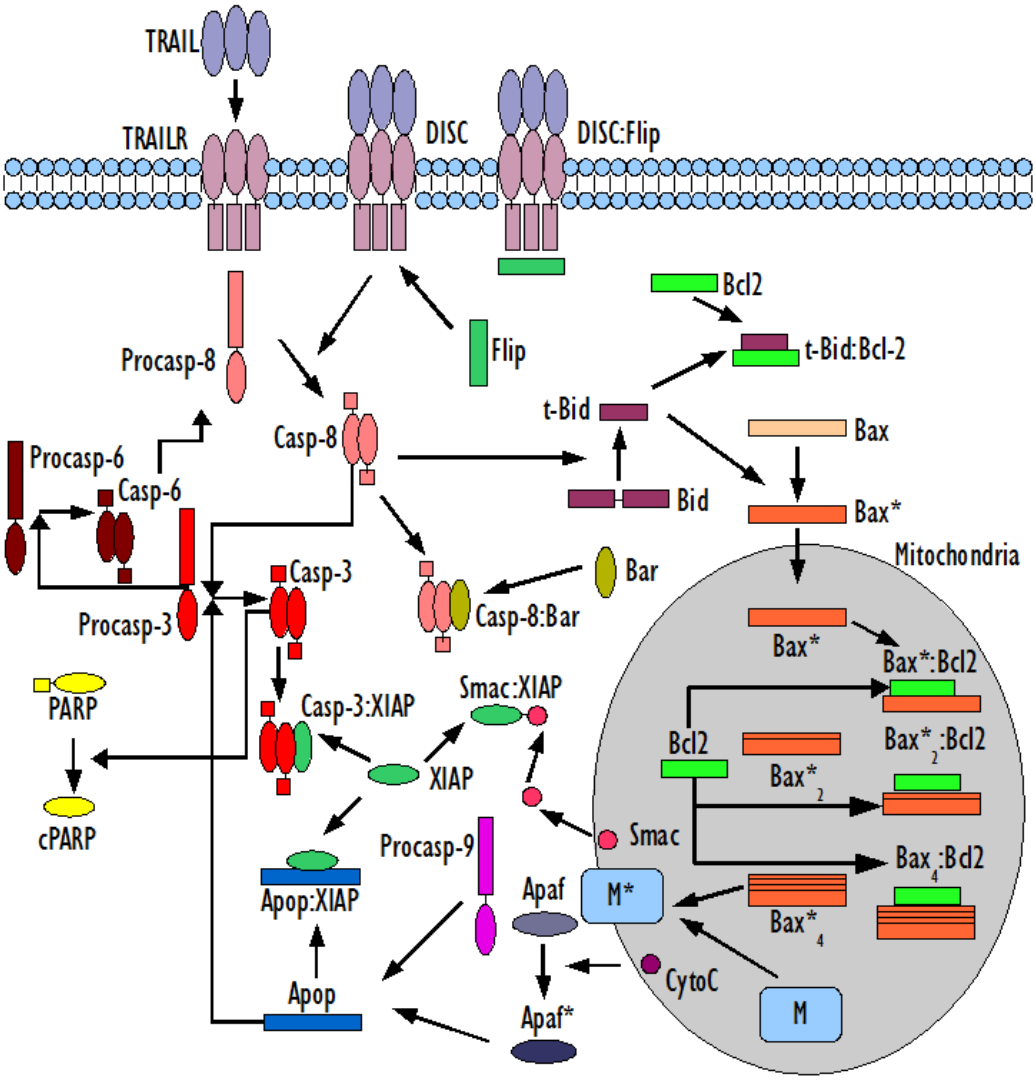

Figure 5.1: Molecule interactions of comprehensive apoptosis model (EARM v1.0)
The model, EARM v1.0 developed by Albeck et al. was used to study mechanisms during apoptosis

5.2.2 Models used for the evaluation of the second approach:

a) EGFR-internalization model

In order to test the developed method for model analysis explained in section 2.3, a model of EGF-receptor internalization developed by Schmidt-Glenewinkel et al. [85] was used, where an analytical solution was derived by the authors and therefore the performance of the developed model analysis method could be evaluated.

The EGFR-model contains a well known motive for competitive binding [16] which has been adapted and formulated for investigation of the sorting mechanism of the EGF-receptor into two different internalization pathways by Schmidt-Glenewinkel et al [85]. In this model, the ligand activated receptor can be internalized either via the clathrin dependent pathway (CDE) or the clathrin independent pathway (CIE).

The model consists of seven molecular species of which the ligand, the receptor and two adaptors abbreviated by CDE and CIE require an initial value greater than zero. Upon ligand activation, the EGF-receptor component either internalizes via the clathrin-dependent pathway (CDE-pathway) or the clathrin-independent pathway (CIE-pathway), resulting in the model components L:R:CIE and L:R:CDE as the most downstream components (Fig. 5.2)

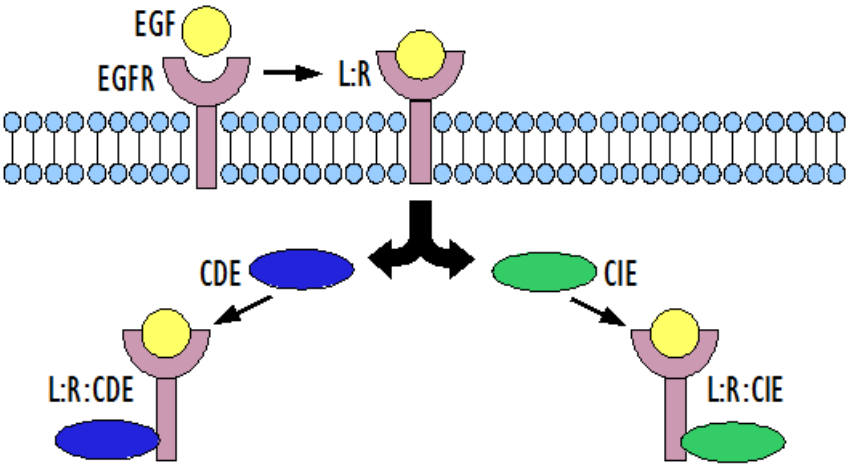

Figure 5.2: Schematic representation of EGF-receptor internalization pathways

b) Caspase-3 activation model

For comparison of a related multivariate method, called direct Lyapunov exponent (DLE) analysis and its results with the developed method applying decision trees for model analysis, a model for caspase activation was analyzed. This model consists of eight components and was investigated for conditions leading to apoptosis caused by high caspase-3 activation by Aldridge et al. [4].

The caspase model contains eight molecular species including pro-apoptotic procaspase-3, caspase-3, procaspase-8 and caspase-8 and pro-survival factor XIAP as well as three intermediate products (Fig. 5.3).

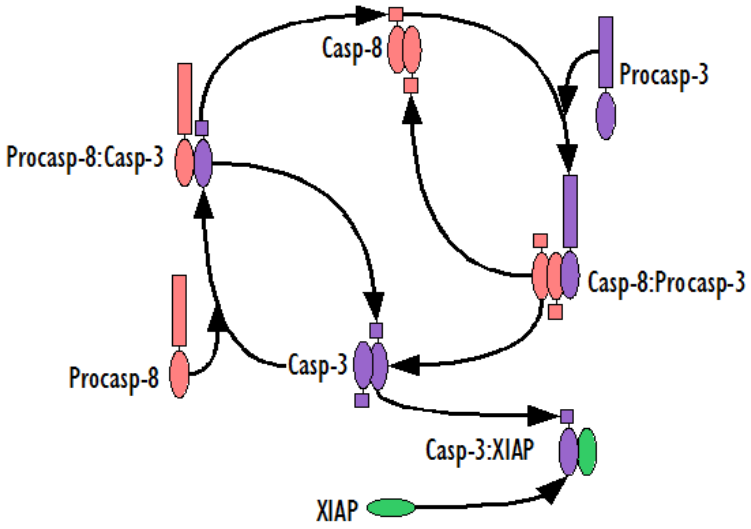

Figure 5.3: Network topology of caspase-3 activation model

c) CD95-mediated apoptosis model

CD95-mediated apoptosis has been studied experimentally and mathematically by simulation of several models [11] [73] [36]. The study of Neumann et al. focused on the crosstalk of the CD95-mediated extrinsic apoptosis with NFκB-pathway activation by upstream formation of p43-FLIP (Fig. 5.4). The ODE-based model was regressed to quantified western blot time series data and consists of 23 equations, eight of which represent molecular species with non-zero initial values including pro-apoptotic components represented by caspase-3 and -8 as well as two different species of cFLIP-isoforms previously associated with inhibition of apoptosis [60].

In order to predict biologically relevant relations of a well studied aspect of the apoptosis pathway, we applied our decision tree approach for model analysis to the model of Neumann et al. [73]. This model was chosen for analysis, since it has a reduced number of components focussing on the essential part of CD95-mediated apoptosis and NFκB activation, but still reflects experimental data well [73]. Due to the reduced number of components representing a experimentally and computationally well studied network, we expect our method for model analysis to predict biologically relevant relations of only few molecules deciding on cell fate.

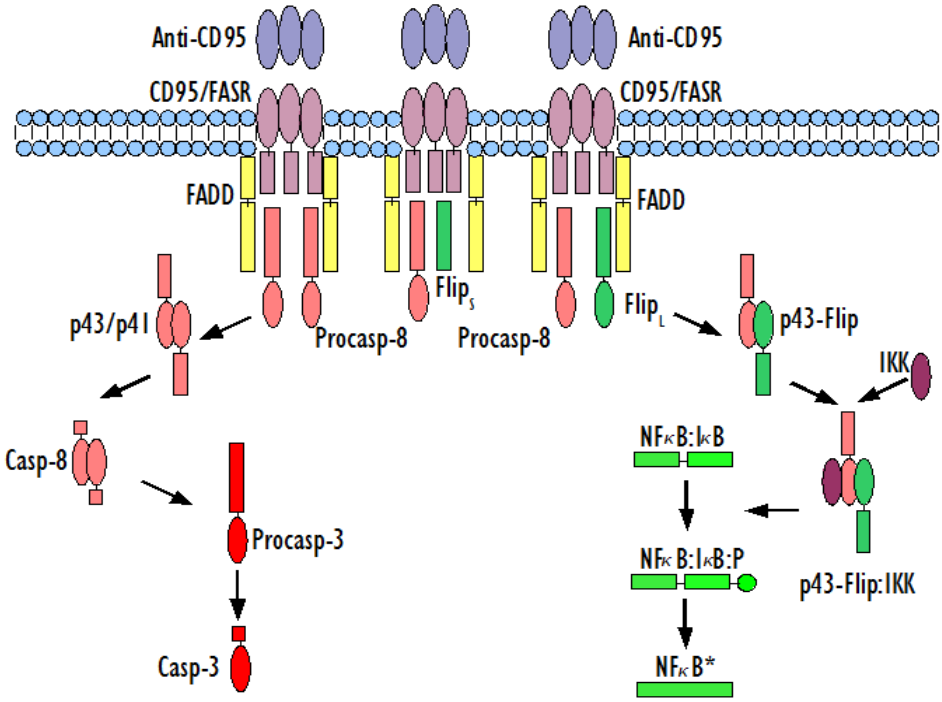

Figure 5.4: Network topology of CD95/Fas mediated apoptosis model

6 Results

Previous to data and model analysis, model components representing un-cleaved, un-complexed proteins which correspond to a certain gene were mapped to their corresponding probe identifiers, EntrezGene identifiers and symbols as listed in Fig. 6.1. Here, identifiers and symbols for caspase genes correspond to the protein procaspases.

Table 6.1: Model components and corresponding symbols, names and identifiers

Component	Gene symbol (Entrez Gene)	Official Name (Entrez Gene)	Gene ID
TRAIL-R1/DR4	TNFRSF10A	*tumor necrosis factor receptor superfamily, member 10a*	8797
Flip	CFLAR	*CASP8 and FADD-like apoptosis regulator*	8837
Procaspase 8	CASP8	*caspase 8, apoptosis-related cysteine peptidase*	841
Bar	BFAR	*bifunctional apoptosis regulator*	51283
Procaspase 3	CASP3	*caspase 3, apoptosis-related cysteine peptidase*	836
Procaspase 6	CASP6	*caspase 6, apoptosis-related cysteine peptidase*	839
XIAP	XIAP	*X-linked inhibitor of apoptosis*	331
PARP	PARP1	*poly (ADP-ribose) polymerase 1*	142
Bid	BID	*BH3 interacting domain death agonist*	637
Bcl-2	BCL-2	*B-cell CLL/Lymphoma*	596
Bax	BAX	*BCL2-associated X protein*	581
Cytochrom C	CYCS	*cytochrome c*	54205
Smac	DIABLO	*diablo homolog (Drosophila)*	56616
Procaspase 9	CASP9	*caspase 9, apoptosis-related cysteine peptidase*	842
Apaf-1	APAF	*apoptotic peptidase activating factor 1*	317
TRAIL	TNFSF10	*tumor necrosis factor (ligand) superfamily, member 10*	8743

6.1 Perturbation of individual genes

We first investigated which single genes are over- or under- expressed in tumor samples compared to normal samples. To this aim we analysed each of the eight gene expression data sets for the fold change of every probe corresponding to a model component. Some components are represented by several probes, hence Fig. 6.1 contains repeated labels. In the OralTongue data set (Fig. 6.1 A), the fold change value of TRAIL ligand (TNFSF1) is four fold over-expressed. Also pro-apoptotic BAX is double as strongly expressed in tumors compared to normal samples. In contrast, the fold change (*FC)* of all anti-apoptotic Bcl-2 probe sets are slightly lower than 1. The remaining *FC*-values are in the range of 0.9 to 1.4.

All fold changes of the BreastCancer data set are close to 1 and therefore regarded as not differentially expressed. *FC*-values of procaspase-3, cytochrome C and some probe sets corresponding to the ligand are exceeding 1.2 (Fig 6.1 B).

In case of the Lung data set (Fig. 6.1 C), genes for TRAIL ligand (TNFSF10), Flip (CFLAR) and one of the Bcl-2 probe sets is lower expressed in tumors than in normal samples, while PARP and cytochrome C are over-expressed (*FC* = 1.26 and *FC* = 1.5 respectively).

The probe sets of the ProstateSingh data set (Fig. 6.1 D) are mostly balanced between tumors and normal samples. Ligand TNFSF10 and cytochrome C however, are over-expressed in tumors (*FC* = 1.5 and *FC* = 1.25, respectively).

Prostate cancer patients in the ProstateErnst data set (Fig. 6.2 E) show a Bax up-regulation (*FC* = 1.4) for the same probe set as identified in the OralTongue data set. Most other probe sets are within a range of 0.8 to 1.2. Furthermore, a slight reduction of one of the BCL-2 probe sets can be observed in tumor.

CML patients in blast crisis (Fig 6.2 F) have higher gene expression values for TRAIL (TNFSF10), procaspase-8, PARP, Bax and BCl-2 (probe set ID: 203685_at) with FC-values of about 1.5. Another characteristic of this data set is a high variability (*FC* of 0.6-1.15) of Flip fold changes (CFLAR).

In the case of the Head And Neck data set, procaspases are slightly over-expressed (*FC* = 1.2), and PARP is up-regulated (*FC* = 1.4) in tumors. Higher fold change values of 1.4 and 1.5 were are observed for Bax probes (Fig. 6.2 G).

The Microdissected data set (Fig. 6.2 H) shows *FC*-values close to 1 for all probes, except for Bcl-2 (probe set ID: 203685_at), which is reduced by almost 50% reduction in tumor.

In summary, the fold changes of all data sets are not considered as highly up- or down-regulated, since most *FC*-values are in a range between 0.8 and 1.6. Only few exceptions are exceeding this range: down regulated probes below a fold change of 0.8 are found in the Lung data set for TRAIL-L (*FC* = 0.48) and CFLAR (*FC* = 0.65), the latter corresponding to the homologues of the FLIP-family. Within the CML data set, CFALR (*FC* values of 0.6 and 0.75) and Bcl-2 in Prostate Microdissected (*FC*= 0.55) are found to be below 0.8. The only component considered as strongly over-expressed in tumor samples is ligand TRAIL-l (TNFSF10) with a *FC*-value of 4.5 in the OralTongue data set.

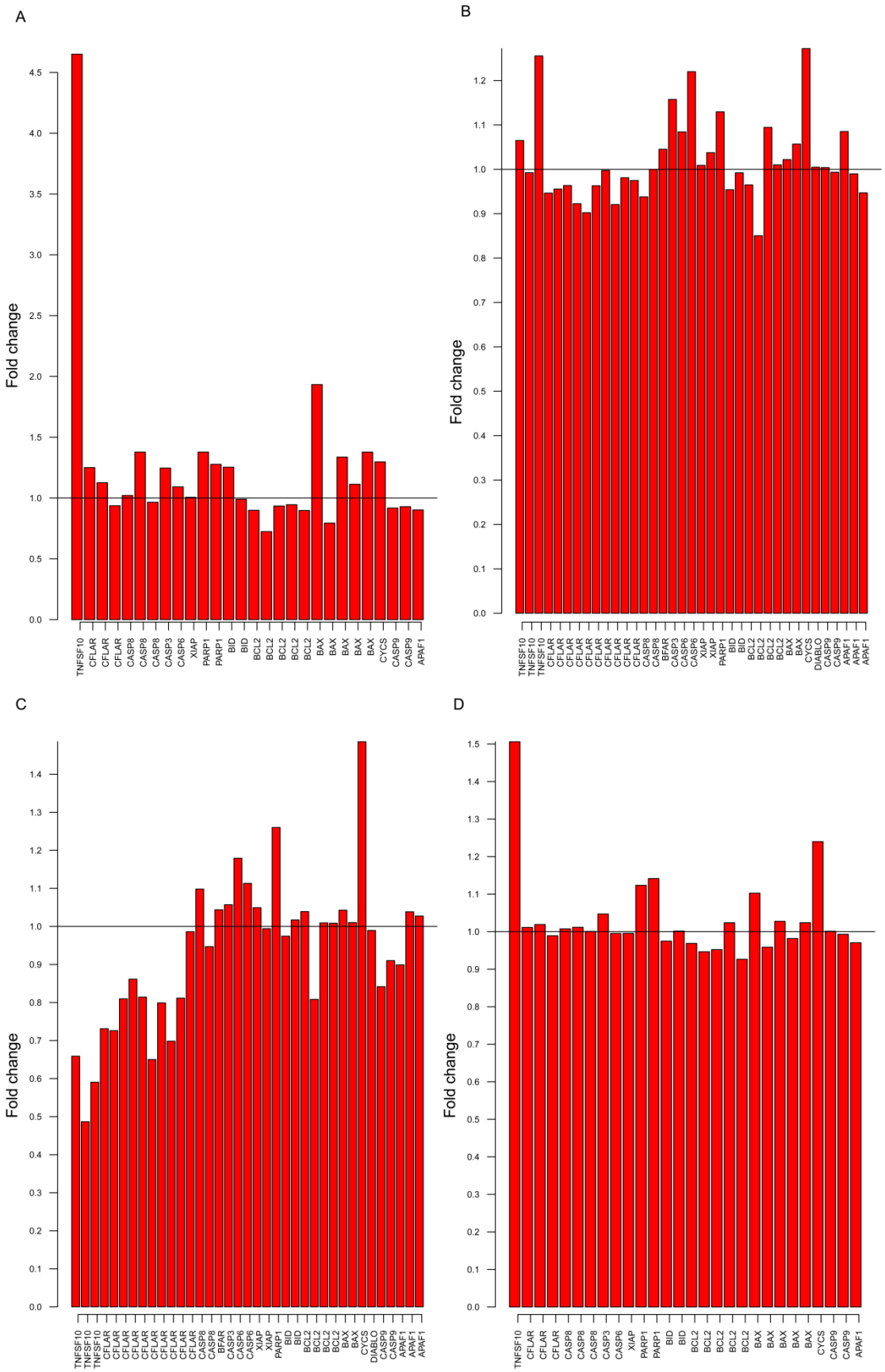

Figure 6.1: Fold changes of gene expression data A-D
A: OralTongue, B: Breast, C: Lung, D: ProstateSingh

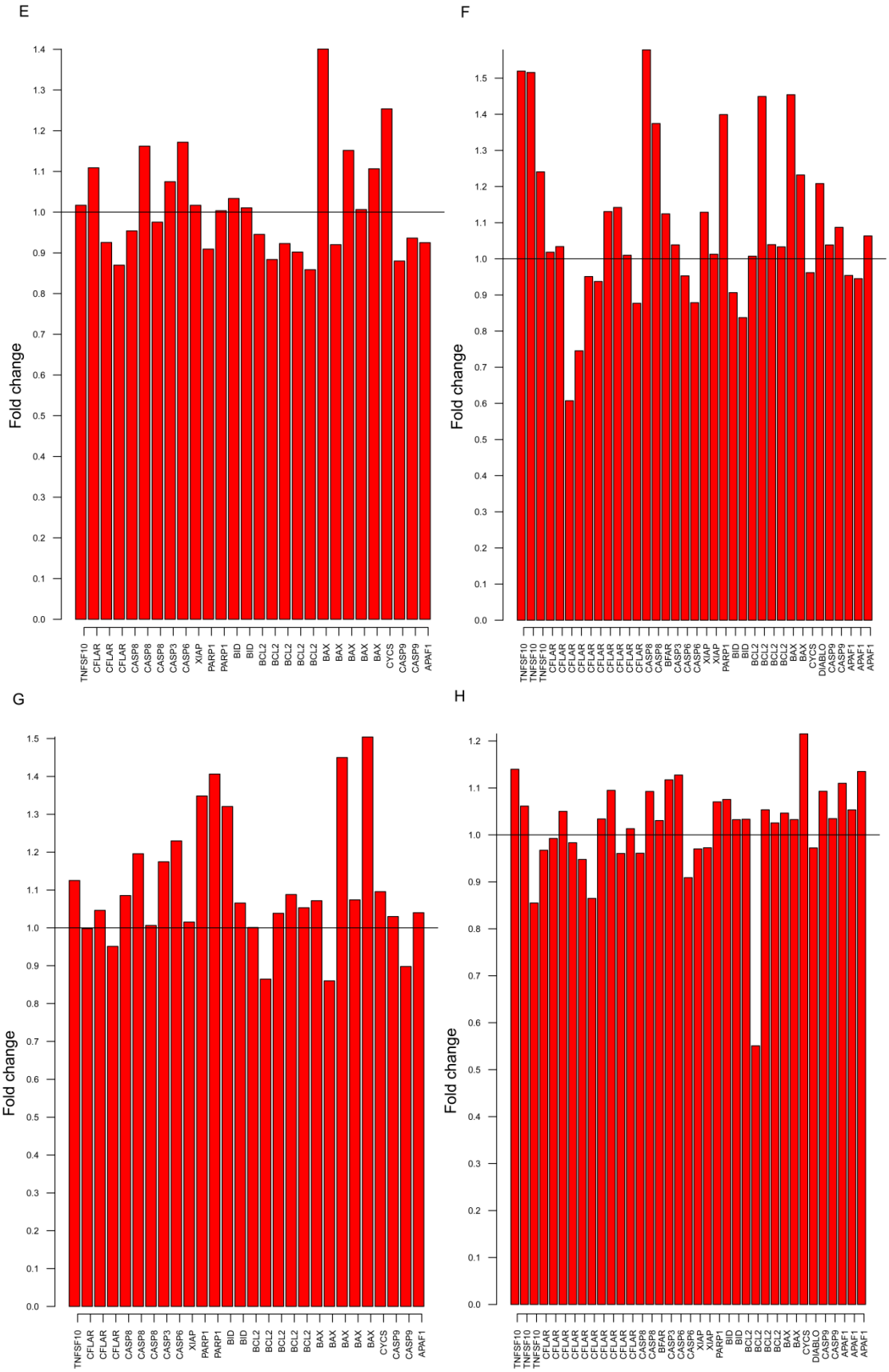

Figure 6.2: Fold changes of gene expression data E-H

E: ProstateErnst; F: CML ; G: HeadAndNeck ; H: Microdissected

6.2 Apoptosis pathway association with disease state

By means of the *globaltest*, the ability to classify phenotypes based on model components as a set of genes is tested. *P*-values below 0.05 are considered as significant. Hence, an association with the group of genes listed in Table 6.1 to the phenotype can be assumed for the gene expression data sets: OralTongue, BreastCancer, Lung and ProstateSingh (Table 6.2).

Table 6.2 Data sets and respective p-values resulting from a *globaltest*
The *globaltest* analyzes sets of genes towards their ability of distinguishing phenotypes and yields a p-value as significance measure for the association of the tested group of genes with a phenotype or class. The selected apoptosis components tested for data sets OralTongue, BreastCancer, Lung and ProstateSingh result in a p-value of below 0.05. Hence for these data sets, the apoptosis components are regarded as significantly associated with the clinical phenotype.

Data Set	*p*-value
OralTongue	$1.1942 \cdot 10^{-7}$
Breast	$1.336 \cdot 10^{-5}$
Lung	$0.78422 \cdot 10^{-3}$
ProstateSingh	$0.94492 \cdot 10^{-3}$
ProstateErnst	0.05303
CML	0.061471
HeadAndNeck	0.17882
Microdissected	0.48811

6.3 Approach 1: From data to model

The assumption for this approach is that tumor cells are more resistant to apoptosis and have developed strategies to survive, in contrast to normal cells, which are supposed to be more sensitive to death stimuli. The PARP trajectory as output signal of the model of Albeck et al. [3] is therefore assumed to have an earlier or steeper decline for scaling start values according to normal samples.

In the following, only figures according to the OralTongue data set are shown. Results for all other data sets were similar. Moreover, the *globaltest* applied to the OralTongue data set resulted in the lowest p-value of all given gene expression data sets. However, differing results and peculiarities of certain data sets are noted in this chapter and figures for all data set can be found in the Appendix A.

6.3.1 Results for start values scaled by fold change

Each of the eight data sets listed in Tab. 5.1 was used separately for scaling original start values according to tumor and normal data. Time courses of the PARP component resulting from original start values representing the class “tumor” and scaled start values representing the class “normal” as illustrated in Fig 6.3 were visually inspected.

PARP start values after scaling were for most data sets lower than original values and resulted in a delayed response while steepness of the time courses and switching time T_s were similar for all data sets. Scaling according to the data set ProstateErnst yielded a slightly higher initial value for PARP but still resulted in a delayed response. The Lung data set resulted in a lower concentration for PARP and in a slightly earlier response (Appendix A).

However, the scaling of original start values by the fold change did not yield qualitatively different trajectories for the PARP component with regard to T_s as indicator for apoptosis, even when start values were much higher or lower after scaling.

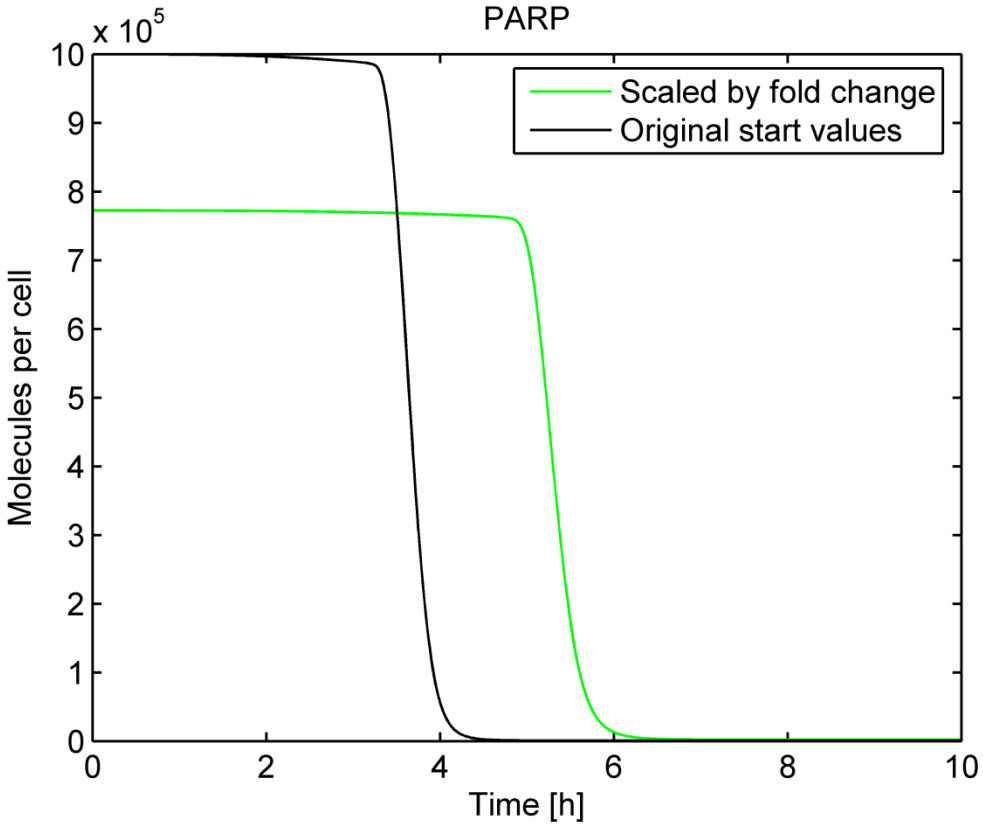

Figure 6.3: PARP trajectories resulting from fold change scaling
PARP trajectories shown for original start values in black and for fold change scaled values representing the class “normal” in green as a result of the OralTongue data set.

6.3.2 Results for start values scaled per sample

Scaling per sample was introduced to avoid averaging of samples of one class and get a more differentiated system response according to individual samples. However, the obtained system response does not show a clear separation of the PARP time courses resulting from scaling per sample with regard to the classes tumor and normal as illustrated in Fig. 6.4 for OralTongue data.

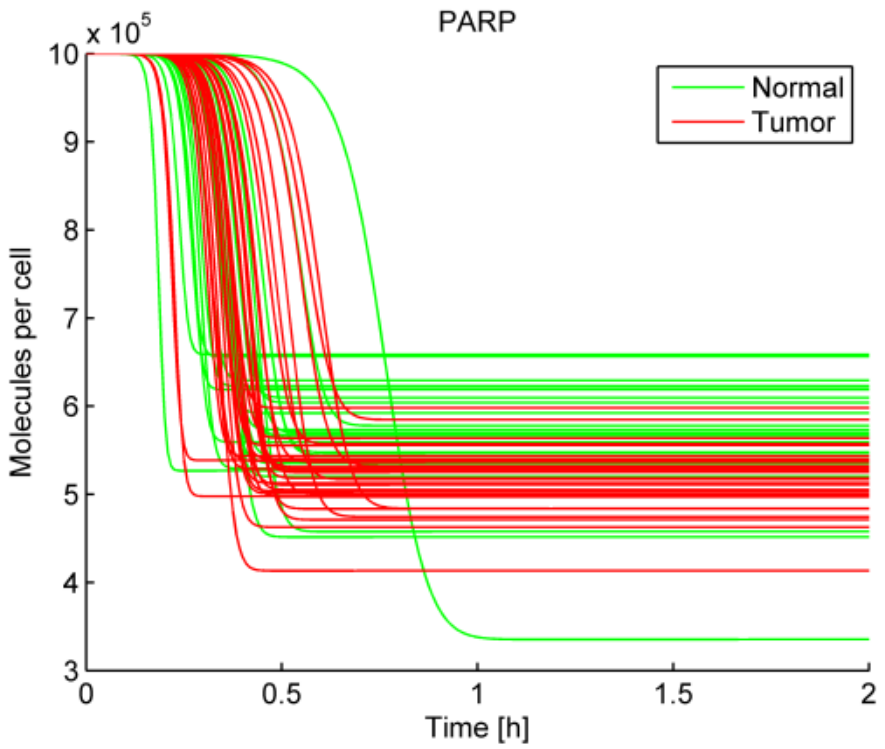

Figure 6.4: PARP trajectories resulting from scaling per sample
Start values were scaled according to microarray data of normal samples (green) and tumor samples (red).

Table 6.3: T_d and T_s of PARP time courses for gene expression data sets
Median values for T_d (delay time) and T_s (switching time) calculated as characteristics of PARP trajectories resulting from scaling according to normal samples (N) and tumor samples (T).

Data set	T_d median N (minutes)	T_d median T (minutes)	T_s median N (minutes)	T_s median T (minutes)
OralTongue	22	23	8	8
Breast	342	339	122	159
Lung	491	459	552	450
Prostate Singh	13	14	4	4
Prostate Ernst	41	31	17	12
CML	216	200	77	50
HeadNeck	40	43	17	20
Prostate Microdiss	360	300	602	285

Calculation of median values of delay and switching times with regard to the system response of each class were performed and showed only minor differences (see Table 6.3, compare column 2 to column 3 and column 4 to column 5).

6.3.3 Relations reflected in gene expressions

The encoded relations of model components were ranked according to their relative frequencies observed for the phenotype classes (tumor/normal) as explained in section 4.1.3. The top 20 component relations of data sets yielding *p*-values below 0.05 for the *globaltest* are listed in Tab. 6.4. Ranking of relations for the remaining data sets are presented in the Appendix A. Generally, for all analysed data sets, in case a pro-apoptotic and an anti-apoptotic component form a relation, the pro-apoptotic gene is found to be higher expressed than the anti-apoptotic components in tumor samples (Tab. 6.4). This appears to be counter-intuitive, since it suggests tumor cells to be more sensitive to apoptosis. Only gene expression data from breast cancer show one anti-apoptotic component (BFAR) in tumor compared to normal to exceed the expression level of a pro-apoptotic component (caspase-8).

These results agrees with the finding of Stegh et al. [88] who showed that BFAR inhibits caspase-8 in the breast cancer cell line MCF-7 and thus prevents the cleavage of Bid and caspase-3. Hence caspase-8 exceeding BFAR in the breast cancer data set suggests that the identified molecule relation contributes to apoptosis resistance.

Furthermore, PARP levels are found to exceed pro-apoptotic components in tumor, but are in this work not considered to contribute to a death or survival decision, since it is the most downstream protein in the model used.

Table 6.4: Pair-wise relations of apoptosis components in gene expression data
Orange background: Relation more often found in tumor, light green background: Relation more often found in normal, Red font: A pro-apoptotic component, Dark green font: A pro-survival/anti-apoptotic component, Blue circle: Only relation were an anti-apoptotic component is higher in tumor samples than a pro-apoptotic component on gene expression level. Same gene symbols within one relation refer to different probe set IDs of the microarray.

Pos.	OralTongue	Breast	Lung	Singh2002
1	TNFSF10 < BAX	CASP8 < PARP1	TNFSF10 < CYCS	TNFSF10 < BAX
2	TNFSF10 < BCL2	CYCS < APAF1	CFLAR < PARP1	BCL2 < BAX
3	TNFSF10 < BCL2	TNFSF10 < CFLAR	CASP3 < BCL2	TNFSF10 < BCL2
4	TNFSF10 < CASP8	TNFSF10 < XIAP	TNFSF10 < CASP8	TNFSF10 < BAX
5	TNFSF10 < PARP1	CFLAR < CYCS	CFLAR < PARP1	TNFSF10 < PARP1
6	TNFSF10 < PARP1	CYCS < APAF1	CASP6 < BCL2	TNFSF10 < BCL2
7	BAX < BAX	PARP1 < CASP9	BCL2 < BCL2	TNFSF10 < CASP9
8	TNFSF10 < CFLAR	CFLAR < BCL2	TNFSF10 < PARP1	TNFSF10 < BID
9	BCL2 < BAX	CFLAR < CFLAR	TNFSF10 < BID	CFLAR < BCL2
10	TNFSF10 < BAX	CFLAR < CYCS	TNFSF10 < PARP1	TNFSF10 < BAX
11	BID < BCL2	CFLAR < BCL2	CFLAR < PARP1	CFLAR < CYCS
12	PARP1 < BAX	TNFSF10 < CFLAR	CFLAR < BFAR	TNFSF10 < CASP8
13	CASP3 < BCL2	TNFSF10 < CFLAR	CFLAR < CASP6	TNFSF10 < BCL2
14	TNFSF10 < BAX	CFLAR < CFLAR	CFLAR < CASP9	CFLAR < BCL2
15	CASP8 < BAX	CASP8 < BFAR	CFLAR < BID	TNFSF10 < PARP1
16	BCL2 < BAX	CASP8 < CYCS	CFLAR < CASP9	TNFSF10 < BCL2
17	CASP3 < APAF1	BCL2 < CYCS	CFLAR < CASP9	TNFSF10 < CASP9
18	CASP8 < BCL2	CFLAR < APAF1	CFLAR < BID	CYCS < CASP9
19	BCL2 < BAX	TNFSF10 < CFLAR	CFLAR < PARP1	PARP1 < BAX
20	CFLAR < BAX	PARP1 < CASP9	BID < BCL2	BID < BCL2

6.3.4 Complementary analysis

The simulation results did not show remarkably different system responses. In contrast *p*-values lower than 0.05 of the data sets OralTongue, Breast, Lung, ProstateSingh (Tab.6.2 and Tab 6.3) resulting from a *globaltest* indicated significance although most fold change values were close to 1, within a range of 0.8 to 1.6.

By visual inspection, fold change plots (see Figures 6.1 and 6.2) were analyzed and components were identified that are up- or down-regulated the most. Only few exceptions are exceeding the fold change range between of 0.8 and 1.6:

The ligand TRAIL-L with a fold change of about 4.5 is exceptionally over-expressed in tumor samples of the OralTongue data set. Down-regulated probes below a fold change of 0.8 are found in the Lung data set for TRAIL-L (fold change of 0.48) and CFLAR (fold change of 0.65), the latter corresponding to homologues of the Flip-family. Within the

CML data set, CFLAR (fold changes of 0.6 and 0.75), and Bcl-2 in the Microdissected data set (fold change of 0.55) are found to be below 0.8.

Note that fold change values for all data sets identified that most pro-apoptotic components are higher expressed in cancer higher than in normal samples, and most anti-apoptotic pathway components to be lower expressed in normal samples. Exceptions here can be observed in the data sets Lung and ProstateErnst where pro-apoptotic procaspase-9 is down-regulated in cancer. Another exception is up-regulation of anti-apoptotic Bcl-2 in CML.

For some data sets, either the intrinsic or the extrinsic pathway was more likely to be affected, as listed in Table 6.5. Based on expression levels, in the case of prostate cancer and head and neck cancer the intrinsic pathway is turned on, whereas in CML rather the extrinsic pathway is turned on. Note that here not solely up-regulation of pro-apoptotic components, but also down-regulation of anti-apoptotic components is considered.

Table 6.5 Inspection of fold change values with regard to apoptotic subpathways

Visual inspection of the PARP trajectories showed either degradation in one step or stepwise in two steps (column 1). Plots of fold changes (FC) showed low differences (0.8-1.5), although yielding a low p-value for the *globaltest* (GT) (column one). Model components which yield higher or lower FC in tumor (T) are listed in columns two and three. Also, fold changes for pathway components were found to either affect rather the extrinsic (extr.) or the intrinsic (intr.). The arrows (↑ / ↓) indicate subpathway activation of apoptosis in tumor, which not only refers to up-regulation of pro-apoptotic, but also considers down-regulation of anti-apoptotic components. Some data sets showed a heterogeneous up-/down-regulation of components for subpathways (H).

Data set	GT p-value	FC higher in T	FC lower in T	Patway affected
OralTongue	1.19e-07	L, Bax	Bcl-2	Extr. ↑, intr. ↑
Breast	1.34e-05	L, C6, C3 Cyto C	Bcl-2	Extr. ↑, intr. ↑
Lung	0.78e-03	Cyto C, PARP	L, Flip, C9	Extr. ↓, intr. ↑
Prostate Singh	0.94e-03	L, Cyto C	-	Extr. ↑, intr. ↑
Prostate Ernst	0.53e-01	Bax, Cyto C	C9, Bcl-2, Apaf	Intr. ↑
CML	0.61e-01	L, C8, PARP, Bax, Bcl-2	Flip	Extr. ↑, intr. (H)
HeadNeck	0.18e+00	Bax, PARP, Bid, C3/6/8	Bcl-2, Bax	Extr. (H), intr. ↑
Prostate Microdiss	0.48e+00	Cyto C	Bcl-2	Intr. ↑

6.4 Generating hypotheses from model analysis: Approach 2

The developed method for model analysis presented in section 4.2.1 was applied to ODE-models as explained in section 5.2. In the following, results from using the method to predict component relations that trigger pathway activation will be presented.

6.4.1 Decision tree approach can classify parameter sets

For the small model of EGF-receptor internalization illustrated in Figure 5.2, the pathway of interest is the CIE-pathway. In order to find conditions leading to CIE-pathway activations, the class definition was based on the trajectories of the L:R:CIE- (the ligand activated receptor internalized via CIE) and the L:R:CDE-component (the ligand activated receptor internalized via CDE). The system response was classified as class 1 (pathway "on"), in case the L:R:CIE level reaches 80% of the L:R:CDE level at steady state, else the CIE-pathway was defined to be not sufficiently activated and classified as class 0 (pathway "off").

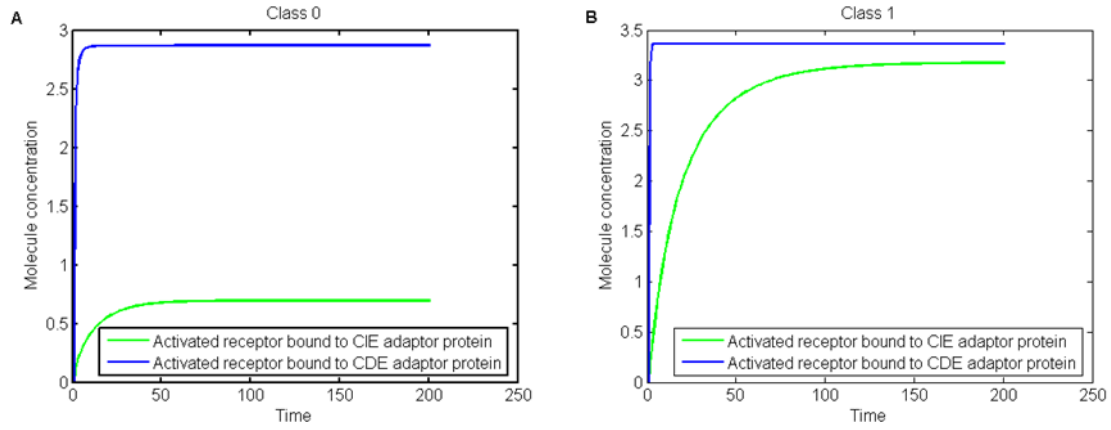

Figure 6.5: Pathway activation classes of clathrin-independent EGFR-internalization.
A: L:R:CIE-trajectories (green) and L:R:CDE-trajectories (blue) representing CIE-pathway "off".
B: L:R:CIE-trajectories (green) and L:R:CDE-trajectories (blue) representing CIE-pathway "on"

Random initial values were generated for the ligand, the receptor and the two adaptors CIE and CDE as specified in Appendix B (Table B1). Out of 10000 generated parameter sets, 1808 in total (904 per class) were chosen as input for the decision tree algorithm resulting in a misclassification trajectory illustrated in Figure 6.7.

The misclassification error obtained from the algorithm decreases with regard to the number of leave nodes for both training and test by cross validation and suggests a decision tree containing four leave nodes (Fig. 6.6) as the tree on the best level. This tree

yields a misclassification error of 0.06 (Fig. 6.7) which is equivalent to an accuracy of 94%. For comparison, the decision tree applied to absolute values instead of parameter relations is illustrated in Fig. 6.8. Here, a pruning level was chosen with regard to a similar misclassification error as for the tree in Fig. 6.6.

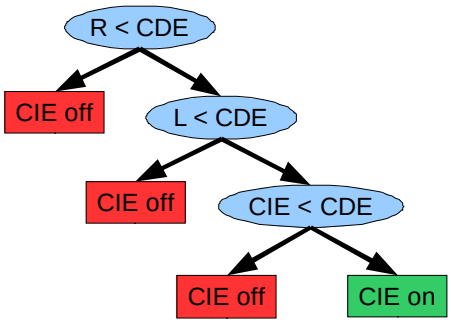

Figure 6.6: Decision tree resulting from model analysis of EGFR-model

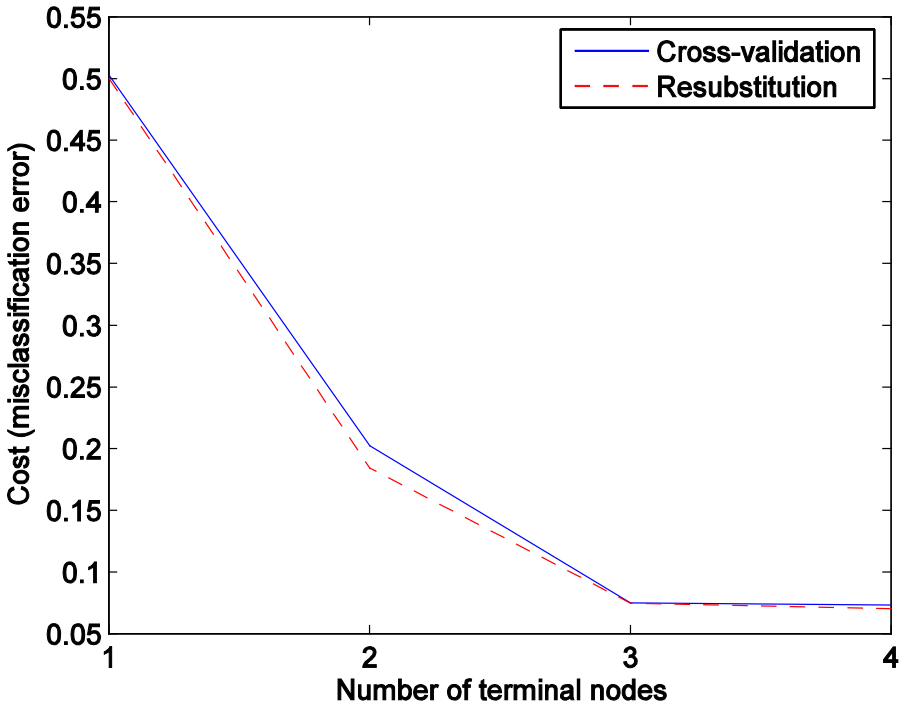

Figure 6.7: Misclassification error for decision tree method applied EGFR-model

Although a decision tree applied to absolute parameter values instead of relations can classify randomly sampled parameter sets equally well, as the respective error trajectory suggests (Fig. 6.9), the corresponding tree (Fig. 6.8) is not easy to interpret.

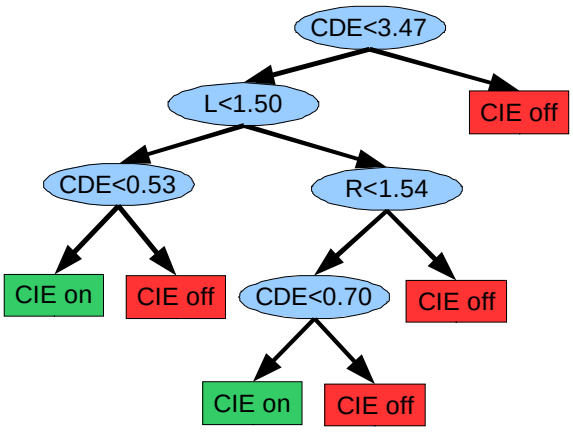

Figure 6.8: Decision tree applied to absolute parameter values of EGFR-model

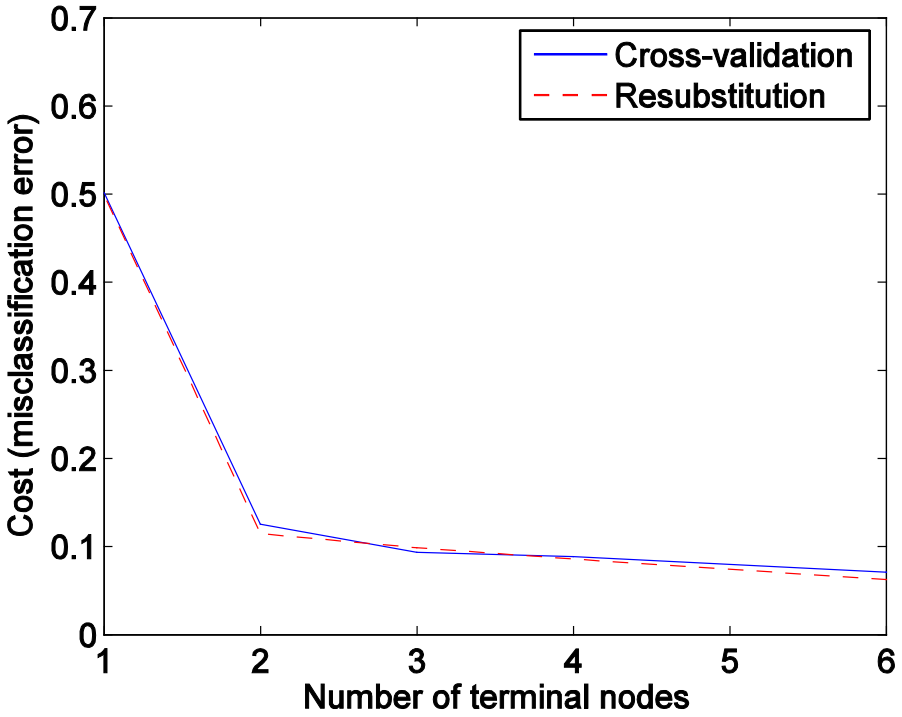

Figure 6.9: Misclassification error for decision tree of Fig. 6.8

The decision tree using absolute values as features (Fig. 6.8) chooses the components and threshold values as nodes which reflect the parameter relations of the tree using component relations (Fig. 6.6). However, the decision tree using relations as features structures the tree according to the assumption that thresholds of molecule abundances are crucial for signaling. This results in a more condensed representation which is easy to interpret.

6.4.2 Analytically determined solution is reflected in decision tree

The focus of interest of Schmidt-Glenewinkel et al. [85] was elucidating which conditions lead to a switching behaviour of the CIE-pathway. This was done by steady-state analysis and parameter scans. Two different steady-state classes were identified which are reached depending on the initial values of molecule species leading to either low, medium or high number of ligand-activated receptors internalized via CIE-adaptors (Tab. 6.6).

The system reaches steady-state class I if all free receptors or all ligand molecules get depleted and all activated receptors (model component EGF:R) will be internalized [85]. This steady state has been associated with no or basal receptor internalization via CIE-adaptors and is reached in case the initial values of either ligand or receptor is lower than the number of CDE-adaptors (condition (S1), Tab. 6.6).

In order to reach steady state class II, neither the ligand nor the receptor is limiting with respect to pathway activation, but both adaptor types are consumed completely.

This steady state is reached, if ligand and receptor concentrations exceed initial values of CDE adaptors leading to a moderate CIE-pathway activation (condition (S2), Tab.6.6). A high level of CIE-receptor internalization is reached if ligand and receptor concentrations exceed the sum of CIE_0 and CDE_0 (condition (S3), Tab.1). X_0 denotes here the initial concentration of component X.

In order to demonstrate how the conditions found by our method and stated in the decision nodes reflect the requirements of Schmidt-Glenewinkel et al. [85], we transformed the decision tree (Fig. 6.6) into a set of rules where each path from root to a leaf node is represented as a rule containing the expressions of each decision node on that path (Tab. 6.7).

The top two decision nodes "$R_0 < CDE_0$" and "$EGF_0 < CDE_0$" reflected in rule (D1) and in the second expression of rule (D2) of Tab. 6.7 indicate that either the ligand or the receptor level is required to be lower than CDE start concentration to result in an inactive CIE-pathway. This directly corresponds to condition (S1) (Tab. 6.6).

Table 6.6: Conditions for pathway activation of EGFR internalization via CIE

Initial values	Steady-state	CIE-internalization
(S1) $\min\{EGF_0,R_0\} < CDE_0$	I	Low
(S2) $\min\{EGF_0,R_0\} > CDE_0$	II	Medium
(S3) $\min\{EGF_0,R_0\} > CDE_0 + CIE_0$	II	High

Table 6.7: Decision rules for CIE-pathway activation
The decision tree obtained from analysis of the EGFR-model was translated into corresponding set of decision rues.

Path	Rule
(D1)	If $R_0 < CDE_0$ then class "CIE off"
(D2)	If $R_0 >= CDE_0$ and $L_0 < CIE_0$ then class "CIE off"
(D3)	If $R_0 >= CDE_0$ and $L_0 >= CDE_0$ and $CIE_0 < CDE_0$ then class "CIE off"
(D4)	If $R_0 >= CDE_0$ and $L_0 >= CDE_0$ and $CIE_0 > CDE_0$ then class "CIE on"

The developed method thus identifies either the ligand or the receptor as the limiting factor for pathway activation, which exactly reproduces the requirements determined by Schmidt-Glenewinkel et al. for steady state class I associated with no or low CIE-internalization [85] (Tab. 6.6).

For a moderate activation of the CIE-pathway, both the ligand and receptor initial values have to exceed the number of CDE-adaptors (condition (S2), Table 6.6), which is reflected in the top two decision nodes combined with each other and more explicitly stated by the first two terms of rule (D4) for pathway activation ("$R_0 >= CDE_0$" and "$L_0 >= CDE_0$", Tab.6.7).

However, for a full activation (condition (S3), Table 6.6) leading the system towards steady state class II, ligand and receptor have to exceed the sum of CDE- and CIE-adaptors according to Schmidt-Glenewinkel et al [85]. This condition is not obvious from the decision tree and the corresponding rule set.

Therefore the influence of component start values on pathway activation was explored by simulation and visual inspection of component levels of receptor internalized via CIE and CDE at steady state. Simulations were performed for initial values in the order of magni-

tude of original values [85] with fixed start values for ligand and receptor ($L_0 = R_0 = 1.5$) and varying start values of adaptor components (CIE_0 and $CDE_0 \in [0, 2.5]$).

Simulation results for L:R:CIE- and L:R:CDE- levels at steady state are illustrated according to the conditions previously identified (Tab.6.6, [85]) in Fig. 6.10 and according to the decision tree results in Fig. 6.11.

The steady state levels of receptors internalized via CIE and classified as 'pathway on' (moderate or full) depicted in Fig.6.10 in yellow and green show a large overlap with the decision tree result of Fig. 6.11 ('pathway on' coloured in green). However, the decision tree performs a binary classification ('pathway on' or 'pathway off' depicted in green and red respectively (Fig. 6.11) instead of a three-fold categorization (full, moderate and off) and is more restrictive for CIE_0 and $CDE_0 \in [0, 1.5]$ i.e. for initial adaptor values lower than ligand and receptor concentration.

Hence the decision tree categorizes the subspace for $CIE_0+CDE_0 < L_0 (= R_0 = 1.5)$ as not activated (Fig. 6.11) other than the analytical solution (Tab.6.6), which identifies this subspace as 'pathway activated' although at fairly low absolute values of L:R:CIE as well as low relative levels compared to L:R:CDE are observed (Fig. 6.10).

Further, pathway activation according to the decision tree was compared to pathway activation according to the predefined criterion for classification of parameter sets (step 1 section 4.2.1). Here, the decision tree is less restrictive than the predefined criterion due to simple rules of our approach used as features. Therefore a larger area is assigned to 'pathway on' as result of the decision tree (Fig. 6.11) compared to the result of the predefined criterion (Fig. 6.12, green area).

The developed method is therefore able to identify the previously published conditions leading to an inactive CIE-pathway as well as conditions resulting in a large overlap for an active CIE-pathway. Moreover, simulation results show that it is crucial which characteristics are regarded as representing pathway activation.

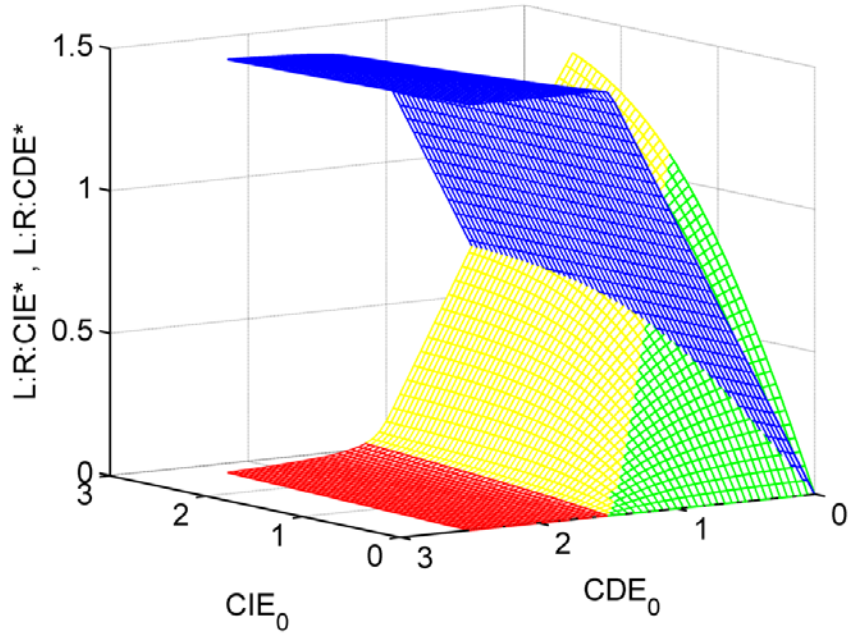

Figure 6.10: Conditions for pathway activation according to steady state analysis
Amount of receptors internalized via CIE and parameters fulfilling conditions identified by steady state analysis of Schmidt-Glenewinkel et al. [85] for full pathway activation in green, for moderate activation in yellow and no activation in red. For comparison, the amount of receptors internalized via CDE is plotted in blue. Results were observed for t = 200 at steady state. Ligand and receptor initial values set to $L_0 = R_0 = 1.5$.

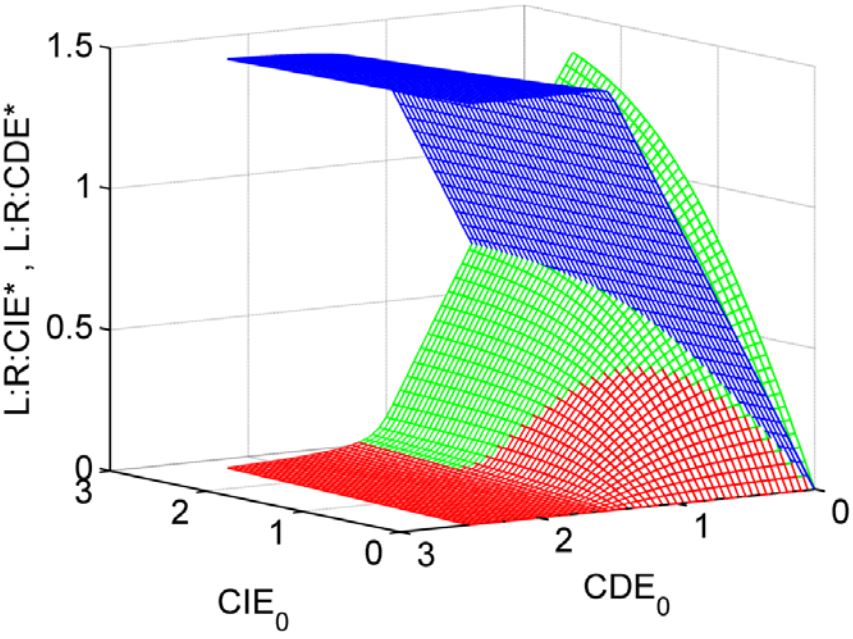

Figure 6.11: Conditions for pathway activation according to the decision tree
Amount of receptors internalized via CIE and parameters fulfilling conditions identified by decision tree analysis for pathway activation colored in green and, in case, conditions are not fulfilled in red.
For comparison, the amount of receptors internalized via CDE is plotted in blue. Results were observed for t = 200 at steady state. Ligand and receptor initial values set to $L_0 = R_0 = 1.5$.

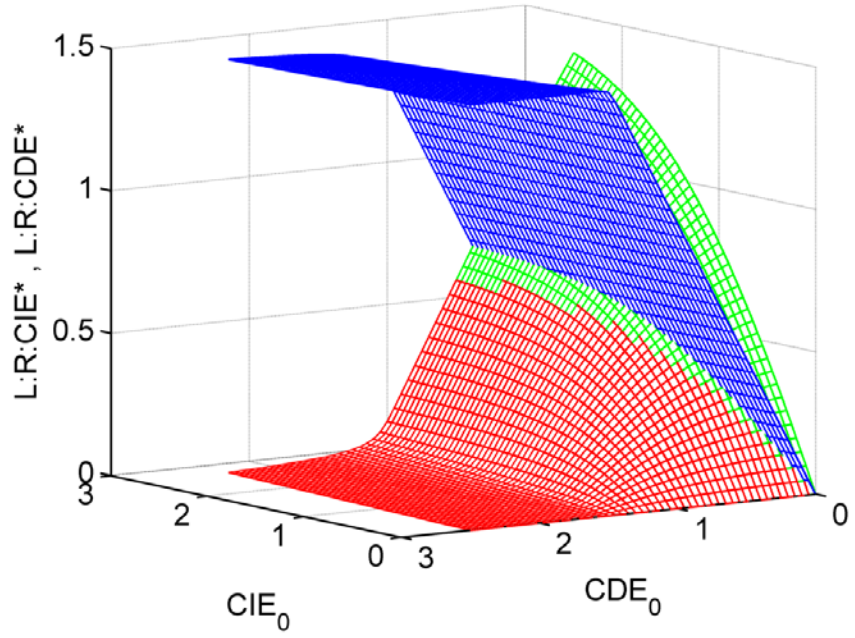

Figure 6.12: Conditions for pathway activation according to predefined criterion
Amount of receptors internalized via CIE according to predefined criterion for classification of parameter sets into class 1 ('pathway on') in green and class 0 ('pathway off') in red. For comparison, the amount of receptors internalized via CDE is plotted in blue. Results were observed for t = 200 at steady state. Ligand and receptor initial values set to $L_0 = R_0 = 1.5$.

6.4.3 Comparing DLE and decision tree approach using a model of caspase activation

The caspase activation model of Aldridge et al. (Fig. 5.3) was used in order to compare the developed method to a different approach of multivariate model analysis for signaling networks, called Direct Lyapunov exponent (DLE) analysis.

Direct Lyapunov exponent (DLE) analysis is a multivariate method for local sensitivity analysis. Inspection of DLE values enables identification of regions in phase space (the set of initial chemical species concentrations) that give rise to different system responses, here survival or apoptosis, respectively [4] (Fig. 6.16). The caspase model contains eight molecular species including pro-apoptotic procaspase-3, caspase-3, procaspase-8 and caspase-8 and pro-survival factor XIAP as well as three intermediate products (Fig. 5.3).

Cells were previously categorized as apoptotic if they exhibit a relatively tall and wide pulse of active caspase-3 [4]. Figure 6.13 shows examples of caspase-3 trajectories representing apoptosis and survival. Accordingly, we set the definition for classification of parameter sets to the class 'apoptosis' (class 1), if their resulting caspase-3 trajectories

exhibit a pulse lasting for at least two hours at a level higher or equal to $1.5 \cdot 10^{+05}$ molecules per cell and to 'survival' (class 0) otherwise.

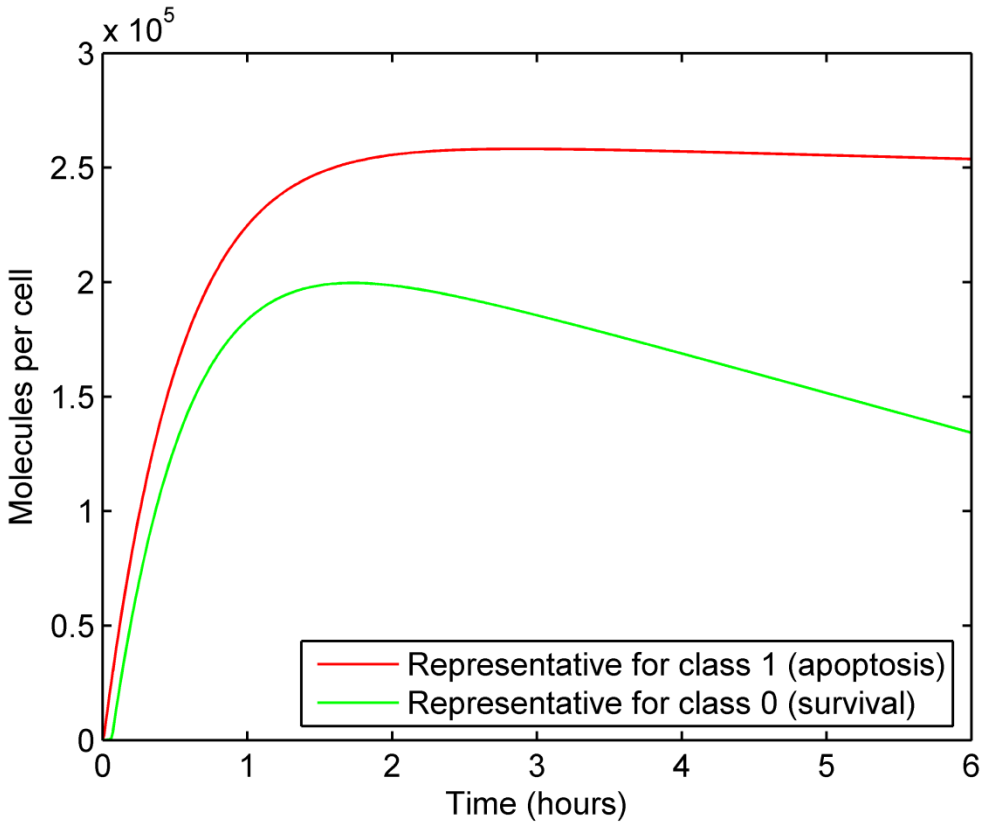

Figure 6.13: Classification criterion for the caspase-3 activation model

Rate constants were set to original values and similar ranges for initial conditions of component concentration were used as originally proposed [4]. While initial values for intermediate products (active caspase-3 and any active caspase bound to a pro-caspase or XIAP) are generally assumed to be zero in our approach, the presence of active caspase-8 was necessary in this case to initiate the caspase cascade and hence was set to the same range as published by Aldridge et al. [4]. In total 50,000 parameter sets were generated for initial concentrations of four components (procaspase-8, active caspase-8, procaspase-3 and XIAP). From these parameter sets, 3012 (1506 per class) were selected for training of the decision tree. Details of the simulation can be found in Appendix B

The misclassification errors of all sub-trees obtained by the decision tree algorithm suggest a tree containing six leaf nodes as the tree on the best pruning level.

However, even decision trees with fewer terminal nodes show good performance as indicated by the decreasing misclassification error (Fig. 6.14).

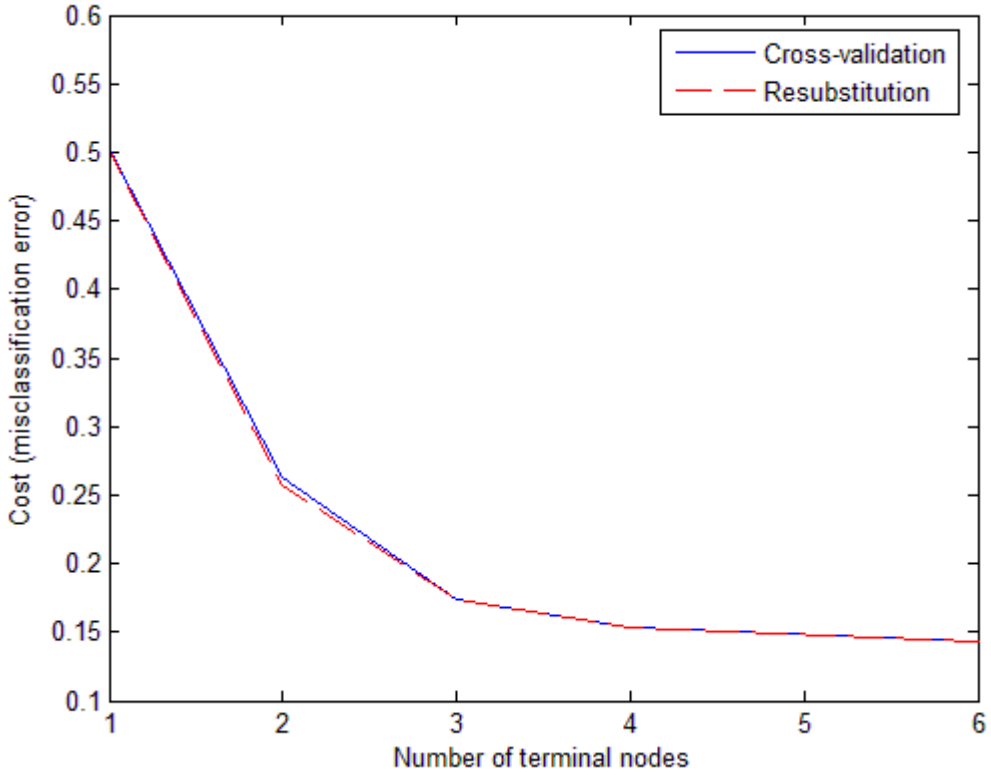

Figure 6.14: Error decline for capsase-3 activation model (Aldridge et al.)

We consider an accuracy of 80% to suffice for identification of conditions for cell fate decision. Therefore we here present decision trees with a corresponding misclassification of about 20% and more accurate trees in the supplementary information. The full tree is shown in Appendix B.

The top decision node suggested by our analysis identifies the relation of active caspase-8 and XIAP as the most important feature (Fig. 6.15) which accounts for the reduction of the misclassification error to 25.7% (Fig. 6.14).

The balance between active caspase-8 and XIAP has also been identified by Aldridge et al. to influence cell fate: The higher the concentration of active caspase-8, the higher XIAP levels are needed to cross the separatrix of large DLE values from death to survival [4] (Fig. 6.16). The same notion is reflected in our results by the top decision node containing the relation of active caspase-8 and XIAP (Fig. 6.15) where a XIAP level exceed-

ing that of active caspase-8 is linked to survival, while caspase-8 concentration higher than XIAP is associated with apoptosis.

The leaf node representing the class “survival” in Fig. 6.15A yields a class probability of 1.0 and therefore corresponds to a pure subset due to the split of the top decision node.

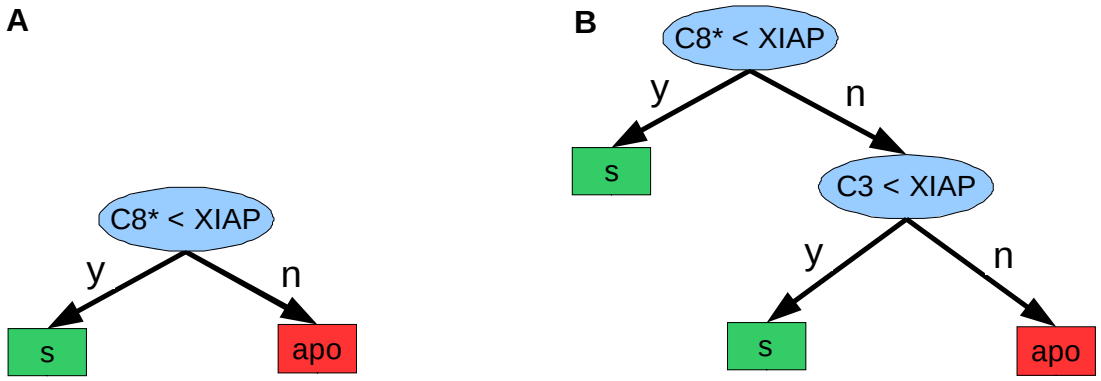

Figure 6.15: Decision trees resulting from analysis of the caspase-3 activation model
Decision trees resulting from analysis of the caspase activation model illustrated in Fig. 6 for a misclassification error of 25.7% (A) and 17.43% (B).

However, caspase-8 exceeding XIAP can still lead to survival in case procaspase-3 concentration is lower than XIAP, as indicated by addition of the second decision node (Fig. 6.15B) leaving 17.43% parameter sets as wrongly classified to cause apoptosis.

The second decision node in combination with the root node reflects the need for enough procaspase-3 as substrate for cleavage by caspase-8 to high levels of active effector caspase-3, which is required for apoptosis.

Although our method suggests a multivariate result by combinations of decision nodes to lead to a decrease of the misclassification error which can be interpreted towards the specified system response, the strongest reduction of misclassification is achieved by the top node able to classify 74.3% of the generated parameter sets correctly. Our method thus selects as most important feature for classification the relation of XIAP and caspase-8, which agrees with the result of Aldridge et al. for analysis of the caspase-3 activation model. Compared to the DLE, our method does not require visualization and exploration of phase space. Instead it directly suggests simple rules yielding the most important molecule relations on top of the decision tree.

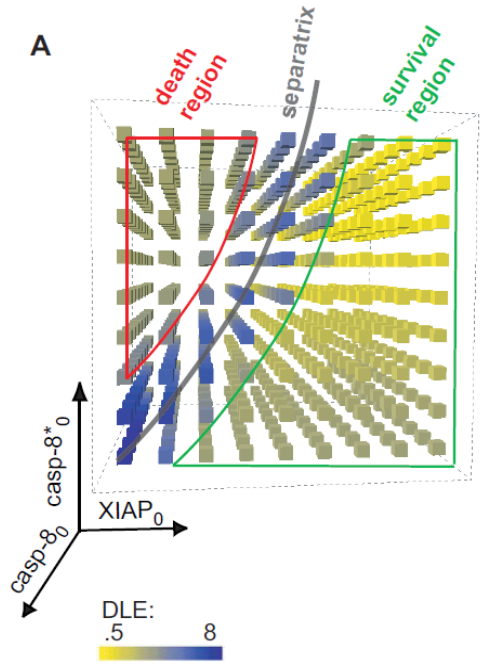

Figure 6.16: Visualization of DLE values (from Aldridge et al. [4])

6.4.4 Predictions for FAS-mediated apoptosis

The decision tree approach for model analysis was applied to the apoptosis model of Neumann et al. [73] illustrated in Fig. 5.4. Experiments showed, that a ligand concentration of 250 ng/ml results in a cell death rate of more than 70% [73]. For a ligand concentration of 250 ng/ml, the trajectory of active caspase-3 (C3*) was inspected which reaches a maximum concentration of 0.05 nM with an initial value for procaspase-3 (C3) of 1.5 nM leading to a ratio of $\max(C3^*)/C3_{y0} = 0.03$. Therefore, class definition was set in collaboration with L. Neumann to:

class 1 (apoptosis), if $\max(C3^*)/C3(t_0) > 0.02$
class 0 (survival) otherwise

The decision tree pruned to the best level requires four decision nodes and five leaf nodes yielding a misclassification cost of 0.048 (Fig. 6.17). Since we consider a misclassification error of 0.2 as sufficient to give an insight into the interactions predicted to lead to cell death, corresponding subtrees will be presented.

The top decision node (Fig. 6.18A) resulting from our model analysis approach suggests the balance of two DED-containing species, procaspase-8 and c-FLIP$_S$ as the most important relation accounting for a reduction of the misclassification error to 0.28.

The decision node on the second level further decreases the misclassification to 0.15 and suggests higher c-FLIP$_L$ compared to the receptor-FADD component to lead to survival, given procaspase-8 levels exceeding c-FLIP$_S$ (Fig. 6.18B).

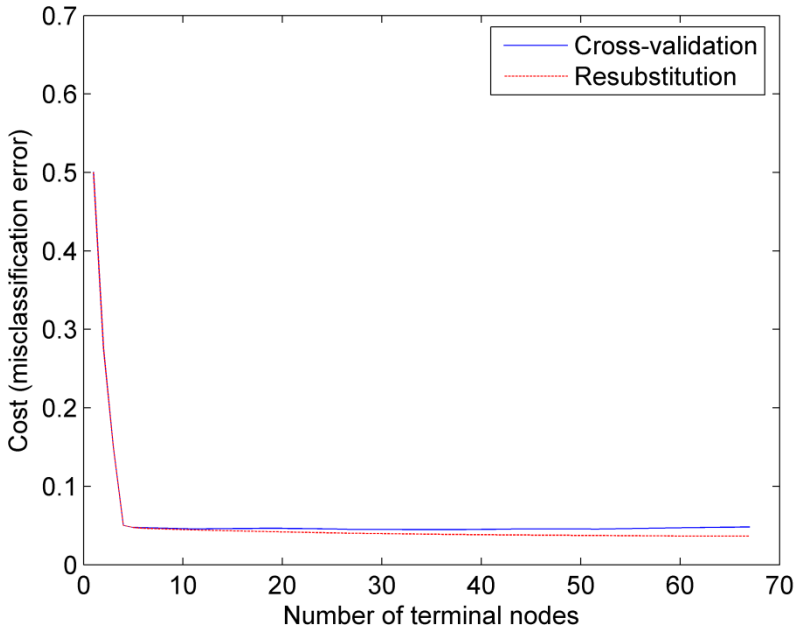

Figure 6.17: Misclassification error for model analysis of CD95/Fas-model

The top decision node suggesting survival for initial concentrations of c-FLIP$_S$ levels higher than procaspase-8 thus reflects the notion of competitive binding of c-FLIP$_S$ and procaspase-8 to the DISC hampering caspase-8 activation.

The role of c-FLIP$_L$ has been discussed to be anti-apoptotic at high concentrations [20] [56] or pro-apoptotic at low concentrations [20]. A recent study of Fricker et al. revealed experimentally, that not solely the amount of c-FLIP$_L$, but receptor stimulation influences the pro- or anti-apoptotic effects of c-FLIP$_L$. [36].

The model component representing receptor stimulation in the analysed CD95-model of Neumann et al. is the receptor-FADD (RF) component [73]. Hence, the relation of RF

and c-FLIP$_L$ as the second decision node represents the dependence of receptor stimulation on c-FLIP$_L$ for cell death.

Further, the top two decision nodes combined with each other reflect the finding that, over-expression of a short c-FLIP-isoform in combination with over-expression of c-FLIP$_L$ strongly inhibits apoptosis enabling an even higher percentage of cells to survive than either of the c-FLIP isoforms alone [36]. Furthermore, DISC components, specifically c-FLIP isoforms have been previously identified by sensitivity analysis as most influential components. Hence my method applied to a model of CD95-mediated apoptosis agrees with previous experimental and computational findings that DISC components, in particular c-FLIP isoforms are influential components with regard to cell death.

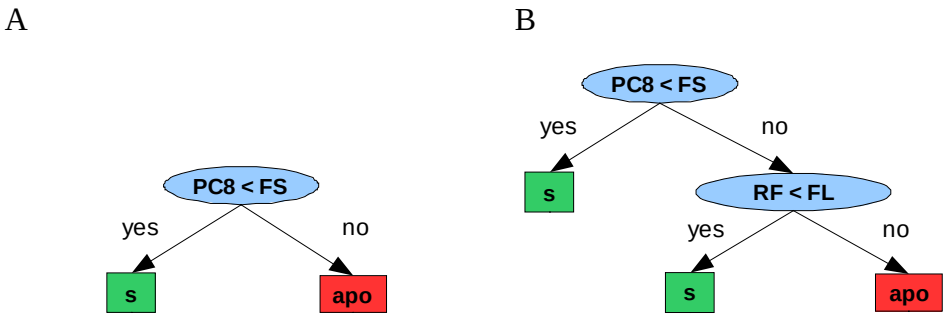

Figure 6.18: Resulting decision trees for analysis of CD95/Fas - model

6.4.5 Prediction for a comprehensive apoptosis model

The decision tree approach for model analysis was applied to the comprehensive model of Albeck et al. [3] with a class definition based on their observations on delay time (T_d) and switching time (T_s) for caspase-3 activity using PARP levels as explained in section 5.2. Additionally, it was required that at least 50% of PARP has to be cleaved for the classification of the system response. Therefore classes for categorizing generated parameter sets were set to:

class 1 (apoptosis) if T_s < 30 minutes and 50% PARP is cleaved
class 0 (survival) otherwise.

We applied our method to EARM v1.0 by generating 100,000 random parameter sets within a range of four orders of magnitude centred at original start values for 18 chemical species which require non-zero start values (see Appendix B for details). We then choose

93,046 parameter sets (46,523 per class), as input for the decision tree algorithm. Misclassification error shows an early decline and a good classification performance for less than ten nodes (Fig. 6.19).

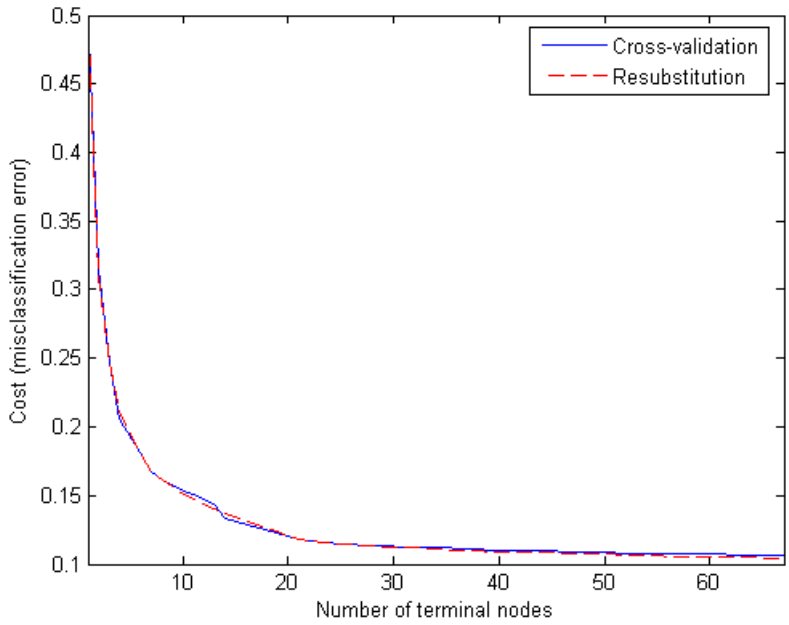

Figure 6.19: Misclassification error for model analysis of EARM v1.0

Again, we are interested in the conditions required for apoptosis instead of building the most accurate classifier and have a closer look at all subtrees up to a misclassification error of 0.2.

Subtree containing one decision node

Our approach identifies the relation of XIAP compared to the model component Smacm representing mitochondrial Smac as the top node accounting for a misclassification error of 0. 309. The top node suggests survival for Smac exceeding XIAP and apoptosis otherwise (Fig. 6.20A).

Due to the model structure of EARM v1.0, the component Smacm is not formed without pore formation of the mitochondria (component M*). Mitochondrial outer membrane permeabilization (MOMP), which has been shown by experiments to be the crucial event

for effective apoptosis [3] [2], therefore is an implicit prerequisite for the top decision node.

Smac/DIABLO (Direct IAP Binding Protein With Low pI) is a protein released from the mitochondria, which physically interacts with IAPs relieving the inhibitory effect of IAPs on caspase-3 and thereby sensitizing cells for apoptosis [92] [31]. It was shown by Verhagen et al. that the ratio of IAP to DIABLO determines whether cells undergo apoptosis or not. Smac and XIAP were integrated into EARM v1.0 as direct interaction partners by Albeck et al. [3]. The top decision node therefore identifies a pathway interaction integrated into the analyzed comprehensive apoptosis model as well as a relation that has been previously shown by experiments to influence cell fate.

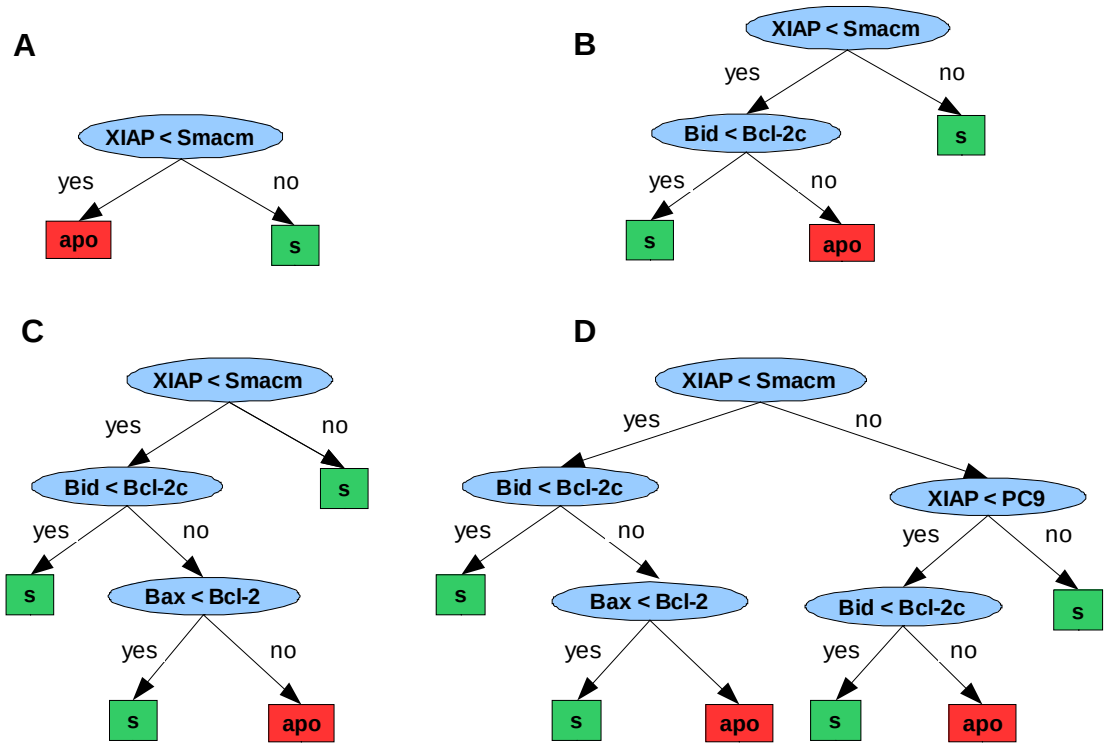

Figure 6.20: Decision trees resulting from analysis of apoptosis model EARM v1.0
Decision trees resulting from analysis of apoptosis model EARM v1.0 (Albeck et al. [3], illustrated in Fig. 5.1) for a misclassification error of 30.85% (A), 24.68% (B), 21.17% (C) and 18.13% (D).

Subtree containing two decision nodes

The second decision node predicts higher Bid compared to Bcl-2c (denoting Bcl-2 molecules in the cytosol) concentrations to lead to cell death, but Bid levels lower than Bcl-2c to be able to save cells from apoptosis when mitochondrial Smac still exceeds XIAP. These two decision nodes account for a misclassification of 0.247. Like the top node, the

second decision node contains known apoptosis antagonists [94] (Fig. 6.20B). Bid was also integrated into the model as a component sequestering cytosolic Bcl2 [13].

Subtree containing three decision nodes

In case both pro-apoptotic conditions are met, Smacm exceeding XIAP as well as Bid exceeding Bcl-2c, concentrations of Bax higher than mitochondrial Bcl-2 are suggested by the decision tree to lead to apoptosis and Bax lower than mitochondrial Bcl-2 to lead to survival. The third level node contains Bax and Bcl-2 (mitochondrial Bcl-2) and accounts for 0.211 of misclassification rate.

As for the first and the second node, the third decision node selected by our method suggests relations of molecular species which have been integrated into the model as direct interaction partners representing antagonists for cell fate based on previous experimental results [74].

Subtree containing five decision nodes

The alternative path from root to bottom assuming XIAP levels higher than Smacm concentrations combines two decision nodes within one pruning step (XIAP versus procaspase-9 and Bid versus Bcl2-c) (Fig. 6.20D). For concentrations of XIAP exceeding both Smacm and procaspase-9, cell survival is predicted. For concentrations of XIAP lower than procaspase-9, again Bid concentration higher than Bcl-2c levels are connected to apoptosis whereas molecule concentrations of Bid lower than Bcl-2c levels are associated with survival resulting in a misclassification rate of 0.181.

Compared to the known antagonists Bid and Bcl-2, the importance of the apoptosis inhibitor XIAP versus pro-apoptotic caspase-9 is less obvious, since these molecules are not direct interaction partners of EARM v1.0. However, within the analyzed apoptosis model, caspase-9 is involved in formation of the apoptosome followed by caspase-3 activation and XIAP is modeled as a direct inhibitor of caspase-3. Hence, the decision node containing the relation of XIAP and procaspase-9 identifies the balance between caspase-3 activation and a caspase-3 inhibition as an important relation for apoptosis.

For XIAP exceeding both Smacm and procaspase-9, the decision tree suggests that Bid concentrations exceeding Bcl-2c can overcome XIAP inhibition and lead to apoptosis by Bid sequestering cytosolic Bcl2. In case active caspase-8 is present, Bid can be cleaved to

tBid and further enhance the activity of the intrinsic pathway for a quick and effective cell death.

Our method applied to the comprehensive apoptosis model EARM v1.0 thus identifies component relations as decision nodes containing molecules acting as antagonists of death or survival.

Decision nodes suggested by our method not only represent crucial interactions within the mathematical model, but also important molecule ratios for cell fate identified by experiments in previous studies as will be explained in the following section.

6.4.6 Predictions proved in experimental data and literature

The relations suggested by the decision tree depicted in Fig. 6.20 were checked computationally in the OralTongue gene expression data set. Analysis of this data set resulted in the lowest *p*-value for the *globaltest* (Tab. 6.2) and hence was expected to most likely reflect decision tree relations. However, relations of the decision tree nodes found to be for anti-apoptotic components more abundant in normal samples than pro-apoptotic components and pro-apoptotic components more abundant in tumor samples than anti-apoptotic components (Fig. 6.21).

Smac was not measured on the microarray used to generate the OralTongue data set, therefore the focus in this section is on predictions with regard to Bcl2-family members.

Since gene expression data does not reflect in which cellular compartment molecules reside, hence model components Bcl2c and Bcl2 were both compared to gene expression levels of probes corresponding to Bcl2.

The relation “Bid < Bcl2” chosen as second and fifth decision node (Fig. 6.20B, Fig. 6.20D) was found in 78% of the normal samples and in 19% of the tumor samples (Fig. 6.21) for one pair of probes. Hence anti-apoptotic Bcl2 exceeds pro-apoptotic Bid more frequently in normal samples compared to tumor cells. Another pair of probes corresponding to the relation “Bid < Bcl2” shows relative frequencies of 15% and 3% for normal and tumor respectively. Here, most samples showed higher values for pro-apoptotic Bid compared to anti-apoptotic Bcl2 in both, normal and tumor. However the relation Bid < Bcl2 associated with apoptosis resistance still was five times more frequently found in normal than in tumor tissue (Fig. 6.21).

Bax and Bcl2 are represented by several pairs of probe sets, where all relations “Bax < Bcl2” are more often found in normal than in tumor samples. Here again anti-apoptotic Bcl2-levels are higher than pro-apoptotic Bax in normal samples (Fig. 6.21).

A further check within other gene expression data sets was conducted by comparing the decision nodes of the tree in Fig. 6.20 with gene expression relations listed in Table 6.4 and in Appendix A.

Again, Bcl2 was higher expressed than pro-apoptotic family members in normal, whereas Bax and Bid exceeded Bcl2 expression in tumor samples.

These findings seem counterintuitive, since due to the gene expression data, gene expression levels of pro-apoptotic factors are up-regulated compared to their ant-apoptotic antagonists in tumor cells. According to these results, tumor cells seem to be more prone to apoptosis than normal cells. Hence the predicted relations could not be confirmed as responsible to apoptosis resistance for the given data sets.

In spite of these findings on the gene expression level, previous experimental findings agree with decision tree relations to control cell fate decisions. For that however, molecule relations controlling death and survival were tested on the protein level.

With regard to the top decision node relation in Fig. 6.17, it was shown by Verhagen et al. that the ratio of IAP to DIABLO could determine whether cells undergo apoptosis or not [92]. Further, Bid has been shown to sequester cytosolic Bcl2 [13]. And in case active caspase-8 is present, Bid can be cleaved to tBid in order to enhance pro-apoptotic signaling by activation of the intrinsic pathway for a quick and effective cell death.

Also, Bax and Bcl-2 (mitochondrial Bcl-2) were previously identified by Oltvai et al. [74] to interact and influence cell death.

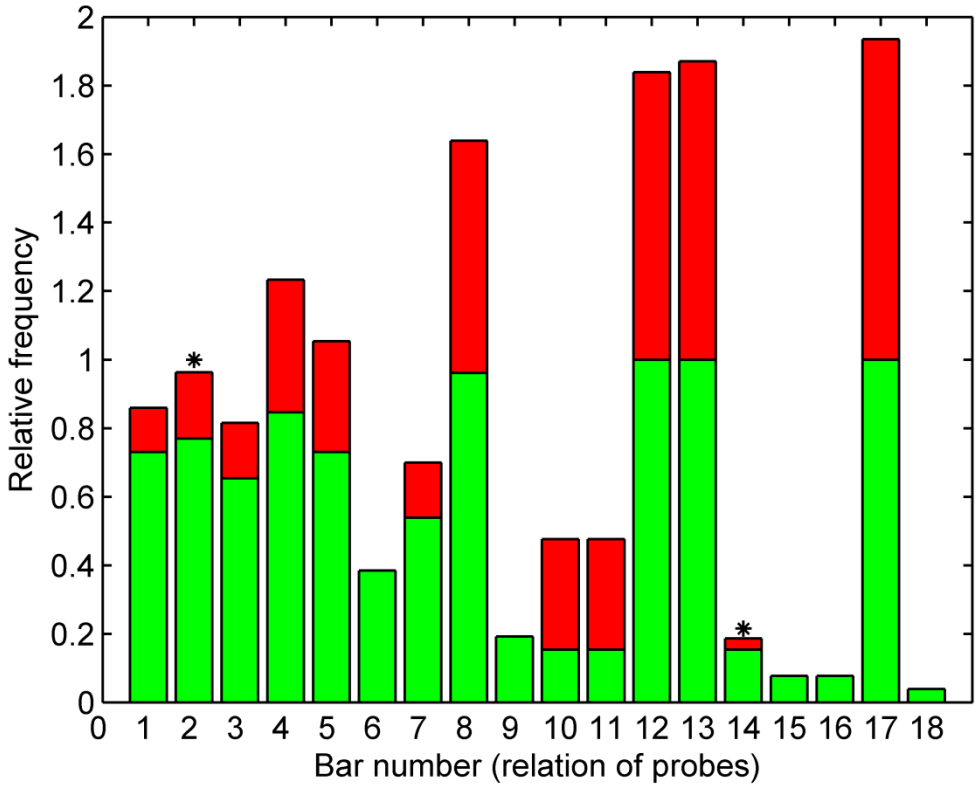

Figure 6.21: Relative frequencies of predicted relations in tumor and normal
Each bar represents a relation of gene expression values of the OralTongue data set. Probes corresponding to each bar are listed in Tab. 6.8. The percentage of normal and the percentage of tumor samples that showed the corresponding relation of expression levels are colored in green and red respectively. Bars marked with an asterix (*) correspond to the relation Bid < Bcl2. All other bars illustrate relative frequencies corresponding to Bax < Bcl2. The figure shows that anti-apoptotic Bcl2 exceeds pro-apoptotic Bcl2-family members (Bax and Bid) in normal cells compared to tumor cells. The reverse notion is found for tumor cells. Here pro-apoptotic Bax and Bid exceed Bcl2. Missing red areas (bars 6, 15, 16 and 18) indicate, that all tumor samples showed the relation Bcl2 < Bax. Probe pairs that do not show any differences between normal and tumor were not considered.

Table 6.8: Probes corresponding bars illustrated in Fig. 6.21
Each bar illustrated in Fig. 6.21 represents a comparison of gene expression values. These values correspond to a relation of certain apoptosis pathway components and are measured as two probe as listed below.

Bar	Relation	Probes	Bar	Relation	Probes
1	BAX < BCL2	1997_s_at < 2038_g_at	10	BAX < BCL2	1998_i_at < 1910_s_at
2	BID < BCL2	32725_at < 2060_at	11	BAX < BCL2	1998_i_at < 2038_g_at
3	BAX < BCL2	1997_s_at < 1910_s_at	12	BAX < BCL2	2066_at < 1910_s_at
4	BAX < BCL2	2067_f_at < 2060_at	13	BAX < BCL2	2066_at < 2038_g_at
5	BAX < BCL2	2065_s_at < 1847_s_at	14	BID < BCL2	32725_at < 1847_s_at
6	BAX < BCL2	1997_s_at < 2060_at	15	BAX < BCL2	2065_s_at < 1909_at
7	BAX < BCL2	2067_f_at < 1847_s_at	16	BAX < BCL2	2067_f_at < 1909_at
8	BAX < BCL2	2065_s_at < 2060_at	17	BAX < BCL2	2067_f_at < 1910_s_at
9	BAX < BCL2	1997_s_at < 1847_s_at	18	BAX < BCL2	1997_s_at < 1909_at

7 Discussion and Outlook

Studies that investigate disease and clinical phenotypes generally use statistical methods on high-throughput data to identify genes or groups of genes associated with the disease state. Researchers interested in cellular phenotypes or processes however, investigate molecule interaction and dynamics in order to uncover mechanism. These two types of study approaches have so far unconnectedly coexisted.

This thesis combined statistical analyses and mechanistic analyses in a new way.

The motivation for the presented work was to understand interactions and mechanisms involved in apoptosis in the context of cancer development. Therefore, analysis of gene expression data from clinical phenotypes was combined with simulation of apoptosis and analysis of the resulting system response. To this end two approaches were followed, which includes the development of two scaling methods and a novel method for model analysis of a system of differential equations.

7.1 Scaling start values according to gene expression data

The first approach investigated if gene expression data reflects pathway disruptions and consequently results in a system response different for values scaled according to tumor and normal samples respectively. Two methods for scaling initial values of model components were developed and start values were scaled according to eight different data sets, which did not yield clear differences with regard to the switching time T_S of PARP time courses, a characteristic associated with apoptosis.

Scaling of initial values of the apoptosis model using the fold change of gene expression values did not yield differences with regard to T_S. Although start values after scaling were much larger or lower than original values, this only had an effect on the delay time T_d. The switching time, however, remained nearly unchanged for all tested data sets. The fold change-scaled start values result from averaging several samples per class. Hence, effects on the gene expression level could have been diminished for obtaining scaling factors. Therefore a second scaling method was developed.

In order to avoid effects from averaging, scaling per sample was applied to original values. Here, for each sample a separate system response could be observed, which similar

did not show remarkable differences between tumor and normal samples with respect to system output of apoptosis.

These results can be explained by different hypothesis. Molecule levels represented by microarray data may not reflect protein concentrations. Apoptosis is a process based on cell signaling on a time scale of minutes or hours, gene transcription sometimes proceeds on a slower scale. In this work, gene expression measurements were assumed to be proportional to steady state levels of apoptotic protein components. However several mechanisms might interfere with this correlation [41].

Reasons for this could be post-transcriptional regulation or protein degradation, both influence the amount of functional signaling molecules and thus may eliminate the correlation between mRNA and protein levels. Using protein data ideally obtained by quantification of single cells could help to elucidate the effect of different molecule abundances on system response, but is currently too challenging to be performed on a genome-wide scale.

Another explanation for the observed results could be that the model of Albeck et al. [3] is optimised for HeLa cells and may not perfectly represent apoptosis in cells within their natural microenvironment, such as in a tumor. The model would be need to be adapted to this situation by exploring time series of protein data from patients and healthy individuals to fit parameters according to this system. Again, this is currently far beyond the scope of available experimental methods.

Complementary analysis of gene expression data by ranking molecule relations revealed anti-apoptotic Bar (BFAR) to be more abundant in cancer samples than pro-apoptotic caspase-8 in the breast cancer data set. In this case an anti-apoptotic component exceeds a pro-apoptotic component in tumor samples, which agrees with the assumption that cancer cells evade apoptosis due to abundant anti-apoptotic components. Together with the finding of Stegh et al. [88] who showed that BFAR inhibits caspase-8 in the breast cancer cell line MCF-7 and further prevents the cleavage of Bid and caspase-3, this suggests that the identified molecule relation contributes to apoptosis resistance.

Previous studies with a focus on signaling pathways have so far emphasized receptor state of cells as crucial for progress of breast cancer [43]. The finding based on gene expression data from cancer patients indicates that the connection between extrinsic and intrinsic pathway represented by Bid as well as the core machinery of apoptosis (caspases)

might be affected by BFAR and thus are responsible for impaired apoptosis in breast cancer. This result was not detectable on the level of fold changes, though, since these were close to 1 for all analyzed apoptosis components for this data set.

Fold changes for all data sets identified most pro-apoptotic components in cancer to be up-regulated, and most anti-apoptotic pathway components to be down-regulated compared to normal samples, which is against intuition given common apoptosis resistance in cancer. Again, this could be due to incomplete correlation between mRNA and protein concentrations, to disruption of apoptosis by mechanisms that are not reflected by expression changes (such as inactivating mutations), or to neglect of our models of further pathways that may cross-talk to apoptosis.

Subpathways in terms of extrinsic and intrinsic components were inspected and in case corresponding expression levels were up- or down-regulated, subpathways were considered as turned on or off, respectively. This analysis revealed the intrinsic pathway to be turned on in prostate cancer as well as in the HeadAndNeck data set, whereas the extrinsic pathway was turned on in CML.

Prostate cancer cells can be effectively killed via the intrinsic pathway by indole-3 [72]. This might be due to up-regulation of the intrinsic pathway in prostate cancer cells. A suggestion for further investigation is to test indole-3 also for head and neck cancer. In the case of CML the extrinsic pathway could be used as selective mechanism similar to the example of prostate cancer.

Although scaling of start values did not lead to a different system response for tumor and normal samples, the approach “from data to model” including complementary gene expression data did give an insight into parts of the apoptosis pathway, which should be investigated in more detail on the protein level. For breast cancer, prostate cancer as well as head and neck cancer, specific points of potential pathway intervention were suggested. Here, data from patients and not from cell lines were used, which increases the chances to find important apoptosis traits in vivo. This might help to develop specific therapies according to the type of cancer, or even for an individual patient.

7.2 Model analysis to generate hypotheses

A novel method for model analysis was developed in the course of this thesis. It suggests conditions for a certain system response given a particular mathematical model of a

pathway, and a definition of the system response. Here, the model is assumed to be a priori correct. Molecule relations instead of absolute values are considered as crucial conditions for pathway activation. The latter is based on the notion that threshold values for molecule abundances exist. In order to find parameter relations which lead to pathway activation, the parameter space for initial concentrations is scanned by randomly generating parameter sets. Sampled parameter sets are used for pathway simulation and classified as pathway-activating or non-activating sets based on a predefined criterion. These parameter sets together with the corresponding, binary system response are then subjected to a decision tree algorithm which learns parameter relations and results by inferring rule sets dissecting the parameter sets into the pathway activation classes.

Learning of ratios instead of concentrations instead of using each absolute concentration has the consequence that the method subdivides the input space by diagonal cuts instead of cuts orthogonal to variable axis (Fig. 3.1). Using probability distributions defined on the entire input space, this method is global and multivariate. It further yields intuitively understandable results and is adaptable to the particular research question. The dynamic analysis is crucial to analyze pathway activation, since it has been shown that steady states are not necessary for pathway activation e.g. for apoptosis [4].

7.2.1 Decision Trees for model analysis compared to alternative methods

My method for multivariate model analysis was applied to two models for technical validation. I compared the results to steady state analysis and calculation of DLEs, two alternative approaches for model analysis. My method was able to find previously identified conditions.

The first case was a model for EGFR internalization. Here, the decision tree method identifies component relations as required for inactivation and activation of the CIE-pathway, which have previously been derived analytically to be responsible for an ultrasensitive sorting mechanism (Schmidt-Glenewinkel et al. [85]). However, comparison of methods showed that it is crucial which criterion is chosen as an indicator for pathway activation. In the case of the EGFR-model, previously identified conditions by Schmidt-Glenewinkel et al. for a steady state associated with full pathway activation do not consider the resulting levels of the internalized receptors. Hence steady state analysis allows very low levels

of internalized receptors to be associated with full pathway activation. Instead, the decision tree method does consider levels of internalized receptors. Moreover, conditions for activation suggested by the decision tree match well with conditions of input region according to the predefined criterion for parameter classification. It thus better conforms to the initial, pre-defined condition for pathway activation than the analytical solution. The decision tree method yields simplified results by clear diagonal separations of the input space. Non-linear relations of components to influence pathway activation cannot be captured. This seems like a drawback, but instead increases the interpretability of the resulting decision tree containing molecule relations within decision nodes.

By using a different machine learning algorithm for analyzing generated parameter sets and its association with a system response, more complex interrelations could be identified, yet the intuitive representation would be lost. In a similar way as accuracy is traded for simplicity to get small decision trees, here we trade accuracy of fitting a predefined criterion for simplicity to get an easy to interpret representation of molecule conditions controlling pathway activation.

In a second use case, the decision tree approach was applied to the caspase activation model of Aldridge et al. [4] to compare the developed method to a different approach of multivariate analysis. The resulting decision tree identified as the top node the balance of active caspase-8 and XIAP to be important for cell fate as previously reported by Aldridge et al. [4]. Although our method suggests a multivariate result by combinations of decision nodes to lead to a decrease of the misclassification error, the strongest reduction to a misclassification of 0.26 is achieved by the top node alone.

Compared to the DLE approach originally used by Aldridge et al. [4], our method is faster and does not require visualization and exploration of phase space. Instead it directly suggests simple rules yielding the most important molecule relations on top of the decision tree.

7.2.2 Analysis of apoptosis models by decision tree method

Two biologically relevant models of apoptosis were analysed which resulted in protein relations previously reported to highly influence death and survival. First, the developed method for model analysis was applied to the CD95/Fas-induced apoptosis model of Neumann et al. [73]. I demonstrated that the method can predict biologically relevant and

experimentally proven relations to be important for cell death and yields a strategy to trigger cell death suggesting the relation of the Fas ligand to $Flip_L$ as crucial.

DISC-proteins, in particular Flip isoforms have been identified previously both by sensitivity analysis and experiments to be the most influential components of CD95/Fas induced apoptosis [73]. Sensitivity analysis however identifies only single proteins as influential. Decision trees using relations as features instead suggest for the CD95/Fas-model interrelations of DISC components, which comply with recent experimental findings of Fricker et al. for receptor stimulation and over-expression of a short in combination with a long c-FLIP-isoform to influence apoptosis [36].

In the case of the comprehensive apoptosis model of Albeck et al. [3] the decision tree method confirms the role of already known antagonists and in addition reveals strategies to trigger cell death of indirectly interacting components. Experimental results from literature agree with the identified direct antagonists. Experiments of Verhagen et al. showed that the ratio of IAP to DIABLO could determine whether cells undergo apoptosis or not [92]. Further, Bid and Bcl2, contained in the second decision node have been shown to interact [13], where Bid is a known mediator between the intrinsic and extrinsic pathway. Also, Bax and mitochondrial Bcl-2 have been previously identified by Oltvai et al. [74] to interact and influence cell death which is suggested by another decision node. How the balance of procaspase-9 and XIAP influence cell fate remains to be proven experimentally. XIAP is considered to be a strong caspase-3. The prediction, that procaspase-9 levels exceeding XIAP in combination with Bax exceeding Bcl-2 will lead to apoptosis implies, that XIAP inhibition can be overcome. Activation of caspase-9 relies on MOMP followed by apoptosome formation. Bax is considered to be responsible for mitochondrial pore formation. The decision tree therefore identifies a feedback relation which includes components that do not directly interact, but nevertheless depend on each other for effective cell death. It would be therefore interesting to experimentally test, if the suggested molecule relations could overcome XIAP inhibition.

7.2.3 From model to data: Conclusion and perspectives

The second approach of in silico identification of a set of molecule relations crucial for a defined system response has been shown to identify important conditions for pathway activation. Here the simplicity of resulting decision tree using molecule relations as split-

ting criteria enables us to interpret system behaviour for its underlying mechanisms in terms of rule sets.

A more complex method for model analysis would be able to capture more complex effects. However, it would be more difficult to interpret the input/output behaviour of a system.

Further improvements could be done by introducing a fuzzy region for component abundances as threshold. Since the parameters are highly unlikely to take the same values if randomly drawn from uniform distributions covering four orders of magnitude, the categorization of relations (step 3 within the method explanation in section 4.2.1) comes down to a binarization. In case crisp threshold values are biologically unlikely or uncertain, a fuzzy region could be introduced for similar values.

Furthermore, a different probability density e.g. a Gaussian distribution, could be used for drawing parameter values. This would turn the approach into a truly global method and additionally could reflect certainty or uncertainty of parameter regions. Hence model analysis could be connected to another idea in systems biology: parameter uncertainly.

One might argue that the predicted interrelations had already been integrated into the model by previous knowledge from literature and therefore the method would not yield new insights. However, assuming the model interactions to be correct does not necessarily imply a higher probability for certain molecule pairs to be chosen as crucial for a certain system response. The generation of parameters drawn from a uniform distribution and selection of equal numbers for activation classes accounts for equal chances for all relation to be selected in a decision node. To argue, that analysis of a model which includes known direct interactions would not yield new insights toward reaching a certain outcome implies that the creator of the model designed the model already towards a certain system response.

Instead, model design integrates current knowledge of model interactions, which the decision tree approach correctly re-identified as important for this cellular process.

The method can of course only identify components integrated into the model, but is not restricted to components involved in the same reaction, as shown by the identified relation of XIAP and procaspase-9 for the comprehensive apoptosis model.

Changing the model structure would further challenge current knowledge and might lead to different molecule relations identified as important for apoptosis e.g. leaving out the

caspase-6 feedback loop. The resulting decision tree predictions could be tested experimentally and in that way contribute to the inherent cycle of system biology: Experiment – Model - Prediction - Experiment.

7.3 Outlook

This thesis aimed at explaining cellular mechanism in a systemic fashion with regard to clinical phenotypes of cancer. Under the assumptions that molecule abundances are important for signal transduction, that measurement of protein levels of single tumor and normal cells can be obtained and a correct mathematical model of the pathway of interest is available, the combination of the two developed approaches can foster further research in the field of cancer systems biology:

The first approach can identify perturbations in a pathway associated with a disease and hence yields characteristics which can be used for selectively kill malignant cells leaving normal cells alive. Complementary to that, the second approach identifies an effective strategy to either activate or inactivate the pathway of interest.

Both approaches in combination enable to identify a selective and effective strategy for targeted therapy which further can contribute to research in the field of personalized medicine.

8 Appendix

This part of the thesis contains additional information, results for every used data set and specifications of every analyzed mathematical model.

Appendix A shows system responses according to all given data sets. Fig. 8.1 shows PARP-trajectories resulting from *FC*-scaled start values. Fig. 8.2 shows PARP-trajectories resulting from start values scaled per sample. Table 8.1 shows the pair-wise relations of expression levels for the data set that did not show a significant *p*-value resulting from a *globaltest*.

Appendix B contains details for model simulation. The developed method for model analysis requires generation of random parameters for initial values. A uniform distribution was assumed for all sampled parameters in ranges as specified in Tabs. 8.2, 8.3, 8.4, 8.5. Sections 8.3, 8.4, 8.5 and 8.6 contain specifications of mathematical models, rate constants and time steps for simulation in Matlab. In section 8.7, trees at the best pruning level as result of the developed decision tree method are illustrated.

8.1 Appendix A

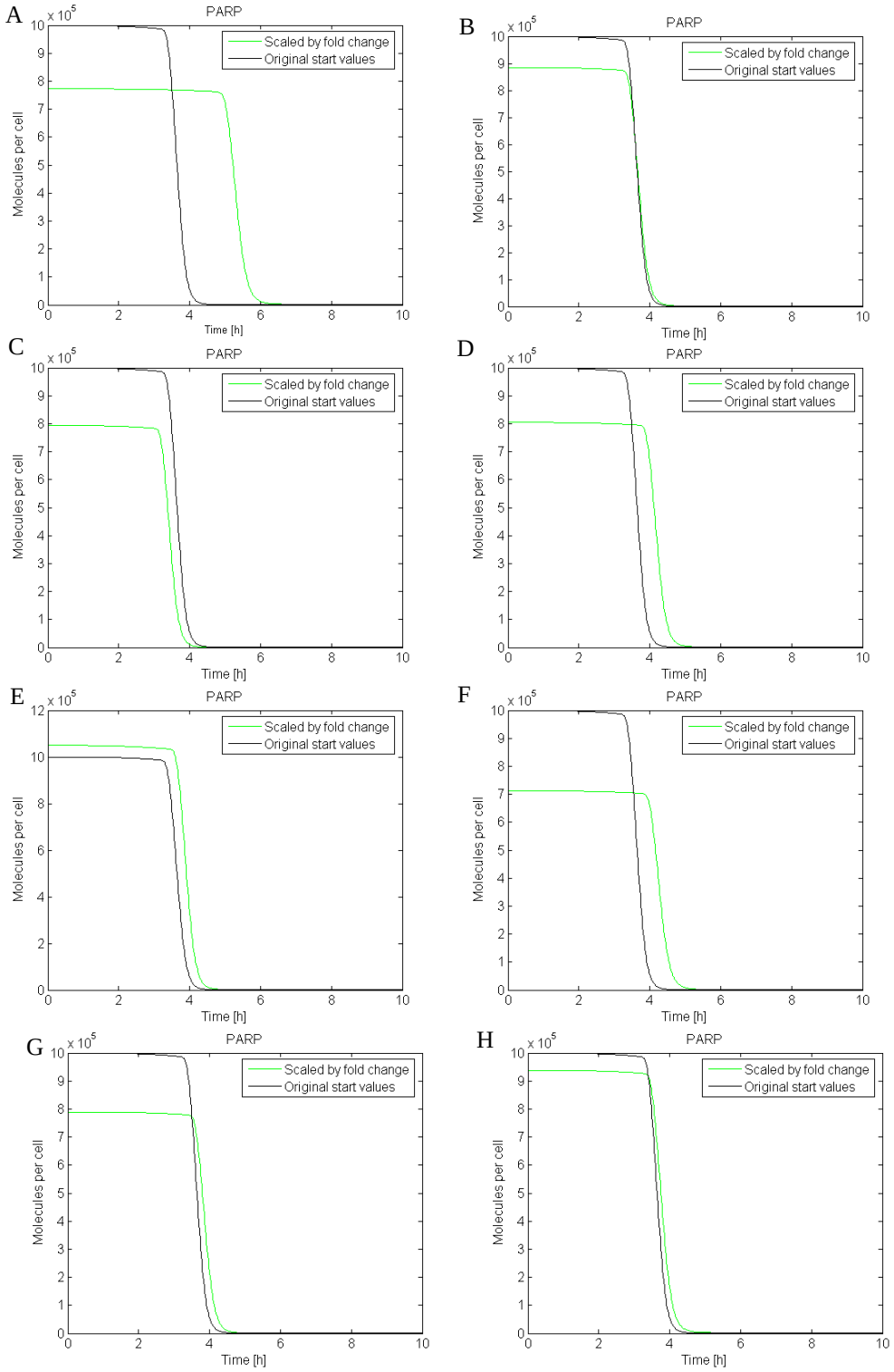

Figure 8.1: PARP-trajectories resulting from *FC*-scaling

For the data sets A: OralTongue, B: Breast, C: Lung, D: ProstateSingh, E: ProstateErnst, F: CML, G: HeadAndNeck, H: Microdissected

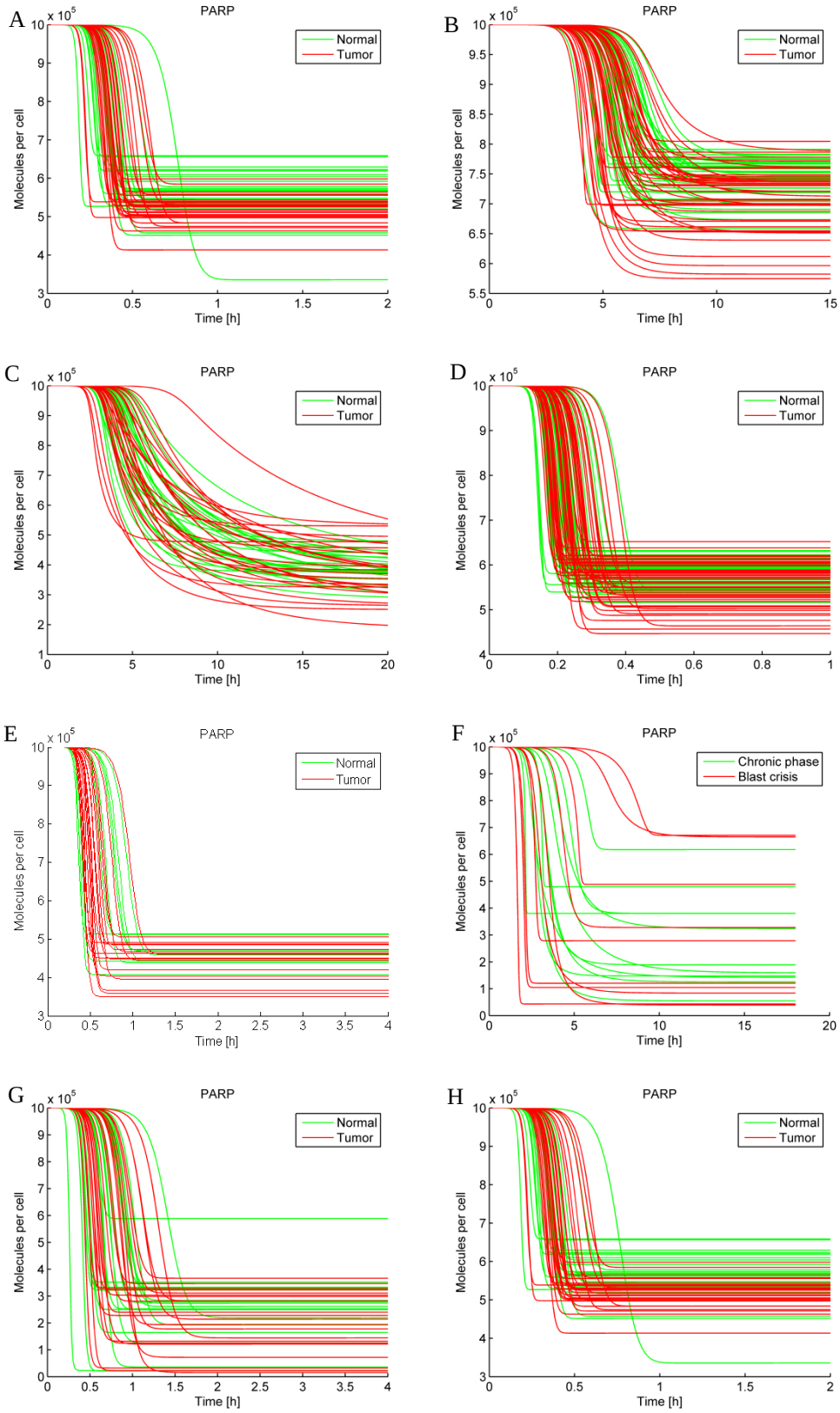

Figure 8.2: PARP-trajectories resulting from per sample-scaling
For the data sets A: OralTongue, B: Breast, C: Lung, D: ProstateSingh, E: ProstateErnst, F: CML, G: HeadAndNeck, H: Microdissected

Table 8.1: Pair-wise relations of apoptosis components in gene expression data
Orange background: Relation more often found in tumor, light green background: Relation more often found in normal, Red font: A pro-apoptotic component, Dark green font: A pro-survival/anti-apoptotic component. Same gene symbols within one relation refer to different probe set IDs of the microarray.

Pos.	ProstateErnst	CML	HeadAndNeck	Microdissected
1	CFLAR < BCL2	CFLAR < CFLAR	BAX < CASP9	CFLAR < CFLAR
2	CFLAR < BAX	CASP8 < BCL2	CASP3 < CASP9	CASP3 < BCL2
3	CFLAR < BAX	CASP8 < BCL2	BAX < BAX	CASP6 < BCL2
4	CASP6 < CASP9	CASP8 < BCL2	BAX < BAX	CFLAR < APAF1
5	BCL2 < BAX	BCL2 < BAX	CASP6 < XIAP	CASP6 < BID
6	CASP8 < CASP6	CFLAR < DIABLO	BAX < CASP9	CFLAR < BCL2
7	CASP8 < BCL2	CASP8 < CASP6	CASP3 < CASP9	CASP8 < BCL2
8	BID < BAX	CASP6 < BCL2	CFLAR < CASP9	CASP6 < CASP6
9	CASP3 < CASP9	CFLAR < CFLAR	BID < BAX	BCL2 < BCL2
10	PARP1 < CYCS	CASP6 < BAX	BCL2 < BAX	CFLAR < APAF1
11	BCL2 < BAX	BCL2 < BCL2	BCL2 < BAX	CASP6 < BAX
12	CFLAR < CASP3	TNFSF10 < CASP6	CFLAR < PARP1	APAF1 < APAF1
13	BID < CASP9	CFLAR < CASP8	CFLAR < BCL2	BID < BCL2
14	BAX < CASP9	CFLAR < CASP8	CFLAR < CASP3	CFLAR < BAX
15	BAX < CASP9	CFLAR < BAX	CASP3 < BCL2	BCL2 < CASP9
16	PARP1 < PARP1	CFLAR < CFLAR	PARP1 < BCL2	CFLAR < BCL2
17	CASP8 < BCL2	BCL2 < BCL2	BID < BCL2	CFLAR < CFLAR
18	CASP3 < CASP6	CFLAR < DIABLO	BCL2 < BAX	TNFSF10 < BCL2
19	CASP6 < APAF1	BCL2 < BAX	BCL2 < BAX	CFLAR < CASP6
20	BCL2 < BAX	BCL2 < BCL2	CFLAR < CASP9	CFLAR < BCL2

8.2 Appendix B

Table 8.2: Simulation details for the decision tree method (EGFR-model)
For the EGFR-model 10,000 parameter sets were generated out of which 1808 (904 per class) were chosen as input to the decision tree algorithm.

Variable	State	Description	Original IC	Range (Min)	Range (Max)
x_1	[EGF]	Ligand	1.5	0.1	10
x_2	[EGFR]	Receptor	1.5	0.1	10
x_3	[L:EGFR]	Ligand-Receptor Complex	0	0	0
x_4	[CIE]	CIE-adaptors	2.5	0.1	10
x_5	[L:R:CIE]	Ligand activated receptor internalized via CIE	0	0	0
x_6	[CDE]	CDE-adaptors	2.5	0.1	10
x_7	[L:R:CDE]	Ligand activated receptor internalized via CDE	0	0	0

Table 8.3: Simulation details for the decision tree method (Caspase-model)
For the Caspase-model 50,000 parameter sets were generated out of which 3012 (1506 per class) were chosen as input to the decision tree algorithm.

Variable	Symbol	Description	Original IC (Min, Max)	Range (Min)	Range (Max)
x_1	[PC8]	Procaspase-8	100, 3.5×10^5	0	3.5×10^5
x_2	[C8]	Active caspase-8	100, 1.0×10^5	0	3.5×10^5
x_3	[PC3]	Procaspase-3	100, 3.5×10^5	0	3.5×10^5
x_4	[C3]	Active caspase-3	100, 1.0×10^5	0	0
x_5	[PC8:C3]	Procaspase-8 bound to active caspase-3	100, 2.5×10^4	0	0
x_6	[C8:PC3]	Active caspase-8 bound to procaspase-3	100, 2.5×10^4	0	0
x_7	[XIAP]	XIAP (X-linked inhibitor of apoptosis)	100, 1.0×10^5	0	3.5e+05
x_8	[C3:XIAP]	Active caspase-3 bound to XIAP	100, 2.5×10^4	0	0

Table 8.4: Simulation details for the decision tree method (CD95/Fas-model)
For the CD95/Fas-model 100,000 parameter sets were generated out of which 6848 (3424 per class) were chosen as input to the decision tree algorithm.

Variable	Symbol	Description	Original IC	Range (Min)	Range (Max)
x_1	[L]	Anti-CD95 ligand	3.1322×10^{2}	3.1322	3.1322×10^{4}
x_2	[RF]	Receptor-FADD Component	91266×10^{1}	91266×10^{-1}	91266×10^{3}
x_3	[CD95L:RF]	Ligand bound to receptor-FADD component	0	0	0
x_4	[PC8]	Procaspase-8	6.4477×10^{1}	6.4477×10^{-1}	6.4477×10^{3}
x_5	[CD95L:RF:C8]	Procaspase-8 bound to ligand-receptor-FADD component	0	0	0
x_6	[FL]	Flip long	7.3986	7.3986×10^{-1}	7.3986×10^{3}
x_7	[L:RF:FL]	Flip long bound to L:RF- component	0	0	0
x_8	[FS]	Flip short	5.0839	5.0839×10^{-1}	5.0839×103
x_9	[L:RF:FS]	Flip short bound to L:RF - component	0	0	0
x_{10}	[p43/p41]	Intermediate product of caspase-8 activation	0	0	0
x_{11}	[p43-FLIP]	Product from cleavage of Flip long and procaspase-8 at the DISC	0	0	0
x_{12}	[L:RF:C8:FS]	Complex of L, RF, C8 and FS	0	0	0
x_{13}	[L:RF:FL:FL]	Flip long bound to L:RF:FL - component	0	0	0
x_{14}	[L:RF:FL:FS]	Complex of L, RF, FL and FS	0	0	0
x_{15}	[L:RF:FS:FS]	Complex of L, RF, FS and FS	0	0	0
x_{16}	[C8]	Active caspase-8	0	0	0
x_{17}	[PC3]	Procaspase-3	1.4434	1.4434×10^{-1}	1.4434×10^{-3}
x_{18}	[C3]	Active caspase-8	0		
x_{19}	[IKK]	IκB Kinase	5.7728	5.7728×10^{-1}	5.7728×10^{-3}
x_{20}	[p43-FLIP:IKK*]	Complex of p43-FLIP and IKK*	0		
x_{21}	[NF-kB:IkB]	Complex of NF-κB and IκB	0	0	0
x_{22}	[NF-kB:IkB:P]	Complex of NF-κB and phosphorylated IκB	0	0	0
x_{23}	[NF-kB*]	Active NF-κB	0	0	0

Table 8.5: Simulation details for the decision tree method (EARM v1.0-model)
For the EARM v1.0-model 100,000 parameter sets were generated out of which 93046 (46523 per class) were chosen as input to the decision tree algorithm.

Variable	Symbol	Description	Original IC	Range (Min)	Range (Max)
x_1	[L]	death ligand	3×10^3	3×10^1	3×10^5
x_2	[R]	inactive receptor complex	2×10^2	2	2×10^4
x_3	[L:R]	ligand-receptor complex	0	0	0
x_4	[R*]	active receptor complex	0	0	0
x_5	[flip]	binds to active receptor acting as inhibitor	1×10^2	1	1×10^4
x_6	[flip:R*]	complex of flip and active receptor	0	0	0
x_7	[C8]	procaspase-8	2×10^4	2×10^2	2×10^6
x_8	[C8:R*]	procaspase-8 bound to active receptor	0	0	0
x_9	[C8*]	procaspase-8 and procaspase-10	0	0	0
x_{10}	[Bar]	binds to C8* and acts as an inhibitor	1×10^3	1×10^1	1×10^5
x_{11}	[C8*:Bar]	complex of active caspase-8/10 and Bar	0	0	0
x_{12}	[C3]	procaspase-3	1×10^4	1×10^2	1×10^6
x_{13}	[C8*:C3]	complex of active caspase-8/10 and procaspase-3	0	0	0
x_{14}	[C3*]	active caspase-3	0	0	0
x_{15}	[C6]	procaspase-6	1×10^4	1×10^2	1×10^6
x_{16}	[C3*:C6]	complex of active caspase-3 and procaspase-6	0	0	0
x_{17}	[C6*]	active caspase-6	0	0	0
x_{18}	[C6*:C8]	complex of active caspase-6 and procaspase-8	0	0	0
x_{19}	[XIAP]	X-linked inhibitor of apoptosis (XIAP)	1×10^5	1×10^3	1×10^7
x_{20}	[XIAP:C3*]	complex of XIAP ans active caspase-3	0	0	0
x_{21}	[PARP]	DNA damage repair enzyme, here represents all substrates of C3*	1×10^6	1×10^4	1×10^8

x_{22}	[C3*:PARP]	complex of active caspase-3 and PARP	0	0	0
x_{23}	[cPARP]	cleaved PARP	0	0	0
x_{24}	[Bid]	substrate of active caspase-8, inactive form of Bid	4×10^4	4×10^2	4×10^6
x_{25}	[C8*:Bid]	compex of active caspase-8 and Bid	0	0	0
x_{26}	[tBid]	cleaved Bid, active form of Bid	0	0	0
x_{27}	[Bcl-2c]	represents the family of anti-apoptotic Bcl-2 proteins in the cellular compartment (CC), it binds to tBid and acts as inhibitor	2×10^4	2×10^2	2×10^6
x_{28}	[tBid:Bcl2c]	complex of cleaved Bid and Bcl2c	0	0	0
x_{29}	[Bax]	substrate of tBid, inactive form	1×10^5	1×10^3	1×10^7
x_{30}	[tBid:Bax]	complex of cleaved Bid and Bax	0	0	0
x_{31}	[Bax*]	active form of Bax	0	0	0
x_{32}	[Bax^*_m]	Bax^*_m in the mitochondrial compartment (MC)	0	0	0
x_{33}	[Bcl2]	represents all anti-apoptotic Bcl-2 proteins in the MC	2×10^4	2×10^2	2×10^6
x_{34}	[Bax^*_m:Bcl2]	complex of Bax*m and Bcl2	0	0	0
x_{35}	[Bax_2]	represents Bax^*_m:Bax^*_m (complex of two Bax^*_m) in the MC	0	0	0
x_{36}	[Bax_2:Bcl2]	complex of Bax_2 and Bcl2	0	0	0
x_{37}	[Bax_4]	represents Bax2:Bax2 (complex of two Bax_2) in the MC	0	0	0
x_{38}	[Bax_4:Bcl2]	complex of Bax_4 and Bcl2	0	0	0
x_{39}	[M]	the number of unoccupied Bcl-2 proteins in the MC	5×10^5	5×10^3	5×10^7
x_{40}	[Bax_4:M]	complex containing Bax_4 and M	0	0	0
x_{41}	[M*]	the number of pores Bax4 created on the outer membrane of the mitochondria	0	0	0
x_{42}	[CyC_m]	cytochrome C inside the mitochondria, in MC	5×10^5	5×10^3	5×10^7

x_{43}	[M*:CyC_m]	complex containing M* and CyC_m	0	0	0
x_{44}	[CyC_r]	cytochrome C released from the mitochondria, but remaining in MC	0	0	0
x_{45}	[$Smac_m$]	Smac/Diablo inside the mitochondria, in MC	1×10^5	1×10^3	1×10^7
x_{46}	[M*:$Smac_m$]	complex containing M* and $Smac_m$	0	0	0
x_{47}	[$Smac_r$]	Smac/Diablo released from the mitochondria, but remaining in MC	0	0	0
x_{48}	[CyC]	cytochrome c in the CC	0	0	0
x_{49}	[Apaf]	Apoptosis Activating Factor (Apaf-1), substrate of CyC, inactive form	1×10^5	1×10^3	1×10^7
x_{50}	[Apaf:CyC]	complex containing Apaf and cytochrome c	0	0	0
x_{51}	[Apaf*]	active form of Apaf-1	0	0	0
x_{52}	[C9]	procaspase-9	1×10^5	1×10^3	1×10^7
x_{53}	[Apop]	the apoptosome, which is the complex Apaf*:C9	0	0	0
x_{54}	[Apop:C3]	complex containing the apoptosome and procaspase-3	0	0	0
x_{55}	[Smac]	Smac/Diablo in the CC	0	0	0
x_{56}	[Apop:XIAP]	complex containing the apoptosome and XIAP	0	0	0
x_{57}	[Smac:XIAP]	complex containing Smac and XIAP	0	0	0
x_{58}	[$C3^*_{Ub}$]	C3* ubiquinated and targeted for degradation, assumed inactive	0	0	0

8.3 Specification of the EGFR-model

```
function responseCurves = runEGFRModel(kMat,yMat,responseVar)
  time = [0:1:200];
  responseCurves =  zeros(length(time),0);

  kf = 1.0;
  kr = 0.01;
  kcie = 0.01;
  kcde = 1.0;

  for (yset = 1:numel(yMat(:,1)) )

       y0=yMat(yset,1);yMat(yset,2);yMat(yset,3);yMat(yset,4);yMat(yset,
       5);yMat(yset,6);yMat(yset,7)];
       options = odeset('AbsTol', 1e-15,'RelTol', 1e-6);
       [t,y] = ode15s(@egfrModel,time,y0,options);
       responseCurve = cat(2,responseCurves,y(:,responseVar));
                    % concatenate column-wise;
       y(:,responseVar)
  end

  function dydt = egfrModel(t,y)
    % EGF
  %  x1 =-kf*EGF*R+kr*R_EGF;
    x1 =-kf*y(1)*y(2)+kr*y(3);

  % R
  %  x2 =-kf*EGF*R+kr*R_EGF;
    x2 =-kf*y(1)*y(2)+kr*y(3);

  % R_EGF
  %  x3 =+kf*EGF*R-kr*R_EGF-kcie*R_EGF*CIE-kcde*R_EGF*CDE;
    x3 =+kf*y(1)*y(2)-kr*y(3)-kcie*y(3)*y(4)-kcde*y(3)*y(6);

  % CIE
  %  x4 =-kcie*R_EGF*CIE;
    x4 =-kcie*y(3)*y(4);

  % R_CIE
  %  x5 =+kcie*R_EGF*CIE;
    x5 =+kcie*y(3)*y(4);

  % CDE
  %  x6 =-kcde*R_EGF*CDE;
    x6 =-kcde*y(3)*y(6);

  % R_CDE
  %  x7 =+kcde*R_EGF*CDE;
    x7 =+kcde*y(3)*y(6);
  dydt = [x1;x2;x3;x4;x5;x6;x7];
  end
end
```

8.4 Caspase-model specification

```
function responseCurves = runCaspaseModel(yMat,responseVar)
  % time span - ending at 6 hours
  % time starts at 0 and ends at 21600 seconds,
  % hence one step corresponds to 43.2866 seconds

  t = linspace(0,6*60*60,500);
  responseCurves =  zeros(length(t),0);

  for (yset = 1:numel(yMat(:,1)) )
    y0                                                                    =
[yMat(yset,1);yMat(yset,2);yMat(yset,3);yMat(yset,4);yMat(yset,5);yMat(y
set,6);yMat(yset,7);yMat(yset,8)];
    options = odeset('AbsTol', 1e-15,'RelTol', 1e-6);
    [t,y] = ode15s(@casp3modelodes,t,y0,options);
    responseCurves = cat(2,responseCurves,y(:,responseVar));  % concate-
nate column-wise; y(:,responseVar)
  end

    function dydt = casp3modelodes(t,y)

    %network species
    c8=y(1);    %casp-8
    c8a=y(2);   %casp-8*
    c3=y(3);    %casp-3
    c3a=y(4);   %casp-3*
    c8c3a=y(5); %casp-8:casp-3*
    c8ac3=y(6); %casp-8*:casp-3
    xiap=y(7);  %XIAP
    c3ax=y(8);  %casp-3*:XIAP

    %parameters
    kd2=.008;
    kd4=1e-3;
    k1=2.667e-9;
    kd1=.01;
    k3=6.8e-8;
    kd3=.05;
    k5=7e-5;
    kd5=1e7/6e11;
    kd6=1/6000;

    %odes
    dc8 = -k1*c3a*c8 + kd1*c8c3a;
    dc8a = kd2*c8c3a -k3*c8a*c3 +kd3*c8ac3 +kd4*c8ac3;
    dc3 = -k3*c8a*c3 +kd3*c8ac3;
    dc3a= kd4*c8ac3-k1*c3a*c8+kd1*c8c3a-k5*xiap*c3a+kd5*c3ax +kd2*c8c3a;
    dc8c3a = -kd2*c8c3a + k1*c3a*c8 - kd1*c8c3a;
    dc8ac3 = -kd4*c8ac3 +k3*c8a*c3 -kd3*c8ac3;
    dxiap = -k5*xiap*c3a +kd5*c3ax +kd6*c3ax;
    dc3ax = k5*xiap*c3a -kd5*c3ax -kd6*c3ax;

    dydt=[dc8;dc8a;dc3;dc3a;dc8c3a;dc8ac3;dxiap;dc3ax];
    end
end
```

8.5 Specification of the CD95/Fas-Model

```
function responseCurves = runNeumannModel(yMat,responseVar)
  t = [0:1:360];
  responseCurves =  zeros(length(t),0);
  for (yset = 1:numel(yMat(:,1)) )
    y0=yMat(yset,1);yMat(yset,2);yMat(yset,3);yMat(yset,4);yMat(yset,5);
    yMat(yset,6);yMat(yset,7);yMat(yset,8);yMat(yset,9);yMat(yset,10);yM
    at(yset,11);yMat(yset,12);yMat(yset,13);yMat(yset,14);yMat(yset,15);
    yMat(yset,16);yMat(yset,17);yMat(yset,18);yMat(yset,19);yMat(yset,20
    );yMat(yset,21);yMat(yset,22);yMat(yset,23)];
   options = odeset('AbsTol', 1e-15,'RelTol', 1e-6);
    [t,y] = ode15s(@neumann_model,t,y0,options);
    responseCurves = cat(2,responseCurves,y(:,responseVar));
                     % concatenate column-wise;
     y(:,responseVar)
  end

   function dydt = neumann_model(t,y)
      k1 = 1.0e+00;
      k2 = 1.28E-04;
      k3 = 6.69E-01;
      k4 = 1.0e-05;
      k5 = 5.95E-04;
      k6 = 1.0e+00;
      k7 = 8.88E-01;
      k8 = 8.04E-04;
      k9 = 2.25E-03;
      k10 = 1.21E-01;
      k11 = 2.89E-02;
      k12 = 1.50E-01;
      k13 = 7.20E-04;
      k14 = 3.59E-01;
      k15 = 3.68E+00;
      k16 = 2.23E-02;
      k17 = 6.42E-03;

    % L
    %  x1 =-k1*L*RF;
      x1 =-k1*y(1)*y(2);

    % RF
    %  x2 =-k1*L*RF;
      x2 =-k1*y(1)*y(2);

    % L:RF
    %  x3 =+k1*L*RF-k2*L:RF*C8-k3*L:RF*FL-k4*L:RF*FS;
      x3 =+k1*y(1)*y(2)-k2*y(3)*y(4)-k3*y(3)*y(6)-k4*y(3)*y(8);

    % C8
   %x4=-k2*L:RF*C8-k5*L:RF:C8*C8-k5*L:RF:FL*C8-k5*L:RF:FS*C8k10*C8*C3*;
   x4=-k2*y(3)*y(4)-k5*y(5)*y(4)-k5*y(7)*y(4)k5*y(9)*y(4)k10*y(4)*y(18);

    % L:RF:C8
    %  x5 =+k2*L:RF*C8-k5*L:RF:C8*C8-k6*L:RF:C8*FL-k7*L:RF:C8*FS;
      x5 =+k2*y(3)*y(4)-k5*y(5)*y(4)-k6*y(5)*y(6)-k7*y(5)*y(8);

    % FL
```

```
%  x6 =-k3*L:RF*FL-k6*L:RF:C8*FL-k6*L:RF:FL*FL-k6*L:RF:FS*FL;
  x6 =-k3*y(3)*y(6)-k6*y(5)*y(6)-k6*y(7)*y(6)-k6*y(9)*y(6);

% L:RF:FL
%  x7 =+k3*L:RF*FL-k5*L:RF:FL*C8-k6*L:RF:FL*FL-k7*L:RF:FL*FS;
  x7 =+k3*y(3)*y(6)-k5*y(7)*y(4)-k6*y(7)*y(6)-k7*y(7)*y(8);

% FS
%  x8 =-k4*L:RF*FS-k7*L:RF:C8*FS-k7*L:RF:FL*FS-k7*L:RF:FS*FS;
  x8 =-k4*y(3)*y(8)-k7*y(5)*y(8)-k7*y(7)*y(8)-k7*y(9)*y(8);

% L:RF:FS
%  x9 =+k4*L:RF*FS-k5*L:RF:FS*C8-k6*L:RF:FS*FL-k7*L:RF:FS*FS;
  x9 =+k4*y(3)*y(8)-k5*y(9)*y(4)-k6*y(9)*y(6)-k7*y(9)*y(8);

% p43/p41
%x10=+k5*L:RF:C8*C8+k5*L:RF:C8*C8-k8*p43/p41*p43/p41-
                                %k8*p43/p41*p43/p41+k10*C8*C3*;
  x10=+k5*y(5)*y(4)+k5*y(5)*y(4)-k8*y(10)*y(10)
                                   k8*y(10)*y(10)+k10*y(4)*y(18);

% p43-FLIP
%  x11 =+k6*L:RF:C8*FL+k5*L:RF:FL*C8-k13*p43-FLIP*IKK;
  x11 =+k6*y(5)*y(6)+k5*y(7)*y(4)-k13*y(11)*y(19);

% L:RF:C8:FS
%  x12 =+k7*L:RF:C8*FS+k5*L:RF:FS*C8;
  x12 =+k7*y(5)*y(8)+k5*y(9)*y(4);

% L:RF:FL:FL
%  x13 =+k6*L:RF:FL*FL;
  x13 =+k6*y(7)*y(6);

% L:RF:FL:FS
%  x14 =+k7*L:RF:FL*FS+k6*L:RF:FS*FL;
  x14 =+k7*y(7)*y(8)+k6*y(9)*y(6);

% L:RF:FS:FS
%  x15 =+k7*L:RF:FS*FS;
  x15 =+k7*y(9)*y(8);

% C8*
%  x16 =+k8*p43/p41*p43/p41-k9*C3*C8*+k9*C3*C8* +0 -k11*y(16);
  x16 =+k8*y(10)*y(10)-k9*y(17)*y(16)+k9*y(17)*y(16) +0 -k11*y(16);

% C3
%  x17 =-k9*C3*C8*;
  x17 =-k9*y(17)*y(16);

% C3*
%  x18 =+k9*C3*C8*-k10*C8*C3*+k10*C8*C3* +0 -k12*y(18);
  x18 =+k9*y(17)*y(16)-k10*y(4)*y(18)+k10*y(4)*y(18) +0 -k12*y(18);

% IKK
%  x19 =-k13*p43-FLIP*IKK;
  x19 =-k13*y(11)*y(19);

% p43-FLIP:IKK*
%  x20 =+k13*p43-FLIP*IKK-k16*y(20);
  x20 =+k13*y(11)*y(19)-k16*y(20);
```

```
    % NF-kB:IkB
    %  x21 =-k14*NF-kB:IkB*p43-FLIP:IKK*;
      x21 =-k14*y(21)*y(20);

    % NF-kB:IkB:P
    %  x22 =+k14*NF-kB:IkB*p43-FLIP:IKK*-k15*NF-kB:IkB:P;
      x22 =+k14*y(21)*y(20)-k15*y(22);

    % NF-kB*
    %  x23 =+k15*NF-kB:IkB:P +0 -k17*y(23);
      x23 =+k15*y(22) +0 -k17*y(23);
    dydt                                                                =
[x1;x2;x3;x4;x5;x6;x7;x8;x9;x10;x11;x12;x13;x14;x15;x16;x17;x18;x19;x20;
x21;x22;x23];
   end
end
```

8.6 Specification of EARM v1.0

```
function responseCurves = runEARM(yMat,responseVar)
  % time span - ending at 6 hours
  % Time settings according to Fig_4B of Albeck et al.
  tf=6*3600; % seconds
  samp_freq=tf/60; % accurate to min
  dt=(tf-0)/samp_freq;
  tt=0:dt:tf;
  options=odeset('AbsTol',1E-10,'RelTol',1E-8,'maxstep',dt);
  responseCurves =  zeros(length(tt),0);

  for (yset = 1:numel(yMat(:,1)) )
    y0 = yMat(yset,:)';
    [t,y] = ode15s(@rhs,tt,y0,options);
    responseCurves = cat(2,responseCurves,y(:,responseVar));  % concate-
nate column-wise; y(:,responseVar)
  end

  function xp=rhs(t,x)
    % ODE with dimensions

    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    %                    Rate constants                              %
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

    transloc=.01;% rate of translocation between the cytosolic and mito-
chondrial compartments
    v=.07; % mitochondria compartment volume/cell volume

    % L + pR <--> L:pR --> R*
    k(1)=4E-7; k_(1)=1E-3; kc(1)=1E-5;

    % flip + DISC <-->  flip:DISC
    k(2)=1E-6; k_(2)=1E-3;

    % pC8 + DISC <--> DISC:pC8 --> C8 + DISC
    k(3)=1E-6; k_(3)=1E-3; kc(3)=1;

    % C8 + BAR <--> BAR:C8
    k(4)=1E-6; k_(4)=1E-3;

    % pC3 + C8 <--> pC3:C8 --> C3 + C8
    k(5)=1E-7; k_(5)=1E-3; kc(5)=1;

    % pC6 + C3 <--> pC6:C3 --> C6 + C3
    k(6)=1E-6; k_(6)=1E-3; kc(6)=1;

    % pC8 + C6 <--> pC8:C6 --> C8 + C6
    k(7)=3E-8; k_(7)=1E-3; kc(7)=1;

    % XIAP + C3 <--> XIAP:C3 --> XIAP + C3_U
    k(8)=2E-6; k_(8)=1E-3;  kc(8)=.1;

    % PARP + C3 <--> PARP:C3 --> CPARP + C3
    k(9)=1E-6; k_(9)=1E-2; kc(9)=1;

    % Bid + C8 <--> Bid:C8 --> tBid + C8
    k(10)=1E-7; k_(10)=1E-3; kc(10)=1;
```

```
% tBid + Bcl2c <-->  tBid:Bcl2c
k(11)=1E-6; k_(11)=1E-3;

% Bax + tBid <--> Bax:tBid --> aBax + tBid
k(12)=1E-7; k_(12)=1E-3; kc(12)=1;

% aBax <-->  MBax
k(13)=transloc; k_(13)=transloc;

% MBax + Bcl2 <-->  MBax:Bcl2
k(14)=1E-6; k_(14)=1E-3;

% MBax + MBax <-->  MBax:MBax == Bax2
k(15)=1E-6; k_(15)=1E-3;

% Bax2 + Bcl2 <-->  MBax2:Bcl2
k(16)=1E-6; k_(16)=1E-3;

% Bax2 + Bax2 <-->  Bax2:Bax2 == Bax4
k(17)=1E-6; k_(17)=1E-3;

% Bax4 + Bcl2 <-->  MBax4:Bcl2
k(18)=1E-6; k_(18)=1E-3;

% Bax4 + Mit0 <-->  Bax4:Mito -->  AMito
k(19)=1E-6; k_(19)=1E-3; kc(19)=1;

% AMit0 + mCtoC <-->  AMito:mCytoC --> AMito + ACytoC
k(20)=2E-6; k_(20)=1E-3; kc(20)=10;

% AMit0 + mSMac <-->  AMito:mSmac --> AMito + ASMAC
k(21)=2E-6; k_(21)=1E-3; kc(21)=10;

% ACytoC <-->  cCytoC
k(22)=transloc; k_(22)=transloc;

% Apaf + cCytoC <-->  Apaf:cCytoC
k(23)=5E-7; k_(23)=1E-3; kc(23)=1;

% Apaf:cCytoC + Procasp9 <-->  Apoptosome
k(24)=5E-8; k_(24)=1E-3;

% Apop + pCasp3 <-->  Apop:cCasp3 --> Apop + Casp3
k(25)=5E-9; k_(25)=1E-3; kc(25)=1;

% ASmac <-->  cSmac
k(26)=transloc; k_(26)=transloc;

% Apop + XIAP <-->  Apop:XIAP
k(27)=2E-6; k_(27)=1E-3;

% cSmac + XIAP <-->  cSmac:XIAP
k(28)=7E-6; k_(28)=1E-3;


%------------------------------------------------------------

xp=double(x);
```

```
xp(1) = -k(1)*x(1)*x(2) +k_(1)*x(3) ; % Ligand

xp(2) = -k(1)*x(1)*x(2) +k_(1)*x(3) ; % R

xp(3) =  k(1)*x(1)*x(2) -k_(1)*x(3) -kc(1)*x(3) ; % L:R complex

xp(4) = kc(1)*x(3) +...
    -k(2)*x(4)*x(5) +k_(2)*x(6) +...
    -k(3)*x(4)*x(7) +k_(3)*x(8) +kc(3)*x(8); % R*

xp(5)= -k(2)*x(4)*x(5) +k_(2)*x(6) ; % flip

xp(6)=  k(2)*x(4)*x(5) -k_(2)*x(6) ; % flip:R*

xp(7) = -k(3)*x(4)*x(7) +k_(3)*x(8) +...
    -k(7)*x(7)*x(17) +k_(7)*x(18) ; % pC8

xp(8) =  k(3)*x(4)*x(7) -k_(3)*x(8) -kc(3)*x(8) ; % R*:pC8

xp(9) =  kc(3)*x(8) +...
    -k(4)*x(9)*x(10) +k_(4)*x(11) +...
    -k(5)*x(9)*x(12) +k_(5)*x(13) +kc(5)*x(13) +...
    +kc(7)*x(18) +...
    -k(10)*x(9)*x(24) +k_(10)*x(25) +kc(10)*x(25) ; % C8

xp(10) = -k(4)*x(9)*x(10) +k_(4)*x(11) ; % Bar

xp(11) =  k(4)*x(9)*x(10) -k_(4)*x(11) ; % Bar:C8

xp(12)= -k(5)*x(9)*x(12) +k_(5)*x(13) +...
    -k(25)*x(12)*x(53) +k_(25)*x(54) ; % pC3

xp(13)=  k(5)*x(9)*x(12) -k_(5)*x(13) -kc(5)*x(13) ; % C8:pC3

xp(14)=  kc(5)*x(13) +...
    -k(6)*x(14)*x(15) +k_(6)*x(16) +kc(6)*x(16) +...
    -k(8)*x(14)*x(19) +k_(8)*x(20) +...
    -k(9)*x(14)*x(21) +k_(9)*x(22) +kc(9)*x(22) +...
    +kc(25)*x(54) ; % C3

xp(15)= -k(6)*x(14)*x(15) +k_(6)*x(16) ; % pC6

xp(16)=  k(6)*x(14)*x(15) -k_(6)*x(16) -kc(6)*x(16) ; % C3:pC6

xp(17)=  kc(6)*x(16) +...
    -k(7)*x(7)*x(17) +k_(7)*x(18) +kc(7)*x(18) ; % C6

xp(18)=  k(7)*x(7)*x(17) -k_(7)*x(18) -kc(7)*x(18) ; % C6:pC8

xp(19)= -k(8)*x(14)*x(19) +k_(8)*x(20) +kc(8)*x(20) +...
    -k(27)*x(19)*x(53) +k_(27)*x(56) +...
    -k(28)*x(19)*x(55) +k_(28)*x(57) ; % XIAP

xp(20)=  k(8)*x(14)*x(19) -k_(8)*x(20) -kc(8)*x(20) ; % XIAP:C3

xp(21)= -k(9)*x(14)*x(21) +k_(9)*x(22) ; % PARP

xp(22)=  k(9)*x(14)*x(21) -k_(9)*x(22) -kc(9)*x(22) ; % C3:PARP

xp(23)= kc(9)*x(22) ; % CPARP
```

```
xp(24)= -k(10)*x(9)*x(24) +k_(10)*x(25) ; % Bid

xp(25)=  k(10)*x(9)*x(24) -k_(10)*x(25) -kc(10)*x(25) ; % C8:Bid

xp(26)=  kc(10)*x(25) +...
    -k(11)*x(26)*x(27) +k_(11)*x(28) +...
    -k(12)*x(26)*x(29) +k_(12)*x(30) + kc(12)*x(30); % tBid

xp(27)= -k(11)*x(26)*x(27) +k_(11)*x(28) ; % Bcl2c

xp(28)= +k(11)*x(26)*x(27) -k_(11)*x(28) ; % Bcl2c:tBid

xp(29)= -k(12)*x(26)*x(29) +k_(12)*x(30) ; % Bax

xp(30)=  k(12)*x(26)*x(29) -k_(12)*x(30) - kc(12)*x(30) ; % tBid:Bax

xp(31)=  kc(12)*x(30) +...
    -k(13)*x(31) + k_(13)*x(32) ; % Bax*

xp(32)=  k(13)*x(31) - k_(13)*x(32) +...
    -1/v*k(14)*x(32)*x(33) +k_(14)*x(34) +...
    -1/v*2*k(15)*x(32)^2 +2*k_(15)*x(35) ; % Baxm

xp(33)= -1/v*k(14)*x(32)*x(33) +k_(14)*x(34) +...
    -1/v*k(16)*x(33)*x(35) +k_(16)*x(36) +...
    -1/v*k(18)*x(33)*x(37) +k_(18)*x(38) ; % Bcl2

xp(34)=  1/v*k(14)*x(32)*x(33) -k_(14)*x(34) ; % Baxm:Bcl2

xp(35)=  1/v*k(15)*x(32)^2 -k_(15)*x(35) +...
    -1/v*k(16)*x(33)*x(35) +k_(16)*x(36) +...
    -2/v*k(17)*x(35)^2 +2*k_(17)*x(37) ; % Bax2

xp(36)=  1/v*k(16)*x(33)*x(35) -k_(16)*x(36) ; % Bax2:Bcl2

xp(37)= 1/v*k(17)*x(35)^2 -k_(17)*x(37)+...
    -1/v*k(18)*x(33)*x(37) +k_(18)*x(38) +...
    -1/v*k(19)*x(39)*x(37) +k_(19)*x(40) ; % Bax4

xp(38)= 1/v*k(18)*x(33)*x(37) -k_(18)*x(38) ; % Bax4:Bcl2

xp(39)= -1/v*k(19)*x(39)*x(37) +k_(19)*x(40); % M

xp(40)=  1/v*k(19)*x(39)*x(37) -k_(19)*x(40) -kc(19)*x(40); % Bax4:M

xp(41)=  kc(19)*x(40) +...
    -1/v*k(20)*x(41)*x(42) +k_(20)*x(43) +kc(20)*x(43) +...
    -1/v*k(21)*x(41)*x(45) +k_(21)*x(46) +kc(21)*x(46) ; % M*

xp(42)= -1/v*k(20)*x(41)*x(42) +k_(20)*x(43) ; % CytoCm

xp(43)=1/v*k(20)*x(41)*x(42)-k_(20)*x(43)-kc(20)*x(43) ; % M*:CytoCm

xp(44)=  kc(20)*x(43) +...
    -k(22)*x(44) +k_(22)*x(48) ; % CytoCr  % CytoC released

xp(45)= -1/v*k(21)*x(41)*x(45) +k_(21)*x(46) ; % Smacm

xp(46)= 1/v*k(21)*x(41)*x(45)-k_(21)*x(46)-kc(21)*x(46) ; % M*:Smacm
```

```
    xp(47)=  kc(21)*x(46) +...
        -k(26)*x(47) +k_(26)*x(55) ; % Smacr

    xp(48)=  k(22)*x(44) -k_(22)*x(48) +...
        -k(23)*x(48)*x(49) +k_(23)*x(50) +kc(23)*x(50) ; % CytoC in CC

    xp(49)= -k(23)*x(48)*x(49) +k_(23)*x(50) ; % Apaf
    xp(50)= k(23)*x(48)*x(49) -k_(23)*x(50) -kc(23)*x(50) ; % Apaf:CytoC

    xp(51)= kc(23)*x(50) +...
        -k(24)*x(51)*x(52) +k_(24)*x(53); % Apaf*

    xp(52)= -k(24)*x(51)*x(52) +k_(24)*x(53) ; % pC9

    xp(53)=  k(24)*x(51)*x(52) -k_(24)*x(53) +...
        -k(25)*x(12)*x(53) +k_(25)*x(54) +kc(25)*x(54) +...
        -k(27)*x(19)*x(53) +k_(27)*x(56) ; % Apop

    xp(54)=  k(25)*x(12)*x(53) -k_(25)*x(54) -kc(25)*x(54) ; % Apop:pC3

    xp(55)=  k(26)*x(47) -k_(26)*x(55) +...
        -k(28)*x(19)*x(55) +k_(28)*x(57) ; % Smac

    xp(56)=  k(27)*x(19)*x(53) -k_(27)*x(56) ; % Apop:XIAP

    xp(57)=  k(28)*x(19)*x(55) -k_(28)*x(57) ; % Smac:XIap

    xp(58)=kc(8)*x(20); % C3_Ub

dydt=[xp(1);xp(2);xp(3);xp(4);xp(5);xp(6);xp(7);xp(8);xp(9);xp(10);xp(11
);xp(12);xp(13);xp(14);xp(15);xp(16);xp(17);xp(18);xp(19);xp(20);xp(21);
xp(22);xp(23);xp(24);xp(25);xp(26);xp(27);xp(28);xp(29);xp(30);xp(31);xp
(32);xp(33);xp(34);xp(35);xp(36);xp(37);xp(38);xp(39);xp(40);xp(41);xp(4
2);xp(43);xp(44);xp(45);xp(46);xp(47);xp(48);xp(49);xp(50);xp(51);xp(52)
;xp(53);xp(54);xp(55);xp(56);xp(57);xp(58)];

  end
end
```

8.7 Trees at best pruning level

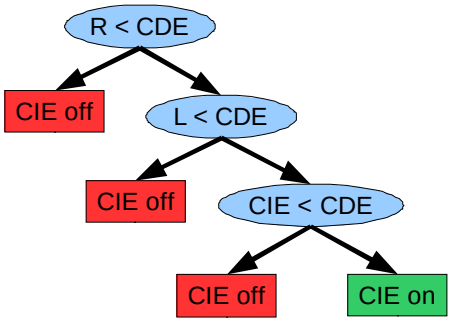

Figure: 8.1: Tree of best level as result of model analysis of the EGFR-model

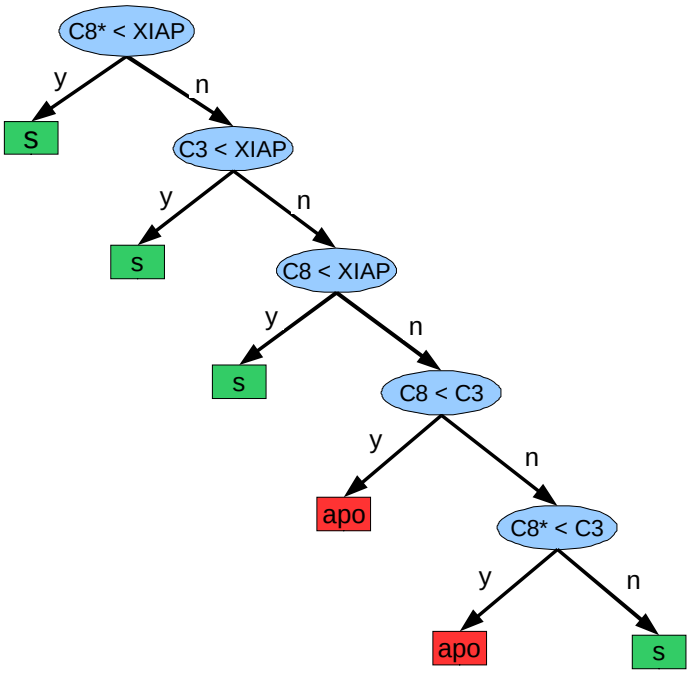

Figure: 8.2: Tree of best level as result of model analysis of the Caspase-model

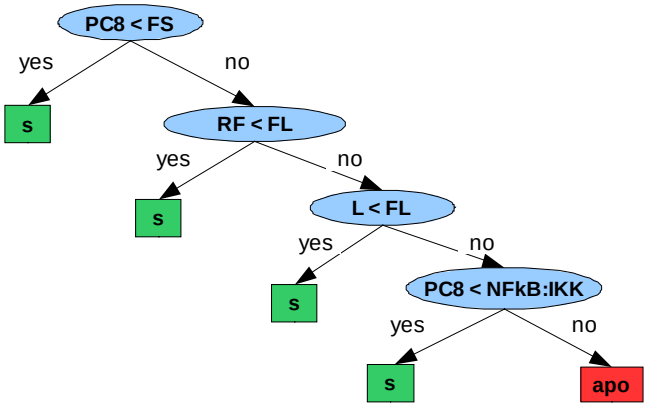

Figure: 8.3: Tree of best level as result of model analysis of the CD95/Fas-model

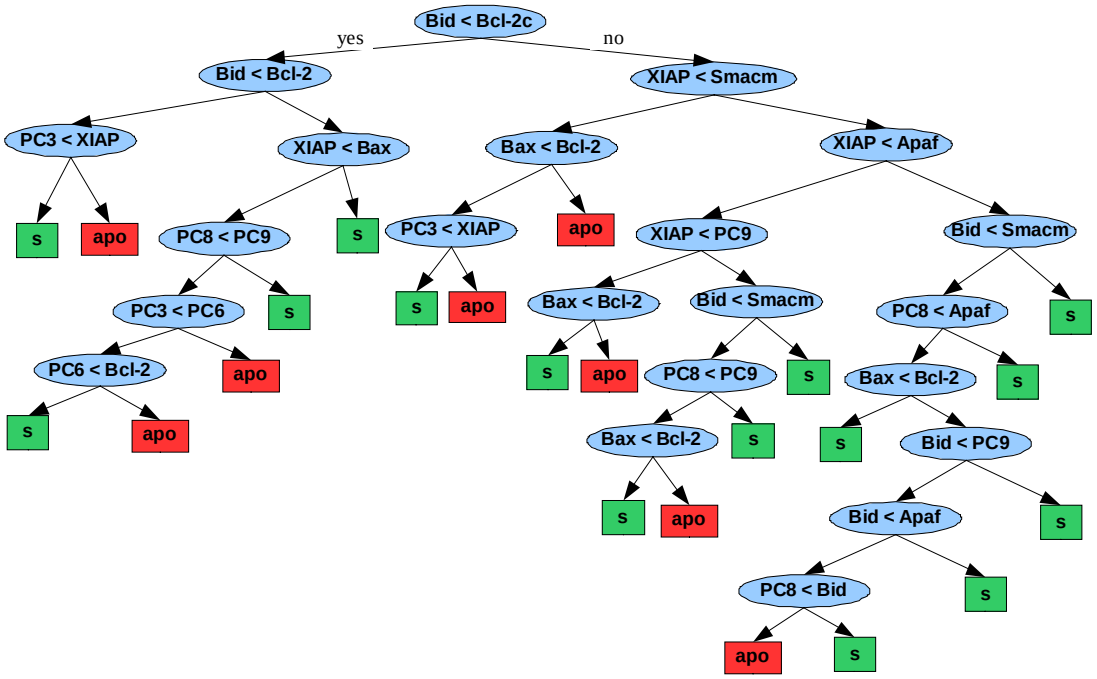

Figure: 8.4: Tree of best level as result of model analysis of the EARM v1.0-model

References

[1] J. M. Adams and S. Cory. The bcl-2 apoptotic switch in cancer development and therapy. *Oncogene*, 26(9):1324–1337, Feb 2007.

[2] John G Albeck, John M Burke, Bree B Aldridge, Mingsheng Zhang, Douglas A Lauffenburger, and Peter K Sorger. Quantitative analysis of pathways controlling extrinsic apoptosis in single cells. *Mol Cell*, 30(1):11–25, Apr 2008.

[3] John G Albeck, John M Burke, Sabrina L Spencer, Douglas A Lauffenburger, and Peter K Sorger. Modeling a snap-action, variable-delay switch controlling extrinsic cell death. *PLoS Biol*, 6(12):2831–2852, Dec 2008.

[4] B. B. Aldridge, G. Haller, P. K. Sorger, and D. A. Lauffenburger. Direct lyapunov exponent analysis enables parametric study of transient signalling governing cell behaviour. *Syst Biol (Stevenage)*, 153(6):425–432, Nov 2006.

[5] Bree B Aldridge, John M Burke, Douglas A Lauffenburger, and Peter K Sorger. Physicochemical modelling of cell signalling pathways. *Nat Cell Biol*, 8(11):1195–1203, Nov 2006.

[6] David Angeli, James E Ferrell, and Eduardo D 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, Feb 2004.

[7] P. W. Atkins and J. De Paula. *Kurzlehrbuch der physikalischen Chemie*. Wiley-VCH, Weinheim, 4. auflage edition, 2008.

[8] E. Z. Bagci, Y. Vodovotz, T. R. Billiar, G. B. Ermentrout, and I. Bahar. Bistability in apoptosis: roles of bax, bcl-2, and mitochondrial permeability transition pores. *Biophys J*, 90(5):1546–1559, Mar 2006.

[9] Elife Z Bagci, Yoram Vodovotz, Timothy R Billiar, Bard Ermentrout, and Ivet Bahar. Computational insights on the competing effects of nitric oxide in regulating apoptosis. *PLoS One*, 3(5):e2249, 2008.

[10] Q. Bao and Y. Shi. Apoptosome: a platform for the activation of initiator caspases. *Cell Death Differ*, 14(1):56–65, Jan 2007.

[11] M. Bentele, I. Lavrik, M. Ulrich, S. Stösser, D. W. Heermann, H. Kalthoff, P. H. Krammer, and R. Eils. Mathematical modeling reveals threshold mechanism in cd95-induced apoptosis. *J Cell Biol*, 166(6):839–851, Sep 2004.

[12] Kelly M Boatright, Martin Renatus, Fiona L Scott, Sabina Sperandio, Hwain Shin, Irene M Pedersen, Jean Ehrland Ricci, Wade A Edris, Daniel P Sutherlin, Douglas R Green, and Guy S Salvesen. A unified model for apical caspase activation. *Molecular cell*, 11(2):529–41, 2003.

[13] Leo Breiman, Jerome Friedman, R Olshen, and Charles J. Stone. *Classification and Regression Trees*. Boca Raton,FL: CRC Press, 1984.

[14] Alexander Bürkle. *Poly(ADP-Ribosyl)ation*. Landes Bioscience, 2006.

[15] Frank J Bruggeman and Hans V Westerhoff. The nature of systems biology. *Trends Microbiol*, 15(1):45–50, Jan 2007.

[16] Nicolas E Buchler and Frederick R Cross. Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Mol Syst Biol*, 5:272, 2009.

[17] Ralph C Budd, Wen-Chen Yeh, and Jürg Tschopp. cflip regulation of lymphocyte activation and development. *Nat Rev Immunol*, 6(3):196–204, Mar 2006.

[18] B. A. Callus and D. L. Vaux. Caspase inhibitors: viral, cellular and chemical. *Cell Death Differ*, 14(1):73–78, Jan 2007.

[19] Victoria Cepeda, Miguel A Fuertes, Josefina Castilla, Carlos Alonso, Celia Quevedo, Manual Soto, and José M Pérez. Poly(adp-ribose) polymerase-1 (parp-1) inhibitors in cancer chemotherapy. *Recent Pat Anticancer Drug Discov*, 1(1):39–53, Jan 2006.

[20] David W Chang, Zheng Xing, Yi Pan, Alicia Algeciras-Schimnich, Bryan C Barnhart, Shoshanit Yaish-Ohad, Marcus E Peter, and Xiaolu Yang. c-flip(l) is a dual function regulator for caspase-8 activation and cd95-mediated apoptosis. *EMBO J*, 21(14):3704–3714, Jul 2002.

[21] E. H. Cheng, M. C. Wei, S. Weiler, R. A. Flavell, T. W. Mak, T. Lindsten, and S. J. Korsmeyer. Bcl-2, bcl-x(l) sequester bh3 domain-only molecules preventing bax- and bak-mediated mitochondrial apoptosis. *Mol Cell*, 8(3):705–711, Sep 2001.

[22] J. E. Chipuk, L. Bouchier-Hayes, and D. R. Green. Mitochondrial outer membrane permeabilization during apoptosis: the innocent bystander scenario. *Cell Death Differ*, 13(8):1396–1402, Aug 2006.

[23] Jerry E Chipuk and Douglas R Green. How do bcl-2 proteins induce mitochondrial outer membrane permeabilization? *Trends Cell Biol*, 18(4):157–164, Apr 2008.

[24] Nika N Danial and Stanley J Korsmeyer. Cell death: critical control points. *Cell*, 116(2):205–219, Jan 2004.

[25] A. O. de Graaf, T. de Witte, and J. H. Jansen. Inhibitor of apoptosis proteins: new therapeutic targets in hematological cancer? *Leukemia*, 18(11):1751–1759, Nov 2004.

[26] Hidde de Jong. Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol*, 9(1):67–103, 2002.

[27] A de Thonel and J E Eriksson. Regulation of death receptors–relevance in cancer therapies. *Toxicology and applied pharmacology*, 207(2 Suppl):123–32, 2005.

[28] Jean-Bernard Denault and Guy S Salvesen. Caspases. *Curr Protoc Protein Sci*, Chapter 21:Unit 21.8, Feb 2002.

[29] J L DeRisi, V R Iyer, and P O Brown. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science (New York, N.Y.)*, 278(5338):680–6, 1997.

[30] Jacqueline M Dresch, Xiaozhou Liu, David N Arnosti, and Ahmet Ay. Thermodynamic modeling of transcription: sensitivity analysis differentiates biological mechanism from mathematical model-induced effects. *BMC Syst Biol*, 4:142, 2010.

[31] C. Du, M. Fang, Y. Li, L. Li, and X. Wang. Smac, a mitochondrial protein that promotes cytochrome c-dependent caspase activation by eliminating iap inhibition. *Cell*, 102(1):33–42, Jul 2000. Identification of Smac.

[32] Thomas Eissing, Holger Conzelmann, Ernst D Gilles, Frank Allgöwer, Eric Bullinger, and Peter Scheurich. Bistability analyses of a caspase activation model for receptor-induced apoptosis. *J Biol Chem*, 279(35):36892–36897, Aug 2004.

[33] S. W. Fesik. Insights into programmed cell death through structural biology. *Cell*, 103(2):273–282, Oct 2000.

[34] Matthias Fischer, André Oberthuer, Benedikt Brors, Yvonne Kahlert, Matthias Skowron, Harald Voth, Patrick Warnat, Karen Ernestus, Barbara Hero, and Frank Berthold. Differential expression of neuronal genes defines subtypes of disseminated neuroblastoma with favorable and unfavorable outcome. *Clin Cancer Res*, 12(17):5118–5128, Sep 2006.

[35] Jasmin Fisher and Thomas A Henzinger. Executable cell biology. *Nat Biotechnol*, 25(11):1239–1249, Nov 2007.

[36] Nicolai Fricker, Joel Beaudouin, Petra Richter, Roland Eils, Peter H Krammer, and Inna N Lavrik. Model-based dissection of cd95 signaling dynamics reveals both a pro- and antiapoptotic role of c-flipl. *J Cell Biol*, 190(3):377–389, Aug 2010.

[37] N. Friedman, M. Linial, I. Nachman, and D. Pe'er. Using bayesian networks to analyze expression data. *J Comput Biol*, 7(3-4):601–620, 2000.

[38] M. Giam, D. C S Huang, and P. Bouillet. Bh3-only proteins and their roles in programmed cell death. *Oncogene*, 27 Suppl 1:S128–S136, Dec 2008.

[39] Jelle J Goeman and Jan Oosting. Testing association of a pathway with a clinical variable. 2007.

[40] Jelle J Goeman, Sara A van de Geer, Floor de Kort, and Hans C van Houwelingen. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics*, 20(1):93–99, Jan 2004.

[41] S. P. Gygi, Y. Rochon, B. R. Franza, and R. Aebersold. Correlation between protein and mrna abundance in yeast. *Mol Cell Biol*, 19(3):1720–1730, Mar 1999.

[42] D. Hanahan and R. A. Weinberg. The hallmarks of cancer. *Cell*, 100(1):57–70, Jan 2000.

[43] A. L. Harris, S. Nicholson, J. R. Sainsbury, J. Farndon, and C. Wright. Epidermal growth factor receptors in breast cancer: association with early relapse and death, poor response to hormones and interactions with neu. *J Steroid Biochem*, 34(1-6):123–131, 1989.

[44] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.

[45] Jorrit J Hornberg, Bernd Binder, Frank J Bruggeman, Birgit Schoeberl, Reinhart Heinrich, and Hans V Westerhoff. Control of mapk signalling: from complexity to what really matters. *Oncogene*, 24(36):5533–5542, Aug 2005.

[46] Wolfgang Huber, Anja von Heydebreck, Holger Sültmann, Annemarie Poustka, and Martin Vingron. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, 18 Suppl 1:S96–104, 2002.

[47] Alan Julian Izenman. *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. Springer, New York, 2008.

[48] Prakash Jagtap and Csaba Szabó. Poly(adp-ribose) polymerase and the therapeutic effects of its inhibitors. *Nat Rev Drug Discov*, 4(5):421–440, May 2005.

[49] Ian B Jeffery, Desmond G Higgins, and Aedín C Culhane. Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. *BMC Bioinformatics*, 7:359, 2006.

[50] J F Kerr, A H Wyllie, and A R Currie. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *British journal of cancer*, 26(4):239–57, 1972.

[51] F. C. Kischkel, S. Hellbardt, I. Behrmann, M. Germer, M. Pawlita, P. H. Krammer, and M. E. Peter. Cytotoxicity-dependent apo-1 (fas/cd95)-associated proteins form a death-inducing signaling complex (disc) with the receptor. *EMBO J*, 14(22):5579–5588, Nov 1995.

[52] Hiroaki Kitano. Systems biology: a brief overview. *Science*, 295(5560):1662–1664, Mar 2002.

[53] Steffen Klamt, Julio Saez-Rodriguez, and Ernst D Gilles. Structural and functional analysis of cellular networks with cellnetanalyzer. *BMC Syst Biol*, 1:2, 2007.

[54] S. Knudsen. *Cancer Diagnostics with DNA Microarrays*. John Wiley & Sons, Hoboken, New Jersey, 1. auflage edition, 2006.

[55] David W Koh, Ted M Dawson, and Valina L Dawson. Mediation of cell death by poly(ADP-ribose) polymerase-1. *Pharmacological research : the official journal of the Italian Pharmacological Society*, 52(1):5–14, 2005.

[56] A. Krueger, I. Schmitz, S. Baumann, P. H. Krammer, and S. Kirchhoff. Cellular flice-inhibitory protein splice variants inhibit different steps of caspase-8 activation at the cd95 death-inducing signaling complex. *J Biol Chem*, 276(23):20633–20640, Jun 2001.

[57] Sharad Kumar and Dimitrios Cakouros. Transcriptional control of the core cell-death machinery. *Trends in biochemical sciences*, 29(4):193–9, 2004.

[58] M Lamkanfi, N Festjens, W Declercq, T Vanden Berghe, and P Vandenabeele. Caspases in cell survival, proliferation and differentiation. *Cell death and differentiation*, 14(1):44–55, 2007.

[59] Inna Lavrik, Alexander Golks, and Peter H Krammer. Death receptor signaling. *J Cell Sci*, 118(Pt 2):265–267, Jan 2005.

[60] Inna N Lavrik, Alexander Golks, Dagmar Riess, Martin Bentele, Roland Eils, and Peter H Krammer. Analysis of cd95 threshold signaling: triggering of cd95 (fas/apo-1) at low concentrations primarily results in survival signaling. *J Biol Chem*, 282(18):13664–13671, May 2007.

[61] Alfons Lawen. Apoptosis-an introduction. *Bioessays*, 25(9):888–896, Sep 2003.

[62] Brian Leber, Jialing Lin, and David W Andrews. Embedded together: the life and death consequences of interaction of the bcl-2 family with membranes. *Apoptosis*, 12(5):897–911, May 2007.

[63] Stefan Legewie, Nils Blüthgen, and Hanspeter Herzel. Mathematical modeling identifies inhibitors of apoptosis as mediators of positive feedback and bistability. *PLoS Comput Biol*, 2(9):e120, Sep 2006.

[64] H Li, H Zhu, C J Xu, and J Yuan. Cleavage of BID by caspase 8 mediates the mitochondrial damage in the Fas pathway of apoptosis. *Cell*, 94(4):491–501, 1998.

[65] X. Luo, I. Budihardjo, H. Zou, C. Slaughter, and X. Wang. Bid, a bcl2 interacting protein, mediates cytochrome c release from mitochondria in response to activation of cell surface death receptors. *Cell*, 94(4):481–490, Aug 1998.

[66] A U Lüthi and S J Martin. The CASBAH: a searchable database of caspase substrates. *Cell death and differentiation*, 14(4):641–50, 2007.

[67] Alborz Mahdavi, Ryan E Davey, Patrick Bhola, Ting Yin, and Peter W Zandstra. Sensitivity analysis of intracellular signaling pathway kinetics predicts targets for stem cell fate control. *PLoS Comput Biol*, 3(7):e130, Jul 2007.

[68] Oded Maimon and Lior Rokach. *Data mining with decision trees, theory and applications*. World Scientific Pub Co, 2008.

[69] Pascal Meier and Karen H Vousden. Lucifer's labyrinth–ten years of path finding in cell death. *Molecular cell*, 28(5):746–54, 2007.

[70] Thomas Millat, Sree N Sreenath, Radina P Soebiyanto, Jayant Avva, Kwang-Hyun Cho, and Olaf Wolkenhauer. The role of dynamic stimulation pattern in the analysis of bistable intracellular networks. *Biosystems*, 92(3):270–281, Jun 2008.

[71] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, Oct 2002.

[72] M. Nachshon-Kedmi, S. Yannai, A. Haj, and F. A. Fares. Indole-3-carbinol and 3,3'-diindolylmethane induce apoptosis in human prostate cancer cells. *Food Chem Toxicol*, 41(6):745–752, Jun 2003.

[73] Leo Neumann, Carina Pforr, Joel Beaudouin, Alexander Pappa, Nicolai Fricker, Peter H Krammer, Inna N Lavrik, and Roland Eils. Dynamics within the cd95 death-inducing signaling complex decide life and death of cells. *Mol Syst Biol*, 6:352, 2010.

[74] Z. N. Oltvai, C. L. Milliman, and S. J. Korsmeyer. Bcl-2 heterodimerizes in vivo with a conserved homolog, bax, that accelerates programmed cell death. *Cell*, 74(4):609–619, Aug 1993.

[75] S. Pelengaris and M. Khan. *The molecular biology of cancer*. Blackwell Publishing Ltd, 2006.

[76] M. E. Peter, A. E. Heufelder, and M. O. Hengartner. Advances in apoptosis research. *Proc Natl Acad Sci U S A*, 94(24):12736–12737, Nov 1997.

[77] M. E. Peter and P. H. Krammer. The cd95(apo-1/fas) disc and beyond. *Cell Death Differ*, 10(1):26–35, Jan 2003.

[78] J. R. Quinlan. Induction of decision trees. *Mach. Learn.*, 1:81–106, March 1986.

[79] Érdi P. and Tóth J. *Mathematical Models of Chemical Reactions. Theory and Applications of Deterministic and Stochastic Models*. Manchester University Press, 1989.

[80] Markus Rehm, Heiko Dussmann, Reiner U Janicke, Jeremy M Tavare, Donat Kogel, and Jochen H M Prehn. Single-cell fluorescence resonance energy transfer analysis demonstrates that caspase activation during apoptosis is a rapid process. role of caspase-3. *J Biol Chem*, 277(27):24506–24514, Jul 2002.

[81] W. Roth, P. Kermer, M. Krajewska, K. Welsh, S. Davis, S. Krajewski, and J. C. Reed. Bifunctional apoptosis inhibitor (bar) protects neurons from diverse cell death pathways. *Cell Death Differ*, 10(10):1178–1187, Oct 2003.

[82] Dimitrios H Roukos, Samuel Murray, and Evangelos Briasoulis. Molecular genetic tools shape a roadmap towards a more accurate prognostic prediction and personalized management of cancer. *Cancer Biol Ther*, 6(3):308–312, Mar 2007.

[83] Andrea Saltelli, Marco Ratto, Stefano Tarantola, and Francesca Campolongo. Identification and review of sensitivity analysis methods. *Chem Rev*, 105(7):2811–2828, Jul 2005.

[84] Guy S Salvesen and Colin S Duckett. Iap proteins: blocking the road to death's door. *Nat Rev Mol Cell Biol*, 3(6):401–410, Jun 2002.

[85] Hannah Schmidt-Glenewinkel, Ivayla Vacheva, Daniela Hoeller, Ivan Dikic, and Roland Eils. An ultrasensitive sorting mechanism for egf receptor endocytosis. *BMC Syst Biol*, 2:32, 2008.

[86] Yigong Shi. Caspase activation, inhibition, and reactivation: a mechanistic view. *Protein Sci*, 13(8):1979–1987, Aug 2004.

[87] Richard M Siegel. Caspases at the crossroads of immune-cell life and death. *Nat Rev Immunol*, 6(4):308–317, Apr 2006.

[88] Alexander H Stegh, Bryan C Barnhart, Jorg Volkland, Alicia Algeciras-Schimnich, Ning Ke, John C Reed, and Marcus E Peter. Inactivation of caspase-8 on mitochondria of bcl-xl-expressing mcf7-fas cells: role for the bifunctional apoptosis regulator protein. *J Biol Chem*, 277(6):4351–4360, Feb 2002.

[89] A. Strasser, A. W. Harris, D. C. Huang, P. H. Krammer, and S. Cory. Bcl-2 and fas/apo-1 regulate distinct pathways to lymphocyte apoptosis. *EMBO J*, 14(24):6136–6147, Dec 1995.

[90] Roman Timofeev. Classification and regression trees (cart) theory and applications. Master's thesis, Humboldt University, Berlin, Germany, 2004.

[91] Natal A W van Riel. Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. *Brief Bioinform*, 7(4):364–374, Dec 2006.

[92] A. M. Verhagen, P. G. Ekert, M. Pakusch, J. Silke, L. M. Connolly, G. E. Reid, R. L. Moritz, R. J. Simpson, and D. L. Vaux. Identification of diablo, a mammalian protein that promotes apoptosis by binding to and antagonizing iap proteins. *Cell*, 102(1):43–53, Jul 2000.

[93] A. Wagner and D. A. Fell. The small world inside large metabolic networks. *Proc Biol Sci*, 268(1478):1803–1810, Sep 2001.

[94] K. Wang, X. M. Yin, D. T. Chao, C. L. Milliman, and S. J. Korsmeyer. Bid: a novel bh3 domain-only death agonist. *Genes Dev*, 10(22):2859–2869, Nov 1996.

[95] Hans V Westerhoff and Bernhard O Palsson. The evolution of molecular biology into systems biology. *Nat Biotechnol*, 22(10):1249–1252, Oct 2004.

[96] Gen Sheng Wu. TRAIL as a target in anti-cancer therapy. *Cancer letters*, 285(1):1–5, 2009.

[97] Richard J Youle and Andreas Strasser. The bcl-2 protein family: opposing activities that mediate cell death. *Nat Rev Mol Cell Biol*, 9(1):47–59, Jan 2008.

[98] H. Zhang, Q. Xu, S. Krajewski, M. Krajewska, Z. Xie, S. Fuess, S. Kitada, K. Pawlowski, A. Godzik, and J. C. Reed. Bar: An apoptosis regulator at the intersection of caspases and bcl-2 family proteins. *Proc Natl Acad Sci U S A*, 97(6):2597–2602, Mar 2000.