UNIVERSITÉ DU LUXEMBOURG

PhD-FSTC-2015-36
The Faculty of Sciences, Technology and Communication

DISSERTATION

Defense held on 16/07/2015 in Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN BIOLOGIE

by

Panuwat TRAIRATPHISAN
Born on 10th April 1985 in Ubonratchathani (Thailand)

STUDYING SIGNAL TRANSDUCTION NETWORKS WITH A PROBABILISTIC BOOLEAN NETWORK APPROACH

Dissertation defense committee

Dr Thomas Sauter, Dissertation supervisor
Professor, Université du Luxembourg

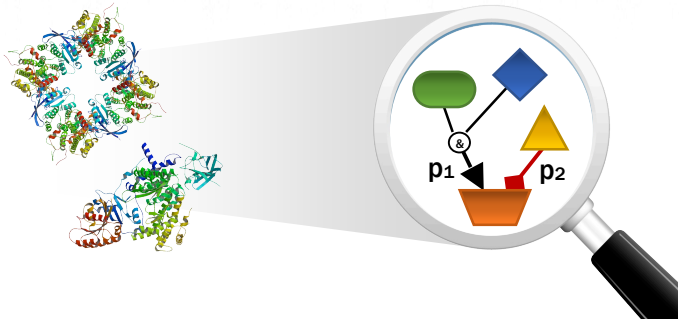
Dr Iris Behrman, Chairman
Professor, Université du Luxembourg

Dr Sjouke Mauw, Vice Chairman
Professor, Université du Luxembourg

Dr Irmgard Merfort, Member
Professor, Albert-Ludwigs-Universität Freiburg

Dr Madalena Chaves, Member
INRIA Sophia Antipolis - Mediterranée

Studying signal transduction networks with a probabilistic Boolean network approach



A dissertation by

Panuwat TRAIRATPHISAN

Life Sciences Research Unit, University of Luxembourg

Supervisor: *Prof. Dr Thomas SAUTER*

Dissertation Defence Committee:

Committee members: *Prof. Dr Iris BEHRMANN (Chair of committee)*

Prof. Dr Sjouke MAUW

Prof. Dr Thomas SAUTER

Prof. Dr Irmgard MERFORT

Dr Madalena CHAVES

AFFIDAVIT

I hereby confirm that the PhD thesis entitled “Studying signal transduction networks with a probabilistic Boolean network approach” has been written independently and without any other sources than cited.

Luxembourg, 16th July 2015

Panuwat Trairatphisan

ACKNOWLEDGEMENT

The almost-four-year PhD time is indeed one of the longest chapters in my life. I would not be able to make it this far without supports from the people surround me whom I would like to say “Thank you” as follows:

I would like to thank my colleagues in the LSRU and the LCSB who make the atmosphere in our research units lively and dynamic. I would also like to give a special thanks to Serge, Monique and Christelle for a great collaboration on the PDGF project, and also to Lis and Maiti for letting me apply my expertise to contribute to the L-plastin project. In parallel, I want to thank all team members of the Systems Biology group for their supports, especially, thanks to Lasse who spent his precious time to comment on my dissertation.

On another side of the UL campus, I would like to thank Prof. Sjouke Mauw for initiating the collaboration which brings computer sciences closer to systems biology. Also, my big thank goes to Andrzej, Jun, and Alex who keep my PhD projects running and reached the finish line with successful publications along my PhD timeline.

Six months of internship at Merrimack changed my life. I learned many “life lessons” through living and working in Boston. Even if I was alone, I rarely felt lonely as I was surrounded by good friends like Linnéa and David, and I had a wonderful working experience under day-to-day supervisions by Jeff and Brian. On this regard, I would like to thank Birgit for offering me this great opportunity.

Friends and family are always the factors that make my day. I am grateful to receive supports from my friends in Thailand and also from my new friends whom I get to know in Luxembourg. Thanks to Alex, Susi, Phú, Lamia, and friends from the CSC and the SnT for great hang-out activities and sports. Also, thanks to the MISB circle which makes me connected to good friends over years through studying and teaching. ขอบคุณม๊า, ป๊า, แม่, และญาติ ๆ ทุก ๆ คนเช่นกันที่ส่งแรงเชียร์และกำลังใจมาให้เสมอมา แม้ว่าจะอยู่ใกล้หรือไกลก็ตาม

Lastly, my PhD journey would have been pre-maturely terminated if I did not receive guidance and support from my supervisor, Prof. Thomas Sauter. I would like to express my deep gratitude for staying with me through the rise and fall and for being very patient when I was frustrated during my PhD time. Thank you for your openness and informality which make me feel comfortable to discuss and share my opinions and perspectives. I do not know if our paths will cross again once I leave the university. Nevertheless, I am grateful that I once have a great person who takes care of me. Danke schön, Professor.

TABLE OF CONTENTS

LIST OF FIGURES

(Note: the list does not include the figures in the published articles and in the manuscripts)

LIST OF ABBREVIATIONS:

4E-BP	eIF4E binding proteins 1,2, and 3
APAF-1	apoptotic protease activating factor-1
ASK1	apoptosis signal-regulating kinase 1
ATF2	activating transcription factor 2
ATM	ataxia telangiectasia mutated
ATR	ataxia and rad3 related
BAD	Bcl-2-associated death promoter
BAK	Bcl-2 homologous antagonist/killer
BAX	Bcl-2-associated X protein
Bcl-xL	B-cell lymphoma-extra large
BID	BH3-interacting domain death agonist
BIM	Bcl-2-like protein 11
BN	Boolean network
Ca^{2+}	calcium ion
CAD	caspase-activated DNAse
CaMKII	Ca^{2+}/calmodulin-dependent protein kinase II
CBM	constrained-based modelling
CDK	cyclin-dependent kinase
CREB	cyclic AMP response element-binding protein
DAG	diacylglycerol
DBN	dynamic Bayesian network
DE	differential evolution (algorithm)
DISC	death-inducing signalling complex
DNA	deoxyribonucleic acid
EA	evolutionary algorithm
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
EMT	epithelial-mesenchymal transition
Epo	erythropoeitin
ERK	extracellular signal-regulated kinase
FADD	Fas-Associated protein with Death Domain
GIST	gastrointestinal stromal tumour
GPU	general propose graphic processing unit
HGF	hepatocyte growth factor

HEK	human embryonic kidney
IAP	inhibitor of apoptosis
ICAD	inhibitor of caspase-activated DNAse
ICAM-1	intercellular adhesion molecule-1
IFNγ	interferon gamma
IκB	inhibitor of kappa-B
IKK	inhibitor of kappa-B kinase
IL-1β	interleukin-1 beta
IL-1R	interleukin-1 receptor
IP_3	inositol 1,4,5-trisphosphate
JAK	Janus kinase
JIP	JNK-interacting protein
JNK	c-Jun NH_2-terminal kinase
KSR	kinase suppressor of Ras-1
L-plastin	lymphocyte cytosolic protein 1
LPS	lipopolysaccharide
LTβR	lymphotoxin beta receptor
MAPK	mitogen-activated protein kinase
MAPK2K (MKK)	mitogen-activated protein kinase kinase
MAPK3K (MKKK)	mitogen-activated protein kinase kinase kinase
Mdm2	mouse double minute 2 homolog
mAb	monoclonal antibody
miR	micro-ribosomal nucleic acid
mRNA	messenger ribosomal nucleic acid
mTOR	mammalian target of rapamycin
mTORC1/2	mammalian target of rapamycin complex 1/2
NEMO	nuclear factor kappa-B essential modulator
NFκB	nuclear factor kappa-B
ODE	ordinary-differential equation
p70S6K	p70 S6 kinase
PBN	probabilistic Boolean network
PC	personal computer
PCA	principal component analysis
PDE	partial-differential equation
PDGF	platelet-derived growth factor
PDGFR	platelet-derived growth factor receptor
PDK1	3-phosphoinositide-dependent protein kinase 1

PHLPP1/2	PH-domain leucine-rich-repeat-containing protein phosphatases
PI3K	phosphoinositide-3-kinases
PIP_2	phosphatidylinositol (3,4)-bis-phosphate
PIP_3	phosphatidylinositol (3,4,5)-tris-phosphate
PK/PD	pharmacokinetic/pharmacodynamics
PKA	protein kinase A
PKB	protein kinase B
PKC	protein kinase C
PLCγ	phospholipase C gamma
PMA	phorbol 12-myristate 13-acetate
PP2A	protein phosphatase 2
PTEN	phosphatase and tensin homolog
RB	retinoblastoma (protein)
RIP-deubi	deubiquitinated receptor associated receptor kinase 1
RNA	ribosomal nucleic acid
RSK	ribosomal protein S6 kinase
RTK	receptor tyrosine kinase
SBTB2	systems biology toolbox 2
SCF	stem cell factor
SDS-PAGE	sodium-dodecyl-sulphate polyacrylamide gel electrophoresis
SH2/SH3	src homology 2/3 (domain)
SHP1/2	shatterproof 1/2
STAT	signal transducer and activator of transcription
TAK1	transforming growth factor-β-activated kinase 1
TGFβ	transforming growth factor beta
TLR	toll-like receptor
TNFα	tumour necrosis factor alpha
TNFR	tumour necrosis factor receptor
TSC1/2	tuberous sclerosis protein 1/2
TSMC	two-state Markov chain
UVB	ultra-violet B (ray)
XIAP	X-linked inhibitor of apoptosis
VEGF	vascular endothelial growth factor

SUMMARY

In recent years, various modelling approaches in systems biology have been applied for the study and analysis of signal transduction networks. However, each modelling approach has its inherent advantages and disadvantages, so the choice has to be made based on research objectives and types of data. In this PhD dissertation, we propose probabilistic Boolean network (PBN) as one of the suitable modelling approaches for studying signal transduction networks with steady-state data. The steady-state distribution of molecular states in PBN can be correlated to the steady-state proteomic profiles generated from wet-lab experiments. In addition, the relevance of interactions within signalling networks can be assessed through the optimised selection probabilities. These features make PBNs ideal for describing the properties of signal transduction networks at steady-state with some uncertainty on network topologies.

To investigate the applicability of PBNs for the study of signal transduction networks, we developed *optPBN*, an optimisation and analysis toolbox in the PBN framework. We demonstrated that *optPBN* can be applied to optimise a large-scale apoptotic network with 96 nodes and 105 interactions. Also, it allows for network contextualisation in a physiological context of primary hepatocytes through the analysis on optimised selection probabilities. Similarly, we also applied *optPBN* to study deregulated signal transduction networks in pathological contexts, i.e. the PDGF signalling in gastrointestinal stromal tumour (GIST) and the L-plastin signalling in breast cancer cell lines. By integrating prior information on network topology from literature with context-specific experimental data, contextualised PBNs can be derived which in turn provide additional insights into biological systems such as the importance of certain crosstalk interactions and the comparative signal flows at steady-state in non-metastatic versus metastatic cancer cell lines.

In addition to the applications on fundamental research, we also explored the applications of PBNs in a pharmaceutical setting where detailed mechanistic models are usually used. Here, we applied *optPBN* as a tool for network contextualisation. A proof-of-concept example on a small model demonstrated that *optPBN* helped to pre-select the suitable network structure according to the provided experimental data prior to the building and optimisation of detailed mechanistic models. Such application is foreseen to be applied in a pharmaceutical setting and to explore additional applications such as combinatorial drugs' effect and toxicity screening.

Chapter 1

INTRODUCTION

1.1. Cellular signal transduction networks

The cell is a basic building block of life. Higher-level complex organisations among different types of cells lead to the formation of tissues and organs with specialised functions that work systematically to maintain the well-being of an organism. On another view, complex and diverse regulatory mechanisms within the cells also operate on multiple cellular components in an orchestrated fashion to preserve the proper functionalities and integrity of each cell.

In eukaryotes, from yeast to human, multi-level cellular organisations and regulations have been established and they are well-conserved through evolution. To shortly recall on the information processing between cellular components within central dogma of molecular biology, the functional genetic components in deoxyribonucleic acid (DNA) are transcribed into messenger ribonucleic acid (mRNA) which are subsequently translated into proteins in diverse forms including structural proteins, transcription factors, signalling proteins and enzymes (Crick, 1970). Each form of these protein has designated functions, i.e. providing structural support and protect the cells from external environment, binding to DNA and regulating transcription, transducing signals from internal and external cues to regulate cellular phenotypes, and catalysing biochemical reactions within the cell, respectively. These diverse effector units are built to supply the need of cells to survive and to undergo through particular cellular events such as cell division or apoptosis according to the incoming changes and perturbations in the environment.

Among the information transfer within central dogma, it is noteworthy that each step is governed by multiple regulatory processes. This includes epigenetic regulations such as chromatin remodelling or DNA methylation, regulation of transcription machinery by various transcription factors, post-transcriptional regulation, e.g. by endogenous microRNAs (miRs), and post-translational modifications of proteins such as phosphorylation or glycosylation. Such regulatory processes modulate and contextualise the information from the genome to create diversified cellular units with specialised properties that function on particular biochemical reactions. Over decades, these regulatory mechanisms were increasingly studied as networks, which depict the interactions or relationships between cellular components. Some of the most well-studied

cellular networks are gene regulatory networks which represent the interactions among genes within genomes, metabolic networks which illustrate the pathways of biochemical reactions of metabolites, and signal transduction networks where the information processing between signalling molecules are rendered (Ryll et al., 2014).

The signal transduction network is one of the core functional layers in a cell. Upon perturbations, it is the first system which receives and integrates information from internal and external stimuli before passing them to various effectors such as transcription factors and metabolic enzymes to further regulate other functional layers including gene regulatory and metabolic networks. In terms of organisation, a signal transduction network comprises multiple cascades of signalling molecules which are classified into several intracellular signalling pathways. Some of the major intracellular signalling pathways are for instance mitogen-activated protein kinases (MAPK), phosphoinositide-3-kinases (PI3K)/AKT (a.k.a. protein kinase B, PKB)/mammalian target of rapamycin (mTOR), phospholipase C gamma/protein kinase C (PLCγ/PKC), Janus kinase and signal transducer and activator of transcription (JAK/STAT) and Caspase pathways (Bos, 2000). Each of these pathways supports various cellular functions such as promoting proliferation, differentiation, senescence, or activating apoptotic machinery depending on the cell types.

In a physiological state, signal transduction networks maintain an optimal signal flow towards effector units. Multiple cues from stimuli such as growth and death factors, cytokines, hormones, and cell-to-cell contacts are sensed by the corresponding receptors including receptor tyrosine kinases (RTKs), cytokine receptors, transmembrane and hormonal receptors, as well as integrins and E-cadherins which in turn pass the signals towards intracellular signalling pathways. The most common route to transduce signals is mediated via the phosphorylation and dephosphorylation of signalling molecules by kinase and phosphatase enzymes. To modulate the signals within the signal transduction processes, a large number of natural regulatory mechanisms are taking place such as the dephosphorylation by cytoplasmic phosphatases or by target-specific phosphatase enzymes, e.g. phosphatidylinositol-3,4,5-triphosphate 3-phosphatase (PTEN), proteosomal degradation following ubiquitination, and the formation of inactive protein complexes, e.g. inhibitor of kappa B (IκB) and nuclear factor kappa B (NFκB) complex. Apart from these regulatory mechanisms, increasing evidence in literature also reports the importance of regulatory effects from crosstalk interactions between intracellular signalling pathways (Su, Mei, & Sinha, 2013; Xia & Storm, 2012). It was shown that crosstalk interactions play a major role in the fine-tuning of signals and they are also involved in the

compensatory mechanism once the signal flows within the canonical pathways are disturbed (Mendoza, Er, & Blenis, 2011).

Generally, the regulatory and compensatory mechanisms are able to cope with small and transient perturbations within signal transduction network. However, for certain perturbations, e.g. the mutations of key signalling molecules such as Ras in MAPK pathway or PTEN in PI3K/AKT/mTOR pathway, the signal flows are deregulated and inappropriate signals were transduced to the effector units which in turn leads to pathological phenotypes of cells. The abnormality in cellular phenotypes subsequently serves as the etiology in many types of diseases including neurodegenerative diseases, metabolic diseases as well as cancers (E. K. Kim & Choi, 2010; Lawrence et al., 2008).

1.2. Deregulation of signalling pathways in cancers

Cancer is one of the major causes of death in human worldwide since decades (Jemal et al., 2011). It is often characterised by the over-proliferation of cancerous cells which invade tissues and cause the failure of homeostatic regulation in the human body. To date, there is no simple treatment which can eradicate all types of cancer due to the high complexity of cellular and tissue organisations within tumours. Thus, a deeper and more comprehensive understanding in the field of cancer research is required to decipher the pathogenesis of cancers and to develop more advanced therapeutic treatments.

In 2000, Hanahan and Weinberg introduced six hallmarks of cancers which hold true still today (Hanahan & Weinberg, 2000). These hallmarks, which are considered to be the common phenotypes in almost all types of cancers, include self-sufficiency in growth signals, insensitivity to anti-growth signals, limitless replication potency, evasion of apoptosis, sustained angiogenesis, and tissue invasion and metastasis. Interestingly, the molecular mechanisms which give rise to the six cancerous phenotypes can be explained by the deregulation of normal physiology in signal transduction networks. A few examples are as follows: the independence of growth signals for cell proliferation can be due to the over-expression of receptor tyrosine kinases, the reduction of apoptotic machinery can be caused by the up-regulation of anti-apoptotic oncogenes such as *bcl-2,* and the changes in integrin and E-cadherin expressions could lead to more invasive cellular phenotypes. From this perspective, it was proposed that the deregulation of signal transduction networks is the central mechanism that drives carcinogenic transformation. The integrated circuit of signal transduction networks comprising multiple intracellular signalling cascades are depicted in Figure 1.

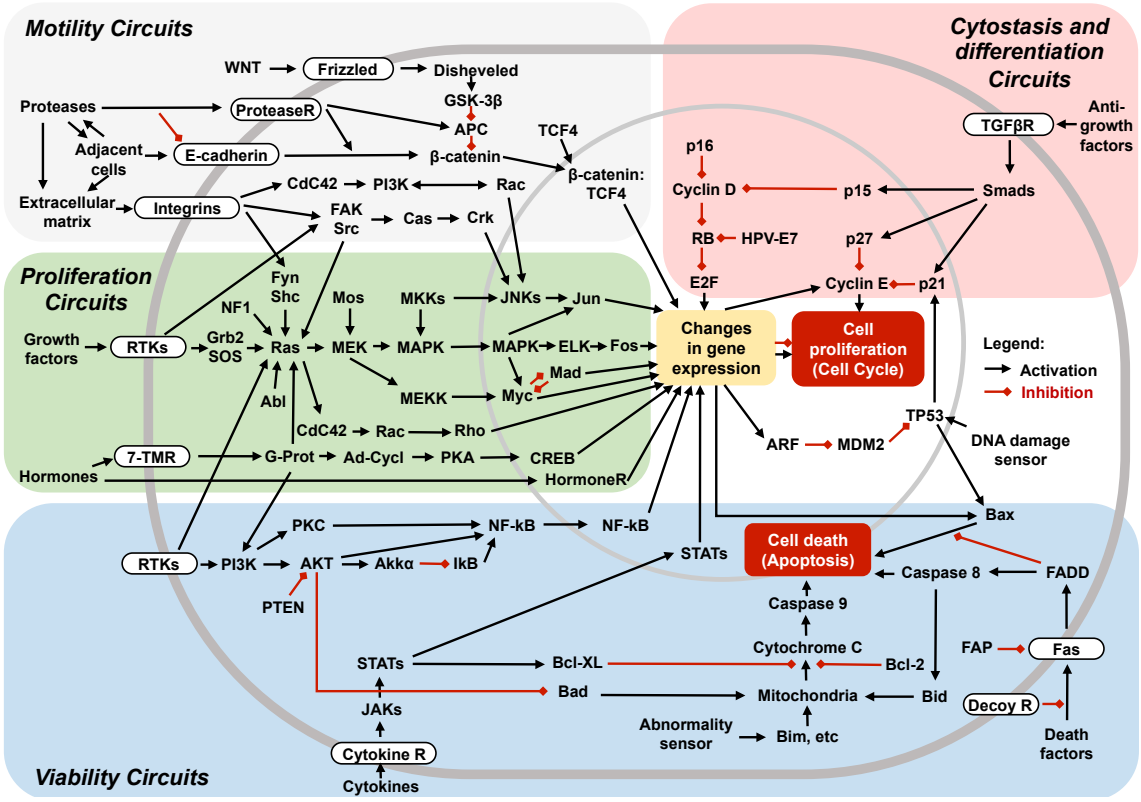


Figure 1. Integrated circuit of a cellular signal transduction network. Multiple intracellular signalling cascades were depicted together with the points of signal integration via crosstalk interactions. Adapted from (Hanahan & Weinberg, 2000) under licensing number 3585911018186 and (Hanahan & Weinberg, 2011).

Since the hallmarks of cancer were established, the body of cancer research based on this conceptual work was largely expanded over the last decade. In 2011, two new cancer hallmarks which are deregulating cellular energetics and avoiding immune destruction together with two enabling characteristics which are tumour-promoting inflammation and genome instability and mutations were additionally proposed to participate along the carcinogenic progression of cancer cells (Hanahan & Weinberg, 2011). It was shown that the reprogramming of glucose metabolism via aerobic glycolysis (Warburg, 1956) could divert glycolytic intermediates towards other biosynthetic pathways (Vander Heiden, Cantley, & Thompson, 2009) and the by-product lactate can also be re-utilised within the lactic cycle (Feron, 2009). In addition, the role of innate immune system-driven inflammation during tumour progression (DeNardo, Andreu, & Coussens, 2010) and the mechanistic details of epithelial-mesenchymal transition (EMT) mediated by a set of transcription factors, e.g. Snail, Slug, Twist, and Zeb1/2 (Klymkowsky & Savagner, 2009) were also illustrated. These advances on identifying additional cancer etiology opened up a new possibility for discovering novel cancer therapeutic treatments such as

metabolic and immunological therapies which were not concretely focused for treating malignancies in the past.

Apart from investigating the molecular transformations of normal cells into cancer cells as a basis to understand carcinogenesis progression, the focus on the tumour microenvironment was also recently emphasised as an important rudimentary unit which allows cancer cells to survive and to further spread to adjacent and distant tissues (Egeblad, Nakasone, & Werb, 2010). The composition of each tumour microenvironment generally comprises cancerous cells together with supporting structures such as cancer-associates fibroblasts, vascular endothelium, pericytes, inflammatory immune cells as well as cancer stem cells. These cells elicit a heterotypic signalling among different cell types to foster their local expansion and invasiveness. In parallel, they also establish a reciprocal interaction with surrounding supporting structures such as stromal reservoirs to exploit the source of nutrients before moving to a distant site. In a causal view, the deregulation of signalling pathways was shown to be one of the principal contributors to the formation of the tumour environment (Hanahan & Weinberg, 2011). For instance, cancer cells up-regulate the expression of vascular endothelial growth factor (VEGF) and its receptors to promote the proliferation of endothelial cells and angiogenesis (Ferrara & Davis-Smyth, 1997). Another example is an over-expressed platelet-derived growth factor (PDGF) and its receptors mediated by tumour cells leading to the recruitment of mesenchymal cells such as fibroblasts to tumour sites which in turn secrete hepatocyte growth factor (HGF) in order to promote tumour metastasis (Hanahan & Weinberg, 2011). At the molecular level, these growth factors serve as ligands for RTKs which activate downstream signalling cascades to transform cellular phenotypes in favour of malignancy. Such line of arguments additionally highlights the importance of deregulated signal transduction networks, both at the levels of receptors and downstream intracellular signalling pathways, which largely contribute to cancer advancement.

To further illustrate the role of major intracellular signalling pathways and their contributions to cancer pathogenesis as summarised in Figure 1, I shortly review the components and their roles in cancer signalling in the scope of this dissertation as follows:

1.2.1. Mitogen-activated protein kinases (MAPK) pathway family

The MAPK pathway comprises a cascade of three or more protein kinases which are MAPK kinase kinases (MAP3Ks), MAPK kinases (MAP2Ks) and MAPKs. The series of molecular activation are modulated by phosphorylation on serine, threonine and tyrosine residues. The mammalian MAPK family consists of 3 downstream effectors which

are extracellular signal-regulated kinase (ERK), p38 and c-Jun NH_2-terminal kinase (JNK) which exist in several forms, i.e. ERK1 to ERK8, p38-α, -β, -γ, -δ, and JNK1 to JNK3. The activation within the MAPK pathway are moderated by either one-to-one interaction between kinase components or by the guidance of scaffold proteins, e.g. kinase suppressor of Ras-1 (KSR), JNK-interacting protein (JIPs) and β-Arrestin (E. K. Kim & Choi, 2010).

On signal processing, the three different MAPK pathways respond to different stimuli with a designated set of molecular players. The JNK pathway responses to cellular stress and inflammatory cytokines by activating apoptosis signal-regulating kinase 1 (ASK1) or transforming growth factor-β-activated kinase 1 (TAK1) which pass the signals towards mitogen-activated protein kinase kinase (MKK) 6 and MKK3 that in turn trigger JNK activity. In parallel, the p38 pathway responds to the same set of stimuli as JNK pathway and it shares the activation on ASK1 and TAK1 downstream. However, this pathway further activates another set of MKK which are MKK4 and MKK7 before eliciting the function of p38. Lastly, the ERK pathway, which is usually activated in response to the activation of the RTK-Growth factor receptor bound protein 2 (Grb2)-Son of Sevenless (SOS) cascade via the members of the Ras family proteins including K-Ras, H-Ras and N-Ras, processes the information through Raf isoforms such as A-Raf, B-Raf or C-Raf (Raf-1), MEK1/2 (MKK1/2) and ERK1/2 successively. Subsequently, the activated MAPKs from each pathway phosphorylate various substrates proteins such as ELK-1, activating transcription factor 2 (ATF2) and TP53 (E. K. Kim & Choi, 2010; Lawrence et al., 2008).

Physiologically, each MAPK pathway is optimally triggered to maintain proper functions in response to corresponding stimuli while pathologic deregulation of MAPK pathways leads to the destabilisation of homeostatic regulation that eventually results in disease. The alterations in MAPK signalling cascades were shown to be the cause of multiple diseases including neurodegenerative diseases such as Alzheimer's disease and Parkinson's disease, chronic and genetic diseases such as Crohn's disease and polycystic kidney disease, and also many types of cancers (E. K. Kim & Choi, 2010). Even though it was shown that the anomalies in JNK and p38 pathways are also the etiology of many diseases, only the aberration in ERK pathway, on which we will further focus in this dissertation, was clearly demonstrated to be the promotor of carcinogenesis (see Fig. 1).

Among the evidence from literature in early years of cancer research, ERK1/2 activity was identified as a crucial contributor to the pathogenesis of cancers (Land, Parada, & Weinberg, 1983). Cancer-associated mutations of the components in ERK

pathway, especially of Ras and Raf, often up-regulate signal flows in this pathway and facilitate the cells to obtain cancerous phenotypes. For instance, the mutations of K-Ras were often found in many human cancers such as lung and colon cancers (Schubbert, Shannon, & Bollag, 2007) while the mutations of B-Raf gene were discovered in more than half of malignant melanomas (Calcagno et al., 2008). Looking at the hallmarks of cancer at molecular level, it was shown that the ERK pathway plays a major role in many steps of tumour development. The key role of this pathway is to transduce growth signals from RTKs into the cells in order to initiate cellular proliferation. Hence, the abnormality of upstream signalling components, such as the over-expression of ligands and RTKs, also contributes to the clonal expansion of cancer cells as well as their supporting structures (Hynes & MacDonald, 2009). In addition, the ERK pathway was also found to contribute to other cancer hallmarks such as increasing cell invasion by inducing the expression of matrix metalloproteinases (Chakraborti, Mandal, & Das, 2003) and promoting survival of cancer cells by regulating the activities of Bcl-2 family proteins (Balmanno & Cook, 2009). Given the broad role of the ERK pathway in different stages of carcinogenesis, it has been now one of the most targeted signalling pathways for the treatment of cancers.

1.2.2. PI3K/AKT/mTOR pathway

The PI3K/AKT/mTOR pathway is highly conserved in mammalian signalling networks and its activation is tightly controlled in multiple steps. Class-I PI3Ks are heterodimers composed of two subunits, a catalytic subunit (p110) and a regulatory subunit (p85). The Class-I PI3Ks are also classified into subclasses, i.e. subclass IA which interacts with RTKs and subclass IB which is activated by receptors coupled with G-protein. Upon the activation of these receptors, the p85 subunit of PI3K binds to the receptor by recognising the phospho-tyrosine residue on the receptors with SH2 domain and activates the enzymatic activity of the p110 subunit which in turn converts phosphatidylinositol (3,4)-bis-phosphate (PIP_2) into phosphatidylinositol (3,4,5)-tris-phosphate (PIP_3) on cell membrane. Active PIP_3 at the cell membrane then recruits AKT to bind, allowing 3-phosphoinositide-dependent protein kinase 1 (PDK1) to access and phosphorylate threonine 308 (T308) residue of AKT which leads to partial AKT activation. Subsequently, partially activated AKT activates the mTOR complex 1/2 (mTORC1/2) and inhibits tuberous sclerosis protein 2 (TSC2) which inhibits the activity of mTORC1/2. The unleashed mTORC1/2 then reciprocally phosphorylates the second regulatory phosphorylation site serine 473 (S473) on the C-terminal tail of AKT which leads to the full activation of AKT activity. The activities of AKT can be counter-regulated by the dephosphorylation of T308 by protein phosphatase 2 (PP2A), the dephosphorylation of S473 by PH-domain leucine-rich-repeat-containing protein phosphatases (PHLPP1/2),

and the reconversion of PIP_3 to PIP_2 mediated by phosphatase and tensin homolog (PTEN). (Fresno Vara et al., 2004; Hemmings & Restuccia, 2012)

Once the downstream signalling molecules of the PI3K/AKT/mTOR pathway, i.e. AKT and mTORC1/2 are fully activated via phosphorylation, they further interact with a large number of their downstream targets and regulate a variety of cellular functions (see Fig.1). To demonstrate some examples, AKT is involved in cell-death regulation by phosphorylating and inhibiting pro-apoptotic factor Bad and procaspase-9. In addition, it also activates an inhibitor of cyclic AMP response element-binding protein (CREB) and kappa-B kinase (IKK) in the NFκB pathway which results in the transcription of anti-apoptotic genes. In parallel, it was shown that AKT also modulates cell cycle progression and cell growth by regulating the localisation of targeting proteins. For instance, Akt inhibits glycogen synthase kinase 3 (GSK3) by phosphorylation which in turn impedes the degradation of β-catenin. This results in the increase of β-catenin translocation into nucleus to bind transcription factors, e.g. TCF/LEF-1 to induce several genes that in turn induce cell cycle progression, e.g. cyclin-dependent kinase (CDK) 1 which inactivates retinoblastoma (RB) protein via hyperphosphorylation. In addition, AKT phosphorylates and inhibits the anti-proliferative proteins p21 and p27 by retaining them in the cytoplasm and it also facilitates nuclear translocation of Mouse double minute 2 homolog (Mdm2) to antagonise the effect of TP53 on cell cycle checkpoints. On the other path, the activated molecules downstream of the mTOR pathway further stimulate protein synthesis by phosphorylating p70 S6 kinase (p70S6K) and eIF4E binding proteins 1,2, and 3 (4E-BPs) where p70S6K increases translation of mRNAs via the phosphorylation of ribosomal protein S6 and 4E-BPs release the initiation factor eIF4E to promote translation of many proteins e.g. cyclin D1, Myc, and VEGF. (Altomare & Testa, 2005; Fresno Vara et al., 2004). To summarize, both AKT and mTOR principally activate various cellular mechanisms to increase the total cell population.

In the pathological context, the deregulation of PI3K/AKT/mTOR pathway also leads to an abnormality in cellular phenotype which is the cause of multiple diseases such as metabolic diseases (Manning & Cantley, 2007) as well as cancers (Altomare & Testa, 2005; Fresno Vara et al., 2004). As mentioned previously, both AKT and mTOR participate in the regulatory circuit of cell survival and cell cycle progression. Hence, this implies that both effector units could also promote cellular proliferation and evasion of apoptosis in cancerous cells. In addition, the by-products of mTOR, e.g. VEGF, also has an angiogenic effect which additionally reinforces the formation of tumours in local and distant sites. Based on the survey on prostate, breast and ovarian cancer, it was shown

that nearly 80% of tumours with activated AKT1 were high grade stage III/IV carcinomas (Sun et al., 2001) while 30-40% of pancreatic and ovarian cancers were found to have high activity of AKT2 (Altomare et al., 2005). In addition, the aberration of upstream signalling molecules of the PI3K/AKT/mTOR axis is also involved in multiple types of cancers. To name a few, the mutation of p110 subunit of PI3K was reported in ovarian carcinoma (Shayesteh et al., 1999) while the mutated gene that codes for the p85 regulatory subunit of PI3K was also shown in a T-cell lymphoma cell line (Jucker et al., 2002). The loss of PTEN function due to somatic mutation also elevates the signal flux in the PI3K/AKT/mTOR pathway via PIP_3 elevation which occurs frequently in glioblastoma, melanoma and cancers of prostate and endometrium (Sansal & Sellers, 2004). Taken together, it appears that the deregulation of the PI3K/AKT/mTOR pathway also plays a central role in cancer progression by activating multiple cancer hallmarks. Hence, their components are worth to be further investigated together with the ones in the MAPK pathway to design more advanced targeted therapies.

1.2.3. PLCγ/PKC pathway

In mammalian cells, up to fourteen PLC-isozymes were characterised and classified into multiple subtypes (Hwang et al., 2005). Among them, one of the well-studied isozymes is PLCγ-1 which has a strong implication in signal transduction. PLCγ-1 has Src homology 2 and 3 (SH2 and SH3) domains, which interact with proteins via phosphotyrosine and proline-rich sequences, respectively. The classical signalling cascade of the PLCγ/PKC pathway starts from the interaction of extracellular signalling molecules such as hormone and growth factors with their receptors on cell surface which in turn triggers a rapid hydrolysis of membrane phospholipid PIP_2 catalysed by PLCγ isozymes. The biochemical reaction leads to the generation of two intracellular messengers, diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP_3) which further activate PKC and promote the release of calcium ion (Ca^{2+}) from intracellular storage. The activated PKC and the gradient of increased Ca^{2+} concentration subsequently modulate the activities of signalling molecules which are the components of other intracellular signalling pathways with diverse functions (Choi, Ryu, & Suh, 2007; Putney, 2002).

The deregulation of the PLCγ/PKC pathway has been investigated for its potential role in promoting tumour formation and metastasis. In one study, PLCγ was shown to mediate together with epidermal growth factor receptor (EGFR) to promote tumour cell invasion in prostate cancer cell lines (Kassis et al., 1999). In other studies, it was shown that the expression level of PLCγ-1 is elevated in cancer tissues comparing to normal counterpart (Nanney et al., 1992) and it also plays an essential role in cellular proliferation

and anti-apoptosis in various cell types (M. J. Kim et al., 2000). Thus, the PLCγ/PKC pathway is one of the important intracellular signalling pathways with direct links to many hallmarks of cancers that needs to be considered and controlled in cancer treatment.

1.2.4. JAK/STAT pathway

The JAK/STAT pathway is known to be ubiquitous among vertebrates but it can also be found in other metazoans such as *Drosophila* (Hombría & Brown, 2002). In human, the JAK family consists of 4 cytoplasmic tyrosine kinases, i.e. JAK1-3 and Tyk2 and 7 proteins were identified in the STAT family, i.e. STAT1-6 with STAT5a and STAT5b. Upon the activation of signalling receptors in response to extracellular growth factors and cytokines, multimerisations of receptor subunits are formed and lead to the activation of JAKs which are bound to receptors via phosphorylation. Subsequently, activated JAKs recruit, phosphorylate, and dimerise STATs which then translocate into nucleus to interact with various regulatory elements on the genome in order to modulate gene expression. The activities of STATs are controlled by negative regulators including tyrosine phosphatases such as SHP1/2, protein inhibitors of activated STATs (PIAS) and suppressors of cytokine signalling (SOCS) proteins (Furqan, Mukhi, Lee, & Liu, 2013; Rawlings, Rosler, & Harrison, 2004).

Various types of ligands such as erythropoietin (Epo), interleukin-6 (IL-6) and interferon-gamma (IFNγ) give rise to multiple combinations of JAK and STAT family activations (Schindler, 2002). The diversities of JAK/STAT organisation subsequently lead to complex regulations on multiple cellular functions including cell proliferation, differentiation, migration, and apoptosis (Rawlings et al., 2004). On disease pathology, the deregulation of JAK/STAT signalling was found to be the etiology for many types of haematological malignancies, such as leukaemia and lymphoma. Among the effector units, the anomaly of certain molecules, e.g. the constitutively active mutants of STAT3, has a key role in the regulation of proliferation as well as oncogenesis in several tissues (Levy & Lee, 2002). Such evidence inevitably points to the involvement of the JAK/STAT pathway in the cancerous signalling circuit.

1.2.5. NFκB pathway

The NFκB pathway is one of the signalling cascades that is conserved through the evolution from *Cnidaria* to humans (Sullivan, Kalaitzidis, Gilmore, & Finnerty, 2007). In total, five members of NFκB family, which serve as transcription factors, were identified including RelA (p65), RelB, c-Rel, NFκB1 (p105) and NFκB2 (p100). In most quiescent

cells, dimers of NFκB family are bound to the inhibitory molecules of the IκB family. To activate the NFκB pathway, 3 routes have been proposed (Gilmore, 2006). In the canonical pathway, inflammatory-related ligands such as tumour necrosis factor alpha (TNFα), lipopolysaccharides (LPS) and interleukin-1 beta (IL-1β) bind to their receptors, i.e. TNF receptor (TNFR), toll-like receptor (TLR) and IL-1 receptor (IL-1R) which in turn activate IκB kinase (IKK) complexes and NFκB essential modulator (NEMO). IKKs subsequently phosphorylate IκBs which leads to proteasomal degradation of IκBs by ubiquitination, allowing the p65/p50 dimers to enter the nucleus and regulate the expression of up to thousand genes. In an alternative pathway, a different set of activated receptors, e.g. lymphotoxin beta-receptor (LTβR) and CD40 induce the activity of NFκB inducing kinase (NIK) which in turn leads to the ubiquitination and partial degradation of p100 into p52 that binds to RelB before entering nucleus. In an atypical pathway which is activated, e.g. by genotoxic stress, the signals are mediated via the signalling kinase ataxia telangiectasia mutated (ATM) which in turn activates NEMO and follows the route of the canonical pathway. The activity of NFκB is controlled by the IκB family as well as by specific NFκB pathway regulators such as A20 (Gilmore, 2006; Hoesel & Schmid, 2013).

In terms of cellular functions, the NFκB pathway is well-known for its involvement in the inflammatory responses mediated primarily by innate immunity. In addition, it was shown that chronic inflammation exerts a pro-tumorigenic effect rather than a protective one, e.g. by up-regulating the expression of anti-apoptotic genes (Dunn, Old, & Schreiber, 2004). In addition, it was shown that the NFκB pathway also contributes to the EMT process (Huber et al., 2004) and neovascularization (Xie, Xia, Zhang, Gong, & Huang, 2010). Even if it is far-fetched to speculate that solid tumours arise from the deregulation of this pathway alone, there is no counter-argument on its involvement in tumourigenesis.

1.2.6. Cell cycle signalling

In response to growth factors, the cell cycle signalling is initiated and drives the cells from the resting state (G_0) through the interphase (G_1/S/G_2) and mitosis (M) cycle, resulting in cell division. The classical model of cell cycle in mammalian cells comprises several cyclin proteins and CDKs which are required at different phases. The most critical steps known as cell cycle checkpoints are G_1/S and G_2/M transitions where they are controlled by cyclin D:CDK4/6 with cyclin E:CDK2 and cyclin A/B:CDK1, respectively. The first checkpoint is controlled by the RB protein which binds to the E2F transcription factor that up-regulates the expression of cyclin E. Additionally, the signalling molecules at the checkpoints can also be regulated by a group of CDK inhibitor families with a broad range of regulatory effects. To mention a few, p15 specifically inhibits cyclin D:CDK4/6 complex

while p21 (*CDKN1A* gene) has wider targets and acts on multiple cyclins:CDKs complexes. The expression of these regulatory molecules can be up-regulated by the activation of anti-growth factors such as transforming growth factor beta (TGFβ) which transduces signals via Smads proteins. Also, TP53, as the chief guardian of genome integrity, can up-regulate p21 and it functions together with other DNA-damage sensor protein kinases such as ATM, ataxia and rad3 related (ATR), and Chk1/2 to ensure the finest quality of DNA sequence before proceeding to DNA replication and mitosis. (Vermeulen, Van Bockstaele, & Berneman, 2003)

Intuitively, the cell cycle signalling is highly involved in carcinogenesis with respect to the control of cell population. The deregulation or mutations of signalling molecules within this signalling pathway, e.g. the mutations of cyclin proteins, CDKs, p15, p21, RB protein, and especially TP53, lead to the alteration of signal flows toward cell cycle and also to the instability of the genomes in the progeny cells. Interestingly, recent studies revealed that CDK1 is the only essential element to drive cell cycle in most mammalian cells while the interphase CDKs, i.e. CDK2/4/6 are only essential in some specific cell types. Such evidence ignites a new line of cancer treatment on cell cycle signalling which can be applied to patients in a more personalised manner (Malumbres & Barbacid, 2009).

1.2.7. Apoptotic signalling

Apoptosis is one of the components in programmed cell death together with autophagy and programmed necrosis (Ouyang et al., 2012). The classical apoptotic pathway is divided into two parts: the extrinsic pathway and the intrinsic pathway where both involve a series of energy-dependent cysteine protease enzymes called caspases. The extrinsic pathway is triggered by the binding of death ligands, e.g. Fas or TNF ligands, to their corresponding death receptors. The ligand-receptor complex recruits adaptor proteins, e.g. Fas-associated protein with death domain (FADD) and forms a death-inducing signalling complex (DISC) which in turn activates caspase 8 and the executioners of apoptosis, i.e. caspases 3,6,7, successively. In parallel, the intrinsic pathway senses perturbations from external and internal stimuli and direct signals towards the Bcl-2 family proteins which are involved in the release of mitochondrial enzymes. Two distinct members of the Bcl-2 family are classified, the pro-apoptotic members such as BAD, BAX, BAK, BID, BIM and the anti-apoptotic members, e.g. Bcl-2, Bcl-xL, Mcl-1. The balancing of the two groups control the permeability of the inner membrane of mitochondria. The disruption of the mitochondrial membrane results in the release of cytochrome c which binds to apoptotic protease activating factor-1 (APAF-1). Activated APAF-1 then leads to the activation of caspase 9 that in turn activates effector caspases

3, 6, 7. Once the executors of apoptosis are activated, they further activate proteases which degrade nuclear and cytoskeleton proteins. The most important executioner is caspase 3 which specifically activates endonuclease caspase-activated DNAse (CAD) by cleaving inhibitor of CAD (ICAD), leading to chromatins breakage and condensation. Subsequently, apoptotic cells externalise phosphatidylserine on their cell surfaces as a marker to be engulfed by phagocytic cells to complete the apoptotic process (Elmore, 2007; Ouyang et al., 2012). Another apoptotic pathway mediated by Perforin and Granzyme stimulated by cytotoxic T-cell is shortly reviewed in (Elmore, 2007).

In the view of cellular functions, the apoptotic pathway is a vital component of multiple cellular processes including embryonic development, normal cell turnover and chemical-induced cell death. The deregulation of this pathway therefore contributes to the pathogenesis of many diseases such as autoimmune diseases, neurodegenerative diseases, and cancer. In the carcinogenic process, evasion of apoptosis is one of the key cancer hallmarks that cancer cells need to acquire which can be achieved by perturbing the regulatory units of this pathway. Apart from the regulation by the Bcl-2 family proteins, there also exist other regulators in the group of inhibitor of apoptosis (IAP), e.g. X-linked IAP (XIAP) which binds to Apaf-1 and inhibits its activity in the intrinsic pathway. In addition, TP53 also plays an important role in sensing cellular threats and controlling the release of mitochondrial cytochrome c via BAX (Benchimol, 2001). The aberration of these regulators thus increases the survival chance of cancerous cells. Furthermore, there are increasing evidences that other signalling cascades also send crosstalk signals to regulate apoptosis, e.g. the PI3K/AKT/mTOR pathway could inhibit the activities of Bad via AKT (Vivanco & Sawyers, 2002) or the increased expression of pro-apoptotic genes can be mediated by NFκB (Karin & Greten, 2005). For that reason, new therapeutic targets in apoptotic signalling in cancers would not be effective if one simply targets only an individual gene or protein components (Ouyang et al., 2012). By contrast, one needs to integrate the interconnected signalling components as a network to unravel the complex regulations of this pathway and to prevent the activation of alternative pathways that contribute to the relapse phase of diseases via crosstalk interactions.

Apart from the short reviews of intracellular signalling pathways as demonstrated, short reviews for other signalling pathways, e.g. Integrin, Wnt/β-catenin, and Src signalling can be found in (Harburger & Calderwood, 2009), (Clevers & Nusse, 2012), and (Parsons & Parsons, 2004), respectively. To further illustrate the connections between aberrant carcinogenic signals and the deregulation of intracellular signalling cascades, we introduce two examples of aberrant signal transduction networks arising from mutated

receptor tyrosine kinases which in turn contribute to the pathogeneses of cancers focused within this dissertation as follows:

1.2.8. Deregulated PDGF signalling in gastrointestinal stromal tumours (GISTs)

Gastrointestinal stromal tumours (GISTs) are the most common primary mesenchymal neoplasia of gastrointestinal tract. These tumours commonly arise from the mutations of either KIT (85-95%) or PDGFRα (5-7%) where the respective mutation sites can be located in extracellular domain, juxtamembrane domain, split kinase domain, or activation loop domain (Tan, Zhi, Shahzad, & Mustacchia, 2012). The first-line therapy for GISTs is surgical resection combined with the administration of Imatinib mesylate (Gleevec®), a tyrosine kinase inhibitor with activities against ABL, BCR-ABL, KIT, and PDGFRα/β. The clinical outcome of the first-line therapy is mostly satisfactory with 35%-49% 9-year survival (Rammohan et al., 2013). However, certain point mutations, e.g. aspartate (D) to valine (V) mutation at amino acid 842 in the *PDGFRA* gene in exon 18 (activation loop domain) are usually associated with drug resistance (Markku & Jerzy, 2006) while the other similar mutations such as aspartate (D) to tyrosine (Y) at the same amino acid or valine (V) to aspartate (D) at amino acid 561 on the *PDGFRA* gene still have good therapeutic outcomes. Such observation requires further investigation at the molecular level to understand how the signals are processed mechanistically.

Once amino acids at the regulatory sites of PDGFRα are mutated in GISTs, these mutations render constitutively active signalling towards downstream intracellular signalling pathways including Ras/Raf/MAPK and PI3K/AKT/mTOR pathways (Gramza, Corless, & Heinrich, 2009). These two main signalling pathways largely contribute to the acquisition of multiple hallmarks of cancers, notably on cell proliferation and cell survival circuits. In addition, PDGFRα was also reported to activate other signalling pathways such as the PLCγ-PKC pathway as well as the tyrosine kinases of the Src family (Heldin, 2014). Both of these pathways have numerous interaction partners within the intracellular signalling circuits, leading to the increase in complexity for further analysis. To add up the challenge, a number of crosstalk interactions between intracellular signalling pathways were proposed to be involved in the regulations within the generic mammalian signal transduction network (Grammer & Blenis, 1997; Lynch & Daly, 2002) as well as within the PDGF-signalling-specific context (C.-C. Wang, Cirit, & Haugh, 2009). At this level of complexity, it seems that only the analyses performed by computational means in the field

of bioinformatics or systems biology sound plausible to derive a global conclusion on the properties of deregulated PDGF signalling pathway at steady-state.

Recently, (Bahlawane et al., 2015) applied a bioinformatics approach to study the deregulated PDGF signalling in GISTs. In their study, they discovered that constitutive signalling via the oncogenic PDGFRα mutants in GISTs leads to the mislocalisation of the receptors which in turn modifies the signalling characteristics via STAT factors in the endoplasmic reticulum. Apart from this work, the importance of crosstalk interactions within the deregulated PDGF signalling in GISTs was investigated and analysed with a systems biology approach in parallel as presented in this dissertation.

1.2.9. Deregulated EGFR and L-plastin signalling in breast cancers

Breast cancer is one of most common forms of malignancy in female worldwide (Tinoco, Warsch, Glück, Avancha, & Montero, 2013) where complex interactions between multiple signalling pathways predominantly contribute to the course of pathogenesis. The three most common signalling receptors which have been characterised to be involved in breast cancer are oestrogen receptor alpha, progesterone receptor, and ErbB2 receptor, while EGFR was thought to be minimally associated. However, recent evidence in literature report that EGFR signalling might play a major role in triple-negative breast tumours as 50%-70% of them exhibit EGFR expression without the presence of the three most common receptors (Burness, Grushko, & Olopade, 2010). Hence, the molecular mechanisms of deregulated EGFR signalling was gaining interests to be further investigated and developed into therapeutic targets in pharmaceutical industry (Foley et al., 2010).

Generally, the deregulation of EGFR signalling is well-characterised as the etiology of epithelial tumours such as colorectal, lung, and skin tumours (Scaltriti, 2006). Besides, the pathological regulations of this signalling pathway in breast cancer were also increasingly reported over years (Foley et al., 2010). Apart from the induction of classical downstream intracellular signalling pathways such as Ras/Raf/MAPK, PI3K/AKT/mTOR, PLCγ/PKC and JAK/STAT pathways which promote the acquisition of cancer hallmarks, it was demonstrated that EGFR can also be directly shunted into nucleus as a transcriptional regulator (Lo, Hsu, & Hung, 2006). Furthermore, the elevated expression levels of EGFR and ErbB2 were associated with bone metastasis and they are considered as markers for disseminated breast tumour cells (Braun et al., 2001). As it is not yet clear why bone is the preferred metastatic site for breast cancer, this observation still needs further investigations to clarify the regulatory mechanism of such cancer phenotypes.

Regarding the role of EGFR signalling in cell invasion and metastasis, it was proposed that downstream targets of this signalling pathway, e.g. ERK1/2 in the MAPK cascade, leads to a more invasive behaviour (Foley et al., 2010). In parallel, the effective cytoskeleton units of the metastatic process might also be mediated by the deregulation of actin-binding proteins such as cofilin, filamin, or Lymphocyte cytosolic protein 1 (L-plastin) (Samstag, Eibert, Klemke, & Wabnitz, 2002). Within the framework of collaboration between the Cytoskeleton and Cell plasticity lab and the Systems Biology lab at the Life Sciences Research Unit of the University of Luxembourg, we applied a systems biology approach to investigate the interconnection between signalling components in MAPK pathways and the surrounding signalling molecules of L-plastin in order to unravel the regulatory mechanisms of cell invasiveness in 4 breast cancer cell lines as presented in this dissertation.

1.3. Model-based data integration approaches in systems biology

Biology is the science of life where its ultimate goal is to understand the structures and dynamics of living species from the molecular to the organism levels. Biological research in the early years focused primarily on identifying and characterising individual cellular components and their distinctive functions with the hope to understand physiology and pathology of life with simplistic causal relationships. However, advanced research revealed that, even if there exist high-throughput technologies which allow for discovering components and functions of individual cellular components at the omics scale, it is usually insufficient to explain the pathophysiology of many diseases as these components function together as a system with underlying regulatory mechanisms. Hence, biological sciences in the 21st century are moving towards the study of biological processes as a network which comprises an ensemble of biological components that perform particular functions (Kitano, 2002).

In order to assemble biological components as a functional system, one needs to integrate information on network structures and interactions derived from literature to experimental data in the context of the study. With this respect, mathematical models, which represent the essential components of the systems in a mathematical form, can be applied as a mean for data integration which allows for illustrating network topology as well as representing the dynamics of the systems depending on the modelling framework.

To date, a large number of modelling approaches were applied to study and analyse biological systems. Some examples of these approaches with their key characteristics are presented in Figure 2.

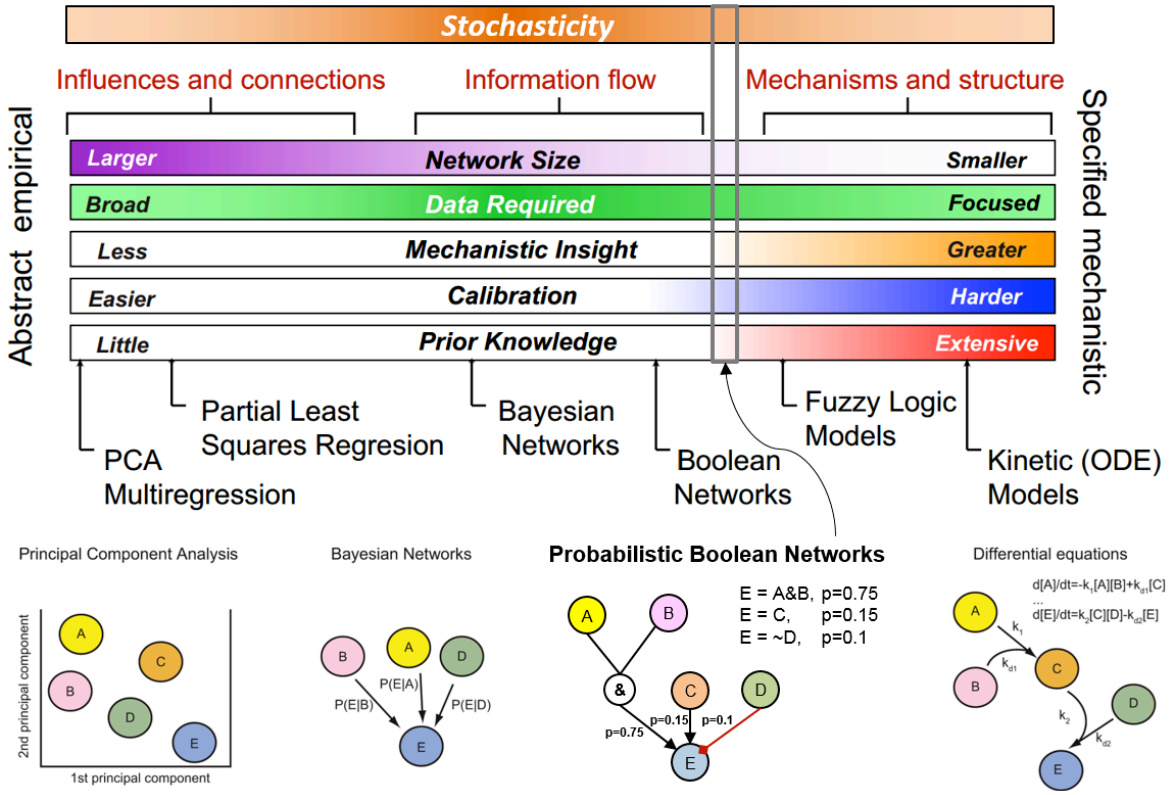


Figure 2. Mathematical models in biology and their characteristics. A large number of modelling approaches in statistics and systems biology were applied to model biological systems. Each of them has certain requirements and features with a trade-off on their advantages. (Adapted from "Spectrum of biological modelling method" by Peter Sorger and (Aldridge, Saez-Rodriguez, Muhlich, Sorger, & Lauffenburger, 2009))

According to the examples in Figure 2, one could classify the modelling approaches into two groups: statistical models and mechanistic models in systems biology. Statistical models such as principal component analysis (PCA) or partial least square regression only provide the information on the connections and influences between biological components without details on causal relationships nor on the mechanisms of regulation. Bayesian networks, a statistical model which could account for causal relationships in terms of conditional probabilities, were extensively used to model biological networks such as gene regulatory networks (Friedman, 2004) and signalling networks (Sachs, Perez, Pe'er, Lauffenburger, & Nolan, 2005). Nevertheless, this modelling approach is still unable to depict the regulatory mechanisms of biological processes which are naturally embedded within the systems.

In contrast to statistical models, the models in systems biology aim to account and represent the interconnections between biological components mechanistically with a set of rules, functions or mathematical equations. In this dissertation, I briefly review several modelling approaches in systems biology and their applications to biological systems followed by a comparison of their advantages and disadvantages.

1.3.1. Qualitative modelling approaches

Qualitative modelling refers to the modelling class which contains qualitative variables within the framework. In systems biology, there exist many qualitative modelling approaches including Boolean networks, Petri net, process algebra and constraint-based modelling. A short review of these approaches can be found in (Machado et al., 2011) and I will shortly introduce the first two approaches in this dissertation.

Boolean networks (BNs) were introduced by Kauffman as a model for gene regulatory networks (Kauffman, 1969). A Boolean model comprises a graph of molecules with binary states which are updated synchronously in each time step following a set of defined Boolean rules. Simple molecular mechanisms such as the formation of protein complex or the redundancy of protein functions can be represented by the Boolean logic gates 'AND' and 'OR', respectively. The steady-state properties of Boolean networks can be derived from attractors where they can also be correlated to the molecular states of biological entities in the network. As the description of this modelling framework is rather simple, multiple variants of Boolean networks were established, e.g. multi-state BN, asynchronous BN, and BN with perturbations, with an intent to capture more realistic biological phenomena (Albert, Thakar, Li, Zhang, & Albert, 2008; Chai et al., 2014; R.-S. Wang, Saadatpour, & Albert, 2012).

Petri nets were originally created to model and analyse concurrent systems (Petri, 1962). The network consists of a graph with two types of nodes, i.e. 'places' which hold 'tokens' connected by 'transitions' which 'fire' the execution. In biology, the concepts of place and token were correlated to the availability of substrates while the concept of transition and firing was correlated to the execution of biochemical reactions. Of note, the firing of transition is not deterministic hence stochasticity is inherited in this approach. Similar to BNs, there exist several Petri net extensions, e.g. coloured, timed, and functional, which allow for more detailed descriptions and analyses of the systems (Mura & Csikász-Nagy, 2008).

1.3.2. Quantitative modelling approaches

In contrast to qualitative models, quantitative modelling approaches take continuous variables into account. The most commonly used quantitative framework in systems biology is the group of differential equation approaches which describe the rates of changes over continuous variables. Therefore, these methods are suitable for the modelling of dynamical systems. Some examples of this framework are ordinary differential equation (ODE)-based models and partial differential equation (PDE)-based models.

ODE-based models contain a set of differential equations which describe the change of dependent variables as a function of one independent variable such as time. This framework can be applied to study many types of biological systems in order to depict detailed quantitative changes of molecular entities after perturbations over the course of an experiment (Chassagnole, Noisommit-Rizzi, Schmid, Mauch, & Reuss, 2002). Also, detailed mechanistic regulations of biochemical reactions can be accounted in this framework by integrating additional formula such as the Hill equation into differential equations to capture complex dynamical properties of biological phenomena.

Once quantitative changes of molecular entities need to be accounted based on more than one variable, PDE-based models which describe the changes of dependent variables as a function of multiple independent variables can be applied. An example of the application of a PDE-based model is the description of changing chemotherapeutic drug concentrations which are varied as a function of time and also as of the radial distance from the core tumour mass. More examples on the applications of PDE-based models in biology can be found in (Rangamani & Iyengar, 2007; Turing, 1952).

1.3.3. Applications of modelling in systems biology on signalling networks

Over recent years, both qualitative and quantitative models in systems biology were applied to study and analyse various biological systems including signal transduction networks. Boolean networks were applied to analyse, e.g. the cell cycle of budding yeast (F. Li, Long, Lu, Ouyang, & Tang, 2004), the signalling properties mediated by T-cell receptor (Saez-Rodriguez et al., 2007), and the crosstalk regulations of apoptotic network in hepatocytes (Schlatter et al., 2009) while Petri nets were applied to examine, e.g. the Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII) regulatory network (Hardy & Robillard, 2008) and also the execution of apoptotic pathways (C. Li, Ge, Nakata,

Matsuno, & Miyano, 2007). Coarse-grained mechanistic insights such as robustness and steady-state properties can be derived from these qualitative networks where some models, e.g. (Schlatter et al., 2009), also serve as a well-established starting topology for further modelling of detailed mechanistic networks (Schmich et al., 2011).

The applications of quantitative ODE-based models on signalling networks cover, e.g. the oscillation and variability in the p53 system (Geva-Zatorsky et al., 2006), the dynamic regulations of NFkB pathways via IkBs (Hoffmann, Levchenko, Scott, & Baltimore, 2002) and the time-course activation of MAPK pathway following EGFR stimulation (Schoeberl, Eichler-Jonsson, Gilles, & Müller, 2002). These models capture quantitative changes of molecular entities over time and thereby offer quantitative analyses and predictions. The mechanistic model of EGFR signalling was also subsequently used in pharmaceutical industry, e.g. to explore the mechanism of drug resistance in breast cancers via the PI3K/AKT/mTOR pathway after inhibition of ErbB2 (Schoeberl et al., 2009) which in turn lead to the discovery of novel therapeutic targets.

1.3.4. Advantages and disadvantages of applying different modelling approaches

According to the characteristics and requirements of qualitative and quantitative models in systems biology as depicted in Figure 2, one could observe a trade-off among the size of the model, the challenge on model calibration, the amount of required prior knowledge and experimental data, and the amount of mechanistic insights which could be derived from the models. Qualitative models such as Boolean networks allows for describing large size networks with minimal calibration. For instance, the Boolean model of (Schlatter et al., 2009) comprising 86 nodes and 125 interaction functions could comprehensively cover a large array of molecular players involved in the apoptotic pathway. Also, this model allows for a very fast computation of the logical steady-state, thereby allowing to compare and predict the effect of combinatorial treatments on apoptosis. However, this approach failed to provide quantitative predictions such as the proportion of apoptotic cells in the different experimental settings because of its inherited qualitative nature of state and time.

On the other hand, quantitative approaches such as ODE-based models grant detailed mechanistic insights from the network such as the precise prediction of molecular concentrations beyond the observed time-course or the sensitivity of molecules upon quantitative perturbations. Nevertheless, ODE-based models require prior knowledge on

biochemical reactions to determine the suitable formula, i.e. the rate laws of differential equations (Machado et al., 2011). In parallel, an extensive set of experimental data is required in order to estimate the unknown kinetic parameters where one could also encounter problems with unidentifiability in case the available datasets are not rich enough. These constraints therefore limit the scale of dynamical models in practice and it is consequently applicable to investigate only small subsets of reactions in the pathways (Blair, Trichler, & Gaille, 2012).

Given the advantages and disadvantages of each modelling approach, one always has to balance between the simplicity of the frameworks versus the expressiveness that the models could offer (Machado et al., 2011). Several factors should be considered, e.g. the research objective, the scope of the model, and the type and availability of data. In certain scenario, one might need to perform a quantitative mechanistic analysis on a large-scale biological network based on a limited amount of dataset, e.g. steady-state data where neither traditional qualitative nor quantitative models are suitable to be applied. With this regard, a class of quantitative mechanistic modelling approach with minimal parameterisation such as probabilistic Boolean networks (PBNs) could be a promising approach in such experimental setting.

1.4. Probabilistic Boolean networks

PBNs were introduced by Shmulevich *et al.* to model gene regulatory networks in the context of uncertainty (Shmulevich, Dougherty, Kim, & Zhang, 2002). To provide a comprehensive overview on the PBN approach, we reviewed the recent development and biomedical applications of PBNs from 2002 to the middle of 2013 (Trairatphisan et al., 2013). The review article covers a broad range of topics including the dynamic, inference, and intervention on PBNs. The relationship between PBNs and dynamic Bayesian networks (DBNs) were also discussed. Furthermore, the applications of PBNs on biological and physiological networks were collected and summarised.

On the study of signal transduction networks in the PBN framework, the steady-state distribution of molecular states in PBN can be correlated to the steady-state proteomic profiles generated from wet-lab experiments. Also, the relevance of interactions within the signalling networks can be accounted in terms of selection probabilities. These features make PBN ideal for describing the properties of signal transduction networks at steady-state with some uncertainty on network topologies derived from literature.

Trairatphisan *et al. Cell Communication and Signaling* 2013, **11**:46
http://www.biosignaling.com/content/11/1/46

CELL COMMUNICATION & SIGNALING

REVIEW Open Access

Recent development and biomedical applications of probabilistic Boolean networks

Panuwat Trairatphisan[1*], Andrzej Mizera[2], Jun Pang[2], Alexandru Adrian Tantar[2,4], Jochen Schneider[3,5] and Thomas Sauter[1]

Abstract

Probabilistic Boolean network (PBN) modelling is a semi-quantitative approach widely used for the study of the topology and dynamic aspects of biological systems. The combined use of rule-based representation and probability makes PBN appealing for large-scale modelling of biological networks where degrees of uncertainty need to be considered.

A considerable expansion of our knowledge in the field of theoretical research on PBN can be observed over the past few years, with a focus on network inference, network intervention and control. With respect to areas of applications, PBN is mainly used for the study of gene regulatory networks though with an increasing emergence in signal transduction, metabolic, and also physiological networks. At the same time, a number of computational tools, facilitating the modelling and analysis of PBNs, are continuously developed.

A concise yet comprehensive review of the state-of-the-art on PBN modelling is offered in this article, including a comparative discussion on PBN versus similar models with respect to concepts and biomedical applications. Due to their many advantages, we consider PBN to stand as a suitable modelling framework for the description and analysis of complex biological systems, ranging from molecular to physiological levels.

Keywords: Probabilistic Boolean networks, Probabilistic graphical models, Qualitative modelling, Systems biology

Background

A large number of formal representation types that exist in Systems Biology are used to construct distinctive mathematical models, each with their own strengths and weaknesses. On one hand, deciphering the complexity of biological systems by quantitative methods, such as ordinary differential equation (ODE) based mathematical models, yields detailed representations with high predictive power. Such an approach is however often hampered by the low availability and/or identifiability of kinetic parameters and experimental data [1]. These limitations often result in the generation of relatively small quantitative network models. On the other hand, qualitative modelling frameworks such as the Boolean Networks (BNs), allow for describing large biological networks while still preserving important properties of the systems [2]. The models pertaining to this latter class fail nevertheless to offer a quantitative determination of the system's dynamics due to their inherent qualitative nature.

Probabilistic Boolean networks (PBNs) were introduced in 2002 by Shmulevich et al. as an extension of the Boolean Network concept and as an alternative for modelling gene regulatory networks [3]. PBNs combine the rule-based modelling of a BN, as introduced by Kauffman [4-7], with uncertainty principles, e.g., as described by a Markov chain [8]. In terms of applications, analogously to the case of traditional BNs, the qualitative nature of state and time in a PBN framework allows for modelling of large-scale networks. The integrated stochastic properties of PBNs additionally enable semi-quantitative properties to be extracted. Existing analytic methods on PBNs allow for gaining a better understanding of how biological systems behave, and offer in addition the means to compare to traditional BNs. Examples are the calculation of influences which represent the quantitative strength of interaction between certain genes [3], or the determination of steady-state

*Correspondence: panuwat.trairatphisan@uni.lu
[1] Life Sciences Research Unit,University of Luxembourg, Luxembourg
Full list of author information is available at the end of the article

distributions to quantitatively predict the activity of certain genes in steady state [8].

It has been shown in the past years that the use of PBNs in the biological field is not limited to the molecular level, but also can potentially be linked to applications in clinic. To name a few, Tay et al. constructed a PBN to demonstrate the interplay between dengue virus and different cytokines which mediate the course of disease in dengue haemorrhagic fever (DHF) [9]. Ma et al. processed functional Magnetic Resonance Imaging (fMRI) signals to infer a brain connectivity network comparing between Parkinson's disease patients and healthy subjects [10]. Even though the research efforts on PBNs in this direction are just sprouting, the results from such PBN studies can provide a first clue on a disease's etiology and progression. As PBNs are highly flexible for data integration and as there exist a number of computational tools for PBN analysis, PBN is a suitable modelling approach to integrate information and derive knowledge from omic scale data which should in turn facilitate a physician's decision-making process in clinic.

For the past decade, PBNs were the object of extensive studies, both theoretical and applied. Among theoretical topics, there are steady-state distribution, e.g., [11-13], network construction and inference, e.g., [14-16], network intervention and control, e.g., [17-19]. Several minor topics were investigated as well, including reachability analysis [20] or sensitivity analysis [21]. Other studies dealt with PBNs in biological systems at multi-level such as gene regulatory networks [22-24], signal transduction networks [25], metabolic networks [26], and also physiological networks [9,10] which could potentially link to medicine as previously mentioned. In parallel, a number of computational tools which facilitate the modelling and analysis of PBNs are also continuously developed [27-29]. Given the continuous development in this area due to the broad on-going range of research on PBNs, we offer a state-of-the-art overview on this modelling framework. A comparison of PBN to other graphical probabilistic modelling approaches is also enclosed, specifically with respect to Bayesian networks. Last but not least, a view of the theoretical and applied research on PBNs as models for the study of multi-level biomedical networks is included.

In order to provide a coherent overview of the recent advances on PBN, we start with several theoretical aspects, organised as follows: an introduction to PBNs and associated dynamics are given in Section 'Introduction to probabilistic Boolean networks and their dynamics', the construction and inference of PBNs as models for gene regulatory networks are presented in Section 'Construction and inference of PBNs as models of gene regulatory networks', structural intervention and external control are discussed in Section 'Structural intervention and control of PBNs', ending with the relationship between PBNs and other probabilistic graphical models in Section 'Relationship between PBNs and other probabilistic graphical models'. Later, in Section 'PBN applications in biological and biomedical studies' we present a broad summary of PBN applications as a representation of biological networks followed by a discussion on the future applications of PBN in Systems Biology and Systems Biomedicine. A short conclusion is given in Section 'Conclusion'.

Introduction to probabilistic Boolean networks and their dynamics

Boolean networks

A *Boolean Network* (BN) $G(V, F)$, as originally introduced by Kauffman [4-7], is defined as a set of binary-valued variables (nodes) $V = \{x_1, x_2, \ldots, x_n\}$ and a vector of Boolean functions $\boldsymbol{f} = (f_1, \ldots, f_n)$. At each updating epoch, referred to as time point t ($t = 0, 1, 2, \ldots$), the *state* of the network is defined by the vector $\boldsymbol{x}(t) = (x_1(t), x_2(t), \ldots, x_n(t))$, where $x_i(t)$ is the value of variable x_i at time t, i.e., $x_i(t) \in \{0, 1\}$ ($i = 1, 2, \ldots, n$). For each variable x_i there exists a *predictor set* $\{x_{i_1}, x_{i_2}, \ldots, x_{i_{k(i)}}\}$ and a Boolean *predictor function* (or simply *predictor*) f_i being the i-th element of $\boldsymbol{f}$ that determines the value of x_i at the next time point, i.e.,

$$x_i(t+1) = f_i(x_{i_1}(t), x_{i_2}(t), \ldots, x_{i_{k(i)}}(t)), \tag{1}$$

where $1 \leq i_1 < i_2 < \cdots < i_{k(i)} \leq n$. Since the predictor functions of $\boldsymbol{f}$ are time-homogenous, the notation can be simplified by writing $f_i(x_{i_1}, x_{i_2}, \ldots, x_{i_{k(i)}})$. Without loss of generality, $k(i)$ can be defined to be a constant equal to n for all i by introducing *fictitious* variables in each function: the variable x_i is fictitious for a function f if $f(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n) = f(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n)$ for all possible values of $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$. A variable that is not fictitious is referred to as *essential*. The $k(i)$ elements of the predictor set $\{x_{i_1}, x_{i_2}, \ldots, x_{i_{k(i)}}\}$ are referred to as the *essential predictors* of variable x_i. The vector $\boldsymbol{f}$ of predictor functions constitutes the *network transition function* (or simply the *network function*). The network function $\boldsymbol{f}$ determines the time evolution of the states of the Boolean network, i.e., $\boldsymbol{x}(t+1) = \boldsymbol{f}(\boldsymbol{x}(t))$. Thus, the BN's dynamics is deterministic. The only potential uncertainty is in the selection of the initial starting state of the network.

Given an initial state, within a finite number of steps, the BN will transition into a fixed state or a set of states through which it will repeatedly cycle forever. In the first case, each such fixed state is called a *singleton attractor*, whereas in the second case, the set of states is referred to as a *cyclic attractor*. An *attractor* is either a singleton or

a cyclic attractor. The number of transitions required to return to a given state in an attractor is the *cycle length* of that attractor. The *attractor structure* of the BN is determined by the particular combination of singleton and cyclic attractors, and by the cycle lengths of the cyclic attractors. The states within an attractor are called *attractor states*. Non-attractor states are called *transient* and are visited at most once on any network trajectory. The states that lead into an attractor constitute its *basin of attraction*. The basins form a partition of the state space of the BN. For example, in Figure 1 the state transition diagrams of four different Boolean networks with three variables are given (in fact all these Boolean networks constitute a probabilistic Boolean network — the framework of probabilistic Boolean networks is presented in Section '5'). For each of these networks attractor states and transient states are indicated and the cyclic- and singleton attractors are given.

A *Boolean Network with perturbations* (BNp) is a BN with an introduced positive probability for which, at any transition, the network can depart from its current trajectory into a randomly chosen state, which becomes an initial state of a new trajectory. Formally, the perturbation mechanism is modelled by introducing a parameter p, $0 < p < 1$, and a so-called *perturbation vector* $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_n)$, where $\gamma_1, \gamma_2, \ldots, \gamma_n$ are independent and identically distributed (i.i.d.) binary-valued random variables [a] such that $\Pr\{\gamma_i = 1\} = \mathrm{p}$, and $\Pr\{\gamma_i = 0\} = 1 - \mathrm{p}$, for all $i = 1, 2, \ldots, n$. For every transition step of the network a new realisation of the perturbation vector is given. If $\boldsymbol{x}(t) \in \{0,1\}^n$ is the state of the network at time t, then the next state $\boldsymbol{x}(t+1)$ is given by either $\boldsymbol{f}(\boldsymbol{x}(t))$ or by

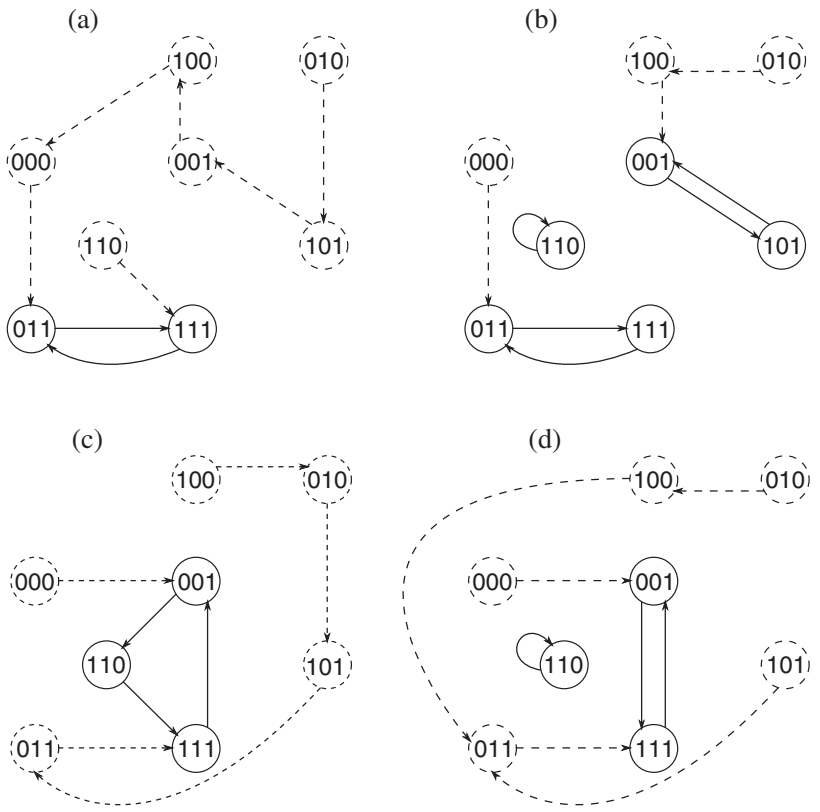


Figure 1 State transition diagrams of the four constituent Boolean networks of the PBN in Figure 2. For each constituent BN the attractor states and the transitions between them are indicated with solid circles and arrows, respectively. The remaining transitions and transient states are indicated with dashed arrows and circles, respectively. **(a)** The constituent BN of the PBN in Figure 2corresponding to transition function $\boldsymbol{f}_1$. There is only one attractor, i.e., {011, 111}, which is a cyclic attractor. **(b)** The constituent BN of the PBN in Figure 2 corresponding to transition function $\boldsymbol{f}_2$. There are two cyclic attractors: {011, 111}, {001, 101} and one singleton attractor: {110}. **(c)** The constituent BN of the PBN in Figure 2 corresponding to transition function $\boldsymbol{f}_3$. {001, 110, 111} is the cyclic attractor. **(d)** The constituent BN of the PBN in Figure 2corresponding to transition function $\boldsymbol{f}_4$. There are two attractors: a cyclic one, i.e., {001, 111} and a singleton one, i.e., {110}.

$\boldsymbol{x}(t) \oplus \boldsymbol{\gamma}(t)$, where $\oplus$ is component-wise addition modulo 2 and $\boldsymbol{\gamma}(t) \in \{0,1\}^n$ is the realisation of the perturbation vector for the current transition. The choice of the state transition rule depends on the current realisation of the perturbation vector. Two cases are distinguished: either $\boldsymbol{\gamma}(t) = \mathbf{0}$ or at least one component of $\boldsymbol{\gamma}(t)$ is 1, i.e., $\boldsymbol{\gamma}(t) \neq \mathbf{0}$. In the first case, which happens with probability $(1-p)^n$, the next state is given by $\boldsymbol{f}(\boldsymbol{x}(t))$. In the second case, given with probability $1-(1-p)^n$, the next state is determined as $\boldsymbol{x}(t) \oplus \boldsymbol{\gamma}(t)$: if $\gamma_i = 1$, then x_i changes its value; otherwise it does not ($i = 1,2,\ldots,n$). Since $\boldsymbol{\gamma}(t) \neq \mathbf{0}$, at least one of the nodes flips its value.

The attractors of a Boolean network characterise its long-run behaviour [8]. However, if random perturbations are incorporated, the network can escape the attractors. In particular, perturbations allow the system to reach any of its states from any current state in one transition. In consequence, the dynamics of the BNp is given by an *ergodic* Markov chain [30], [b] having a unique stationary distribution which simultaneously is its steady-state (limiting) distribution. The steady-state probability distribution, where each state is assigned a non-zero probability, characterises the long-run behaviour of the BNp. Nevertheless, if perturbation probability is very small, the network will remain in the attractors of the original network for most of the time, meaning that attractor states will carry most of the steady-state probability mass [8]. In this way the attractor states remain significant for the description of the long-run behaviour of a Boolean network after adding perturbations. Thus, a BNp inherits the attractor-basin structure from the original BN; however, once an attractor has been reached, the network remains in it until a perturbation occurs that throws the network out of it [31].

Probabilistic Boolean networks

PBNs were introduced in order to overcome the deterministic rigidity of BNs [3,32,33], originally as a model for gene regulatory networks. A PBN consists of a finite collection of BNs, each defined by a fixed network function, and a probability distribution that governs the switching between these BNs.

Formally, a probabilistic Boolean network $G(V,\mathcal{F})$ is defined by a set of binary-valued variables (nodes)[c] $V = \{x_1, x_2, \ldots, x_n\}$ and a list of sets $\mathcal{F} = (F_1, F_2, \ldots, F_n)$. For $i = 1,2,\ldots,n$ the set F_i is given as $\{f_1^{(i)}, f_2^{(i)}, \ldots, f_{l(i)}^{(i)}\}$, where $f_j^{(i)}$, $1 \leq j \leq l(i)$, is a possible Boolean predictor function for the variable x_i, with $l(i)$ the number of possible predictors for x_i. In general, each node x_i can have $l(i)$ different sets of essential predictors, each specified for a particular predictor function in F_i. A *realisation* of the PBN at a given instant of time is determined by a vector of predictor functions, where the ith element of that vector contains the function selected at that time point for x_i. For a PBN with N realisations there are N possible network transition functions $\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_N$ of the form $\boldsymbol{f}_l = (f_{l_1}^{(1)}, f_{l_2}^{(2)}, \ldots, f_{l_n}^{(n)})$, $l = 1,2,\ldots,N$, $1 \leq l_j \leq l(j)$, $f_{l_j}^{(j)} \in F_j$, and $j = 1,2,\ldots,n$. Each network function $\boldsymbol{f}_l$ defines a constituent Boolean network, or *context*, of the PBN.

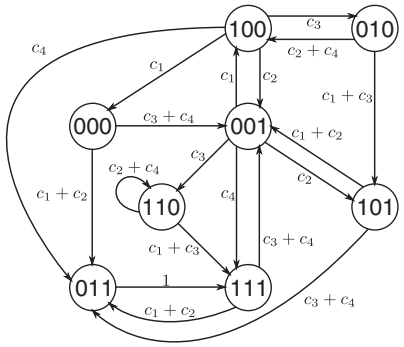
Let $\boldsymbol{f} = (f^{(1)}, f^{(2)}, \ldots, f^{(n)})$ be a random vector taking values in $F_1 \times F_2 \times \cdots \times F_n$; in other words, $\boldsymbol{f}$ is a random vector that acquires as value any of the realisations of the PBN. The probability that the predictor $f_j^{(i)}$, $1 \leq j \leq l(i)$, is selected to determine the value of x_i is given by

$$c_j^{(i)} = \Pr\{f^{(i)} = f_j^{(i)}\} = \sum_{l: f_{l_i}^{(i)} = f_j^{(i)}} \Pr\{\boldsymbol{f} = \boldsymbol{f}_l\}. \qquad (2)$$

It follows that $\sum_{j=1}^{l(i)} c_j^{(i)} = 1$. The PBN is said to be *independent* if the random variables $f^{(1)}, f^{(2)}, \ldots, f^{(n)}$ are independent. Assuming independence, there are $N = \prod_{i=1}^{n} l(i)$ realisations (constituent BNs) of the PBN and the probability distribution on $\boldsymbol{f}$ governing the selection of a particular realisation is given by $\Pr\{\boldsymbol{f} = \boldsymbol{f}_l\} = \prod_{i=1}^{n} c_{l_i}^{(i)}$. An example of a PBN with three nodes is given in Figure2.

At each time point of the PBN's evolution, a decision is made whether to switch the constituent network. This is modelled with a binary random variable ξ: if $\xi = 0$, then the current constituent network is preserved; if $\xi = 1$, then a context is randomly selected from all the constituent networks in accordance with the probability distribution of $\boldsymbol{f}$. Notice that this definition implies that there are two mutually exclusive ways in which the context may remain unchanged: 1) either $\xi = 0$ or 2) $\xi = 1$ and the current network is reselected. The *functional switching* probability $q = \Pr(\xi = 1)$ is a system parameter. Two cases are distinguished in the literature: if $q = 1$, then a switch is made at each updating epoch; if $q < 1$, then the PBN's evolution in consecutive time points proceeds in accordance with a given constituent BN until the random variable ξ calls for a switch. If $q = 1$, as originally introduced in [32], the PBN is said to be *instantaneously random*; if $q < 1$, it is said to be *context-sensitive*. The former models uncertainty in model selection, the latter models the situation where the model is affected by latent variables outside the model [34]. As an example let us consider the PBN given in Figure 2. Let the PBN be instantaneously random, i.e., $q = 1$. The four constituent BNs associated with the four transition functions $\boldsymbol{f}_1$, $\boldsymbol{f}_2$, $\boldsymbol{f}_3$, and $\boldsymbol{f}_4$, are given in Figure 1. Further, let us assume that the initial state is the state 101 and that the consecutive realisations are $\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_4, \boldsymbol{f}_3, \boldsymbol{f}_2, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4, \boldsymbol{f}_4, \ldots$.

$x_1x_2x_3$	$f_1^{(1)}$	$f_1^{(2)}$	$f_2^{(2)}$	$f_1^{(3)}$	$f_2^{(3)}$
000	0	1	0	1	1
001	1	0	1	0	1
010	1	0	0	1	0
011	1	1	1	1	1
100	0	0	1	0	1
101	0	0	1	1	1
110	1	1	1	1	0
111	0	1	0	1	1
$c_j^{(i)}$	1	0.3	0.7	0.4	0.6



$$\mathbf{A} = \begin{bmatrix} 0 & c_3+c_4 & 0 & c_1+c_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & c_1 & c_2 & c_3 & c_4 \\ 0 & 0 & 0 & 0 & c_2+c_4 & c_1+c_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ c_1 & c_2 & c_3 & c_4 & 0 & 0 & 0 & 0 \\ 0 & c_1+c_2 & 0 & c_3+c_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & c_2+c_4 & c_1+c_3 \\ 0 & c_3+c_4 & 0 & c_1+c_2 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Figure 2 An example of truth table, state transition diagram, and transition probability matrix of a PBN. The truth table, the state transition diagram, and the transition probability matrix **A** of a PBN without perturbations consisting of three variables $V = \{x_1, x_2, x_3\}$ and $\boldsymbol{\mathcal{F}} = (F_1, F_2, F_3)$, where $F_1 = \{f_1^{(1)}\}$, $F_2 = \{f_1^{(2)}, f_2^{(2)}\}$, and $F_3 = \{f_1^{(3)}, f_2^{(3)}\}$. Since there is one predictor function for node x_1 and two predictors for nodes x_2 and x_3, there are $1 \cdot 2 \cdot 2 = 4$ realisations of the PBN given by four network transition functions $\boldsymbol{f}_1 = (f_1^{(1)}, f_1^{(2)}, f_1^{(3)})$, $\boldsymbol{f}_2 = (f_1^{(1)}, f_1^{(2)}, f_2^{(3)})$, $\boldsymbol{f}_3 = (f_1^{(1)}, f_2^{(2)}, f_1^{(3)})$, and $\boldsymbol{f}_4 = (f_1^{(1)}, f_2^{(2)}, f_2^{(3)})$ with associated probabilities $c_1 = 0.12$, $c_2 = 0.18$, $c_3 = 0.28$, and $c_4 = 0.42$, respectively. For example, $c_3 = c_1^{(1)} \cdot c_2^{(2)} \cdot c_1^{(3)} = 1 \cdot 0.7 \cdot 0.4 = 0.28$. The edges in the state transition diagram are labelled with the transition probabilities. As can be seen from the state transition diagram, the underlying Markov chain is irreducible and aperiodic, thus ergodic. The steady-state (limiting) distribution for the chosen c_i values, $i = 1..4$, is given by $[\frac{7}{1609}, \frac{3640}{14481}, \frac{49}{4827}, \frac{716}{4827}, \frac{175}{4827}, \frac{238}{4827}, \frac{2548}{14481}, \frac{4696}{14481}]$ (the states are considered in the lexicographical order from 000 to 111).

Then, the corresponding time evolution of the PBN (trajectory) is given by the following sequence of state transitions: $101 \rightarrow 001 \rightarrow 110 \rightarrow 110 \rightarrow 111 \rightarrow 011 \rightarrow 111 \rightarrow 001 \rightarrow 100 \rightarrow 011 \rightarrow \ldots$. Irrespective of which constituent network (realisation) is selected next, the consecutive state in the trajectory is going to be 111 as the probability of moving from 011 to 111 is $c_1+c_2+c_3+c_4=1$.

A *Probabilistic Boolean Network with perturbations* (PBNp) is the variant of the PBN framework in which each constituent network is a BNp with a common perturbation probability parameter p, $0 < p < 1$, and a perturbation vector $\boldsymbol{\gamma}$. If $\boldsymbol{x}(t) \in \{0,1\}^n$ is the current state of the network and $\boldsymbol{\gamma}(t) = \mathbf{0}$, then the next state of the network is determined according to the current network function $\boldsymbol{f}_l$, i.e., $\boldsymbol{x}(t+1) = \boldsymbol{f}_l(\boldsymbol{x}(t))$. If $\boldsymbol{x}(t) \in \{0,1\}^n$ is the current state and $\boldsymbol{\gamma}(t) \neq \mathbf{0}$, then $\boldsymbol{x}(t+1) = \boldsymbol{x}(t) \oplus \boldsymbol{\gamma}(t)$. Whereas a context switch in a PBNp corresponds to a change in latent variables, resulting in a structural change in the functions that govern the PBNp, a random perturbation reflects a transient value change that leaves the network wiring unmodified, as for example in the case of gene activation or inactivation caused by external stimuli such as stress conditions or small molecule inhibitors [8].

The relationship between the four frameworks, i.e., Boolean networks, Boolean networks with perturbations, probabilistic Boolean networks, and probabilistic Boolean networks with perturbations is schematically depicted in Figure 3.

Dynamics of PBNs

A Boolean network with perturbations can be viewed as a homogenous irreducible Markov chain $\boldsymbol{X}_t$, with state space $\mathcal{X} = \{0,1\}^n$, where n is the number of nodes in the BNp. Let $\mathrm{P}_{\boldsymbol{y}}(\boldsymbol{x}) = \Pr[\boldsymbol{X}_{t_0+1} = \boldsymbol{x} | \boldsymbol{X}_{t_0} = \boldsymbol{y}]$ be the Markov chain transition probability from state $\boldsymbol{y}$ to state $\boldsymbol{x}$ at any instant t_0. This probability is a weighted sum of two transition probabilities, one for the BN, with probability $(1-p)^n$, and the other for the perturbations, with probability $1-(1-p)^n$, i.e.,

$$\mathrm{P}_{\boldsymbol{y}}(\boldsymbol{x}) = \mathbf{1}_{[\boldsymbol{f}(\boldsymbol{y})=\boldsymbol{x}]}(1-p)^n + (1-\mathbf{1}_{[\boldsymbol{x}=\boldsymbol{y}]})p^{\eta(\boldsymbol{x},\boldsymbol{y})}(1-p)^{n-\eta(\boldsymbol{x},\boldsymbol{y})}, \tag{3}$$

where p is the perturbation probability, $\mathbf{1}$ is the indicator function ($\mathbf{1}_{[P]} = 1$ if the proposition P is true, and $\mathbf{1}_{[P]} = 0$ otherwise), and $\eta(\boldsymbol{x}, \boldsymbol{y})$ is the Hamming distance between the binary vectors $\boldsymbol{x}$ and $\boldsymbol{y}$.

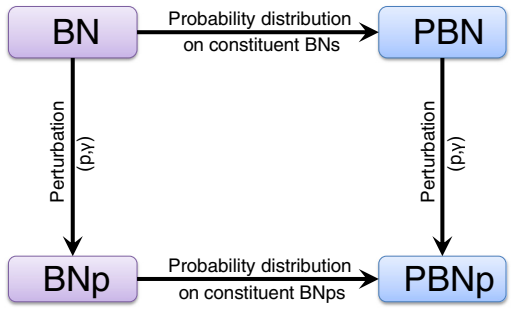


Figure 3 Relationships between the frameworks of Boolean and probabilistic Boolean networks. A Boolean network (BN) can be converted to a Boolean network with perturbations (BNp) by introducing a probability parameter p, $0 < p < 1$, and a perturbation vector ($\boldsymbol{\gamma}$). A probabilistic Boolean network (PBN) is built upon a number of constituent BNs and a probability distribution governing the choice of the Boolean network in accordance with which the next transition is made. Analogically, a PBN can be converted to a probabilistic Boolean network with perturbations (PBNp) by introducing a probability parameter p, $0 < p < 1$, and a perturbation vector ($\boldsymbol{\gamma}$). A probabilistic Boolean network (PBN) is built upon a number of constituent BNps and a probability distribution governing the choice of the BNp in accordance with which the next transition is made.

The Markov chain $\boldsymbol{X}_t$ is *ergodic*, which follows from the fact that it is aperiodic, irreducible, and defined on a finite state space. In other words, it possesses a unique stationary distribution, being simultaneously its steady-state (limiting) distribution. If $\mathrm{P}_{\boldsymbol{y}}^{(t)}(\boldsymbol{x})$ is the probability that the system transitions from $\boldsymbol{y}$ to $\boldsymbol{x}$ in t time steps, i.e., $\mathrm{P}_{\boldsymbol{y}}^{(t)}(\boldsymbol{x}) = \Pr[\boldsymbol{X}_{t_0+t} = \boldsymbol{x} | \boldsymbol{X}_{t_0} = \boldsymbol{y}]$, then the steady-state distribution π of $\boldsymbol{X}_t$ is defined by $\pi(\boldsymbol{x}) = \lim_{t\to\infty} \mathrm{P}_{\boldsymbol{k}}^{(t)}(\boldsymbol{x})$ for any initial state $\boldsymbol{k} \in \mathcal{X}$. For a set of states $B \subseteq \mathcal{X}$ the steady-state probability is given by $\pi(B) = \sum_{\boldsymbol{x}\in B} \pi(\boldsymbol{x}) = \lim_{t\to\infty} \mathrm{P}_{\boldsymbol{k}}^{(t)}(B)$ for any initial state $\boldsymbol{k} \in \mathcal{X}$. For example, the steady-state distribution of the Markov chain given by the transition probability matrix in Figure 2 is $[\frac{7}{1609}, \frac{3640}{14481}, \frac{49}{4827}, \frac{716}{4827}, \frac{175}{4827}, \frac{238}{4827}, \frac{2548}{14481}, \frac{4696}{14481}]$ (the states are considered in the lexicographical order from 000 to 111).

In the case of a probabilistic Boolean network, the transition probabilities $\mathrm{P}_{\boldsymbol{y}}(\boldsymbol{x})$ of the underlying Markov chain $\boldsymbol{X}_t$ depend on the probability of selecting a network transition function $\boldsymbol{f}_k$, $k = 1, 2, \ldots, N$, that determines the transition from $\boldsymbol{y}$ to $\boldsymbol{x}$ i.e.,

$$\mathrm{P}_{\boldsymbol{y}}(\boldsymbol{x}) = \Pr[\boldsymbol{X}_{t+1} = \boldsymbol{x} | \boldsymbol{X}_t = \boldsymbol{y}] = \sum_{k=1}^{N} \mathbf{1}_{[\boldsymbol{f}_k(\boldsymbol{y})=\boldsymbol{x}]} \cdot \Pr\{\boldsymbol{f} = \boldsymbol{f}_k\}, \tag{4}$$

where N, as before, is the number of constituent BNs and $\boldsymbol{f}$ is a random vector determining the PBN's realisation. Letting $\boldsymbol{x}$ and $\boldsymbol{y}$ range all states in $\mathcal{X}$, the transition probability matrix $\mathbf{A}$ of size $2^n \times 2^n$ can be formed and expressed as

$$\mathbf{A} = \sum_{k=1}^{N} \mathbf{A}_k \cdot \Pr\{\boldsymbol{f} = \boldsymbol{f}_k\}, \tag{5}$$

where $\mathbf{A}_k$ is the transition matrix corresponding to the k-th constituent BN.

Now, adding perturbations with probability p makes the underlying finite-space Markov chain $\boldsymbol{X}_t$ of the PBNp aperiodic and irreducible, hence ergodic. This allows the network dynamics of a PBNp to be studied with the use of the rich theory of ergodic Markov chains [30]. In particular, in the case of instantaneously random PBNps, the transition probability matrix $\tilde{A}$ is given by

$$\widetilde{\mathbf{A}} = (1-p)^n \cdot \mathbf{A} + \widetilde{\mathbf{P}}, \tag{6}$$

where $\widetilde{\mathbf{P}}$ is the perturbation matrix of the form

$$\widetilde{\mathbf{P}}_{\boldsymbol{y},\boldsymbol{x}} = (1 - \mathbf{1}_{[\boldsymbol{x}=\boldsymbol{y}]}) p^{\eta(\boldsymbol{x},\boldsymbol{y})} (1-p)^{n-\eta(\boldsymbol{x},\boldsymbol{y})}, \tag{7}$$

where, as before, $\mathbf{1}$ is the indicator function and η is the Hamming distance. As in the case of BNps, the ergodicity of the underlying Markov chain ensures the existence of the unique stationary distribution being the limiting distribution of the chain.

By definition, the set of attractors of a PBN is the union of the sets of attractors of the constituent networks [8]. Notice that whereas in a BN two attractors cannot intersect, attractors from different contexts can intersect in the case of a PBN. Similarly as in the case of Boolean networks, attractors play a major role in the characterisation of the long-run behaviour of a probabilistic Boolean network. If, however, perturbations are incorporated, the long-run behaviour of the network is characterised by its steady-state distribution. Nevertheless, if both the switching and perturbation probabilities are very small, then the attractors still carry most of the steady-state probability mass [8]. From a biological point of view attractors of such networks are interesting as they can be given a clear biological interpretation: they can be used to model cellular states [31]. For example, in the context of gene regulatory networks, it is believed that attractors can be interpreted as cellular phenotypes [7,8]. Thus, the long-run behaviour of the network given by its steady-state probabilities is of a special interest. Specifically, the attractor steady-state probabilities, i.e., $\pi(A)$, where A is an attractor, are important. There are a number of approaches towards the determination and analysis of the steady-state distribution of a PBNp. We review them shortly.

First, one approach to the steady-state analysis is to construct the state transition matrix in some form or another and then apply some numerical methods, e.g., iterative,

decompositional or projection methods [35]. A transition matrix based approach in which the sparse transition matrix is constructed in an efficient way and the so-called power method, which is applied to compute the steady-state probability distribution, is proposed in [36]. Unfortunately, the size of the state space grows exponentially in the number of nodes (genes) and becomes prohibitive for matrix-based numerical analysis of larger networks [11]. In [12], an approximation method for computing the steady-state probability distribution of a PBNp is derived from the approach of [36]. This method neglects some constituent BNps with very small probabilities during the construction of the transition probability matrix. An error analysis is given to demonstrate the effectiveness of this approach. Further, in [13] and [37] a matrix perturbation method for computing the steady-state probability distribution of PBNps is proposed together with its approximation variant. The proposed methods make use of certain properties of the perturbation matrix, $\widetilde{\mathbf{P}}$.

Second, Markov chain Monte Carlo methods [38] represent a feasible alternative to numerical matrix-based methods for obtaining steady-state distributions. Given an ergodic Markov chain, a Monte Carlo simulation method has been proposed: the probability of being in state $\boldsymbol{x}$ in the long run can be estimated empirically by simulating the network for a sufficiently long time and by counting the percentage of time the chain spends in that state regardless of the starting state [8]. A set of examples of Monte Carlo simulations from the PBN example in Figure 2 is shown in Figure 4. However, the question that remains is how to judge whether the simulation time is sufficiently long? The key factor here is the convergence, which in the case of a PBNp is known to depend to a large extent on the perturbation probability p [11]. Several approaches for determining the number of iterations necessary to achieve convergence were developed. A typical class consists of methods based on the second-largest eigenvalue of the transitions probability matrix, but due to reasons already mentioned above, these approaches can be impractical for larger networks. Another method utilises the so-called *minorisation condition* for Markov chains [39] to provide *a priori* bounds on the number of iterations. However, the usefulness of this approach is also limited (see [11] for details). There exist a number of methods for empirically diagnosing convergence to the steady-state distribution [40,41]. In [11] two of them are considered: one, based on the Kolmogorov-Smirnov test, a nonparametric test for the equality of continuous, one-dimensional probability distributions, and, second, the approach proposed in [42] which reduces the study of convergence of the chain to the investigation of the convergence of a two-state Markov chain. For illustration of application of these approaches to PBNs, we refer to [11] where the joint steady-state probabilities of combinations between two genes in human glioma gene expression data set were analysed.

Finally, as shown in [31], analytical expressions for the attractor steady-state probabilities can be derived both for BNps and PBNps. The obtained formulas are further exploited to propose an approximate steady-state computation algorithm.

We just shortly mention here that in the case of probabilistic Boolean networks without perturbations the dynamics is given by a Markov chain that does not necessarily be ergodic, specifically the Markov chain may contain more than one so-called *ergodic set of states*, also referred to as a closed, irreducible set of states in the literature. An ergodic set of states C in a Markov chain is defined as a set of states where all states communicate and no state outside C is reachable from any state in C[d]. The notion of an ergodic set of the corresponding Markov chain in probabilistic Boolean networks is the stochastic analogue of the notion of an attractor in standard Boolean networks [32]. Notice, however, that the ergodic sets and the attractors of a PBN or PBNp may differ. In the case of probabilistic Boolean networks without perturbations where the underlying Markov chain contains more than one ergodic set, considering the ergodic sets rather than the attractors may be more significant for understanding the long-run behaviour of the network. For example, in the context of modelling biological processes with PBNs, cellular phenotypes may in fact be represented by the ergodic sets. For more details see [32,43,44].

A number of other issues related to probabilistic Boolean network dynamics have been considered in the literature. We briefly list them here. In [45,46], the ordering of network switching and state transitions in context-sensitive PBNs are considered and its influence on the steady-state probability distributions is investigated. Algorithms for enumeration of attractors in probabilistic Boolean networks are discussed in [47]. Stability and stabilisation issues of PBNs are covered in [48]. Further, network transformations from one to another without losing some crucial properties, e.g., the steady-state probability distribution, are considered in [49]. For this purpose the concepts of homomorphisms and ϵ-homomorphisms for probabilistic regulatory networks, in particular PBNs, are developed.

Construction and inference of PBNs as models of gene regulatory networks

One approach to the dynamical modelling of gene regulation is based on the construction and analysis of network models. Generally, in the study of dynamical systems, long-run behaviour characteristics are of utter importance and their determination is a main aspect of system

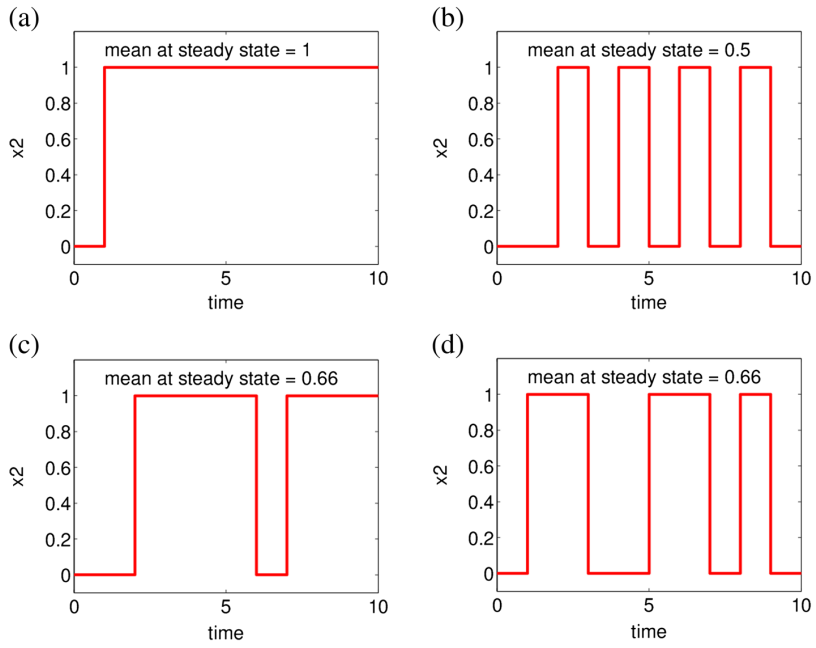


Figure 4 Dynamical simulations of node x_2 of the example network in Figure 2, with initial state $k = 000$. (a) Dynamics of x_2 governed by the constituent BN corresponding to the transition function $\boldsymbol{f}_1$, where $c_1 = 1, c_2 = c_3 = c_4 = 0$. Starting from 000 the periodic attractor $\{011, 111\}$ is reached. The probability of $\{x_2 = 1\}$ given by the stationary distribution is 1. **(b)** Dynamics of x_2 governed by the constituent BN corresponding to the transition function $\boldsymbol{f}_4$, where $c_4 = 1, c_1 = c_2 = c_3 = 0$. Starting from 000 the periodic attractor $\{001, 111\}$ is reached. The probability of $\{x_2 = 1\}$ given by the stationary distribution related to the reached attractor, i.e., $[0, \frac{1}{2}, 0, 0, 0, 0, 0, \frac{1}{2}]$ (the states are considered in the lexicographical order), is 0.5. **(c,d)** Examples of x_2 dynamics in the full PBN as given in Figure 2. Starting from 000 different trajectories are obtained for different simulation runs. The underlying Markov chain is ergodic and a unique stationary distribution, being the steady state (limiting) distribution, exists therefore. The steady state probability of $\{x_2 = 1\}$ is 0.66.

analysis. Reversely, the task of constructing a network possessing a specific set of properties is a subject of system synthesis. However, this *inverse problem* is usually ill-posed, i.e., there may be many models, or none, with the given properties [50]. Here we concentrate on the problem of inference from data in the framework of probabilistic Boolean networks, an inverse problem in which a network is constructed relative to some relationship with the available data. An outline of the workflow in network inference in the PBN framework is shown in Figure 5.

A data-driven approach for model construction consists of inferring the model structure and model parameters from measurement data, which in the case of gene regulation most commonly are gene expression measurements obtained with microarray technology. However, such data are continuous in nature. Thus, prior to the inference of Boolean or other discrete-type models (e.g., ternary) the measurements are usually discretised. The most common discretisation is binary (0 or 1) or ternary (usually -1, 0, 1) [8]. Discretisation is often justified as biological systems commonly exhibit switch-like on/off behaviour. Moreover, there are also a number of pragmatic reasons for quantising the measurements, e.g., it reduces the level of model complexity implying less computation and lower data requirements for model identification, provides a certain level of robustness to noise in the data, and has been shown to substantially reduce error rates in microarray-based classification [8,51-53]. A number of methods for discretisation of gene expression data exist, many of them having their origin in signal processing. One approach to quantisation was proposed in [54]: given some thresholds $\tau_1 < \tau_2 < \ldots$ (e.g., corresponding to limiting cases of a sigmoidal response), a multilevel discrete variable $\mathbf{x}$ is defined as $\mathbf{x} = \varphi(x) = r_k$ for $\tau_k < x \leq \tau_{k+1}$. As mentioned in [8], the thresholds can either come from prior knowledge or be chosen automatically from the data. In fact, there are various ways for optimal selection of the thresholds τ_k. One of the most popular methods is the Lloyd-Max quantizer, which amounts to minimising a so-called mean square quantisation error, see [55] for details. Approaches specific to binarising gene expression data can be found in [56-58]. Recently, Hopfensitz et al. [58] proposed a

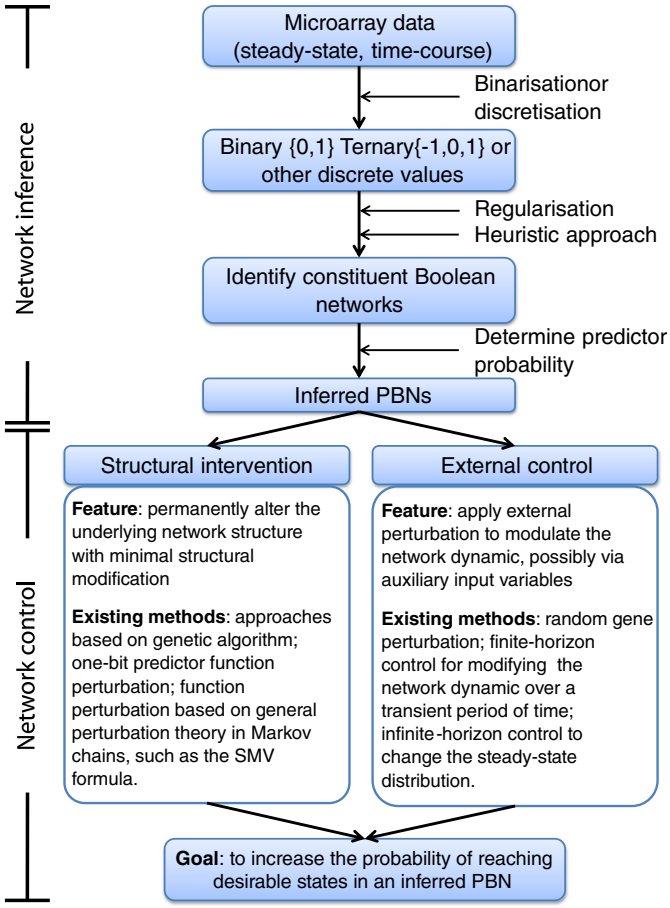


Figure 5 An outline of the workflow in network inference and control in the PBN framework. Microarray data, either from steady-state or time-course measurements, are typically binarised or discretised into discrete values. A heuristic approach, such as using genetic algorithms, is generally applied to identify constituent Boolean networks of the inferred PBN. Regularisation methods can be further applied to improve the accuracy of the inference with use of prior information on the network structure or dynamical rules. A number of well-established methods are subsequently applied to determine the predictor probability of each constituent Boolean network, thus the PBN is inferred. The inferred PBN can subsequently be perturbed with the methods on structural intervention or external control. The goal of network control is to increase the probability of reaching desirable states in the corresponding PBN.

new approach to binarisation which incorporates measurements at multiple resolutions. The method, called Binarization across Multiple Scales, is based on the computation of a series of step functions, detection of the strongest discontinuity in each step function and the estimation of the location and variation of the strongest discontinuities. Two variants of the method are proposed which differ in the approach towards the calculation of the series of step functions. The proposed method allows thresholds determination even with limited number of samples and simultaneously provides a measure of threshold validity – the latter can further be used to restrict network inference only to measurements yielding relevant thresholds. An example of application of binarisation to real data in the context of modelling with PBNs can be found in [10], where a brain connectivity network of Parkinson's disease is analysed. Binarisation is performed on fMRI real-valued data along the method recently proposed in [59].

One of the most straightforward inferential approaches is the *consistency problem* (also referred to as the *extension problem*), that entails a search for a rule from experimental data [8,60-62]. The problem amounts to finding in a specified class of Boolean functions one that complies with

two given sets of "true" and "false" Boolean vectors, i.e., a function that takes the value 1 for each of the "true" vectors and 0 for each of the "false" vectors.

In the case of real experimental data, a consistent extension may not exist either due to measurement noise or due to some underlying latent factors or other external influences not considered in the model [8]. In such case instead of searching for a consistent extension a Boolean function that minimises the number of misclassifications (errors) is considered. This problem is known as the *best-fit extension problem* [61] and is computationally more difficult than the consistency problem, since the latter is a special case of the former.

The application of PBN for modelling of large-scale networks is often impeded by limited sample sizes of experimental data. As mentioned in [63], main challenges in automated network reconstruction arise from the exponential growth of possible model topologies for increasing network size, the high level of variability in measured data often characterised by low signal to noise ratios, and the usually large number of different components that are measured versus relatively small number of different observations under changing conditions, e.g., number of time points or perturbations of the biological system. Together these problems lead to non-identifiability and over-fitting of models [63]. In such cases any prior information on the network structure or dynamical rules is likely to improve the accuracy of the inference [8,64]. This information usually pertains to model complexity and is used to penalise excessively complex models. For this purpose, the so-called *regularisation methods* can be employed. The most popular regularisation assumption in gene regulatory modelling is that the inferred models should be sparse, i.e., the number of regulators acting on a gene is low [65-68] or that the node degree in biological networks is often power law distributed, with only few highly-connected genes, and most genes having small number of interaction partners [63,69]. Regularisation is a well-established inference approach in the framework of Bayesian networks (see, e.g., [63,70,71]) and can be also used in the framework of BNs and PBNs. For example, in the case of inference of Boolean networks, the so-called *sensitivity regularisation method* has been proposed [64]. Due to limited sets of data, the estimates of the errors of a given model in the best-fit extension problem, which themselves depend on the measurements, may be highly variable [64]. The regularisation is built on the observation that the expectation of the state transition error generally depends on a number of terms, among others the sensitivity deviation which is a difference in the sensitivities of the original and the inferred networks. In consequence, as argued in [64], the sensitivity deviation can be incorporated as an additional penalty term to the best-fit objective function, reflecting the hypothesis that the best inference should have a small error in both state transition and sensitivity.

In order to infer a PBN, strong candidates for regular Boolean networks need to be identified first. This can be performed with generic methods mentioned in [72] such as literature data compilation, the gene association networks approach [73,74] or by applying a heuristic approach, e.g., a genetic algorithm, which searches through the model space to find good candidates for the network structure with respect to a specified fitness function. Next, the candidates' predictor functions are combined into a set of network transition functions for the PBN. An example of PBN model selection using heuristics can be found in [75].

A common strategy for determining the predictor probabilities relies on the *coefficient of determination* (CoD) between target and predictor genes [8,32,72,76]. The CoD is a measure of relative decrease in error from estimating transcriptional levels of a target gene via the levels of its predictor genes rather than the best possible prediction in the absence of predictor genes [8]. The CoDs can be then translated to the predictor probabilities. However, as pointed out in [77], for each gene, the maximum number of possible predictors as well as the number of their corresponding probabilities is equal to 2^{2^n}, where n is the number of nodes. This implies that the number of parameters in the PBN model is $O(n2^{2^n})$[e]. Therefore, the applicability of the CoD approach is significantly limited due to the model complexity or imprecisions owing to insufficient data sample size. This hindrance is often surpassed by imposing some constraints on the maximum size of admissible predictors for each gene.

In [50] the authors consider the *attractor inverse problem*, that involves designing Boolean networks given attractor and connectivity information. Two algorithms for solving this problem are proposed. They are based on two assumptions on the biological reality: first, the biological stability, i.e., that most of the steady-state probability mass is concentrated in the attractors and, second, the biological tendency to stably occupy a given state, i.e., attractors are singleton attractor cycles consisting of a single state. The first algorithm operates directly on the truth table, while taking into account simultaneously the information on the attractors and predictor sets. There is however no control on the level-set structure. The second algorithm works on the state transition diagram that satisfies the design requirements on attractor and level-set structures and checks whether the associated truth table has predictor sets that agree with the design goals. The proposed algorithms can be further used in a procedure for designing PBN from data. In the approach described in [50], a collection of BNs is generated by the first algorithm, then some of the BNs are selected

based on the basin sizes criterion and combined in a PBN whose steady-state distribution closely matches the observed data frequency distribution. This design procedure has been applied to gene-expression profiles in a study of 31 malignant melanoma samples in [50].

An inverse PBN construction approach is also described in [78]. This work relies on expressing the probability transition matrix as a weighted sum of Boolean network matrices. A heuristic algorithm with $O(m2^n)$ complexity is proposed, where n, m stand for the number of genes, respectively the number of non-zero entries in the transition matrix. The authors also introduce an entropy based probabilistic extension, both algorithms being analysed against random transition matrices.

Usually, the optimal predictor for a gene will not be perfect as there will be inconsistencies in the data. In [79] it is proposed to model these inconsistencies in a way that mimics context changes in genomic regulation, with the intention to view data inconsistencies as caused by latent variables. The inference procedure of [79] results in PBNs whose contexts model the data in such a way that they are consistent within each context. The key criterion for network design is that the distribution of data states agrees with the distribution of expected long-term state observations for the system.

The probabilities of the system being in a particular context and the number of constituent networks are determined by the data. The approach of [79] can be seen as imposing a structure on a probabilistic Boolean network that resolves inconsistencies in the data arising from mixing of data from several contexts. It should be noted that in this approach the contexts are determined directly by the data, whereas in [32] and [80] constituent networks depend on the number of high-CoD predictor sets or high Bayes-score predictor sets, respectively, and these in turn depend on the designer's choice of a threshold. Moreover, the number of constituent networks is determined by how inconsistencies appear in the data, not the number of states appearing in the data (see [8] for an example). The contextual-design method of [79] has been applied to expression profiles for melanoma genetic network.

We just mention here that also information theoretic approaches were considered for inference of PBN from data. Probably the most widely studied methods are based on the minimum description length (MDL) principle [81]. Descriptions of inference algorithms that utilise this principle can be found, e.g., in [8,82,83].

The manner of inference depends on the kind of experimental data available. There are two cases: 1) time-series data and 2) steady-state data. We proceed with presenting them briefly.

Time-course measurements

It is assumed that the available data are a single temporal sequence of network states. In this case, given a sufficiently long sequence of observations, the goal is to infer a PBN that is one of plausible candidates to have generated the data. Usually, an inference procedure for this type of problem constructs a network that is to some extent consistent with the observed sequence.

In [84,85], the inference in case of context-sensitive PBNs with perturbations is considered, where the probability of switching from the current constituent Boolean network to a different one is assumed to be small. The proposed inference procedure consists of three main steps: first, identification of subsequences in the temporal data sequence that correspond to constituent Boolean networks with use of so-called 'purity functions'; second, determination of *essential predictors* for each subsequence by applying an inference procedure based on the transition counting matrix and a proposed cost function; finally, inference of perturbation, switching, and selection probabilities. However, the amount of temporal data needed for inference with this approach is huge, especially due to the perturbation and switching probabilities: if they are very small, then long periods of time are needed to escape attractors and if they are large, estimation accuracy is harmed. As stated in [85], if one does not wish to infer the perturbation, switching, and selection probabilities, then constituent-network connectivity can be discovered with decent accuracy for relatively small time-course sequences.

A more practical way of inferring PBN parameters from time-course measurements is presented in [77]. The authors propose a multivariate Markov chain model to infer the genetic network, develop techniques for estimating the model parameters and provide an efficient method of estimating PBN parameters from their multivariate Markov chain model. The proposed technique has been tested with synthetic data as well as applied to gene expression data of yeast.

Further, in [86] the problem of PBN context estimation from time-course data is considered. The inference is considered with respect to minimising both the conditional and unconditional mean-square error (MSE). The author proposes a novel state-space signal model for discrete-time Boolean dynamical systems, which includes as special cases distinct Boolean models, one of them being the PBN model. A Boolean Kalman Filter algorithm is employed to provide the optimal PBN context switching inference procedure in accordance to minimisation of MSE.

Steady-state data

Here we consider a long-run inverse problem in the context of probabilistic Boolean networks as models for

gene regulation. On one hand, in the case of microarray-based gene-expression studies it is often assumed that the data are obtained by sampling from a steady state. On the other hand, attractors represent the essential long-run behaviour of the modelled system [31]. Thus, in the modelling framework of Boolean networks it is expected that the observed data states are mostly the attractor states of a model network. In consequence, much of the steady-state distribution mass of the model network should lie in the states observed in the sample data [50,80,87]. In the case of Boolean networks with perturbations or probabilistic Boolean networks with perturbations, the underlying dynamical system is an ergodic Markov chain, hence possesses a steady-state distribution. However, by imposing some mild stability constraints that reflect biological state stability, also in these frameworks most of the steady-state probability mass is carried by the attractors [31].

There are however inherent limitations to the construction of dynamical systems from steady-state data. Although the steady-state behaviour restricts the network dynamics, it does not determine the steady-state behaviour: there may be a collection of compatible networks with a given attractor structure. In particular, it does not determine the Boolean network's basin structure. As a consequence, obtaining good inference relative to the attractor structure does not necessary entail valid inference with respect to the steady-state distribution as the steady-state probabilities of attractor states depend on the basin structure [50,80]. In fact building a dynamical model from steady-state data is a kind of over-fitting [88].

Although the CoD has been used for inference of PBNs from steady-state data in [32], a fundamental problem is that the CoD cannot provide information on the direction of prediction without time-course data. The resulting *bidirectional relationships* can affect the inferred graph topology by introducing spurious connections. Moreover, they can lead to inference of spurious attractor cycles that do not correspond to any biological state [8]. As a consequence, this suppressed the use of the CoD as a inference method for steady-state data.

The inference methods that replaced the CoD approach are primarily based on the attractor structure [50,79] or graph topology [89]. In the former case, the key concern is to infer an attractor structure close to that of the true network. In the latter case, the focus is on the agreement between graph connections, e.g., as measured by the Hamming distance between the regulatory graphs [8]. In [16], an approach that achieves both preservation of attractor structure and connectivity based on strong gene prediction has been proposed.

Another approach to the problem of constructing gene regulatory networks from expression data using the PBNs framework is proposed in [90]. The key element of this method is a non-linear regression technique based on reversible-jump Markov chain Monte Carlo (MCMC) annealing for predictor design. The network construction algorithm consists of the following stages. First, for each target gene x_i $(i = 1, 2, \ldots, n)$ in the network of n genes a collection of predictor sets is determined by applying a clustering technique based on mutual information minimisation. Optimisation [f] is performed with use of the simulated annealing procedure. This step reduces the class of different predictor functions available for each target gene. Next, each predictor set is used to model a predictor function $f_k^{(i)}$ by a perceptron consisting of both a linear and a nonlinear term, where $k = 1, 2, \ldots, l(i)$, with $l(i)$ the number of predictor sets found in the previous step for target gene x_i. A reversible MCMC technique is used to calculate the model order and the parameters. Finally, the CoD is used to compute the probability of selecting different predictors for each gene. For a detailed description of this algorithm and its application to data on transcription levels in the context of investigating responsiveness to genotoxic stresses see [90]. It should be noticed that the proposed reversible-jump MCMC model for predictor design extends the binary nature of PBNs allowing for a more general model containing non-Boolean predictor functions that operate on variables with any finite number of possible discrete values [72].

As an alternative to the technique of [90], a fully Bayesian approach (without the use of CoD) for constructing probabilistic gene regulatory networks, with an emphasis for network topology, is proposed in [80]. In this approach, the predictor sets of each target gene are computed, the corresponding predictors are determined, and the associated probabilities, based on the nonlinear perceptron model of [90], are calculated by relying on a reversible jump MCMC. Then, a MCMC method is used to search for the network configurations that maximise the Bayesian scores to construct the network. As stated in [8], this method produces models whose steady-state distribution contains attractors that are either identical or very similar to the states observed in the data. Moreover, many of the attractors are singleton attractors, which reflect the biological propensity to stably occupy a given state. The approach of [90] has been applied to gene-expression profiles resulting from the study of 31 malignant melanoma samples presented in [91].

In [92] the inverse problem of constructing instantaneously random PBNs from a given stationary distribution and a set of given Boolean networks is considered. Due to large size of this problem, it is formulated in terms of constrained least squares and a heuristic method based on Conjugate Gradient is proposed as a solution.

In [93], the inverse problem of PBNs with perturbations is considered, where a modified Newton method is

proposed for computing the perturbation probability p where the transition probability matrix $\tilde{\boldsymbol{A}}$ and the steady-state probability of the PBNp $\tilde{\boldsymbol{x}}$ are known. The new algorithm makes use of certain properties of the set of steady-state nonlinear equations, i.e., $\tilde{\boldsymbol{A}}\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}} = \boldsymbol{0}$, with p as the unknown variable. Considering these properties improves the computational efficiency with respect to a direct approach in which every of the 2^n equations (n being the number of nodes) is solved and common solutions are reported.

Structural intervention and control of PBNs

Using PBNs for the modelling and analysis of biological systems can lead to a deeper understanding of the dynamics and behaviour of these systems (see Section 'Dynamics of PBNs'), paving the way for different methods used for system structure inference and data measurement (see Section 'Construction and inference of PBNs as models of gene regulatory networks'). Another major objective of such studies is to predict the effect a perturbation or an intervention has on the system structure, e.g., allowing to identify potential targets for therapeutic intervention in diseases such as cancer. Intervention strategies in PBNs, e.g., as to change the long-run behavior of networks in order to decrease the probability of entering some undesired state, rely on two different kinds of direction – *structural intervention* [8,33] and *external control* [8,18]. While the first approach can alter the underlying network structure permanently, the second one uses external control to modulate the network dynamics. A classification of network control methods in the PBN framework is shown in Figure 5.

Structural intervention

The problem of performing a structural intervention in a PBN looks at how the steady-state probability of certain states can be changed with only minimal structural modifications [8,33]. A more formal description is offered in the following. Given a PBN and two subsets A and B of its states, the associated steady-state probabilities $\pi(A)$, $\pi(B)$, have to be modified such as to approach some given values λ_A, respectively λ_B. This can be achieved by replacing the predictor function f_{ik} (of gene i in context k) with a new function g_{ik}, while keeping all other network parameters unchanged. We denote the steady-state distribution of the resulting PBN as μ. Then, it is possible to interpret the problem as an optimisation one: given the state sets A, B, and two values $\lambda_A \geq 0$, $\lambda_B \geq 0$, with $\lambda_A + \lambda_B \leq 1$, find a context k, a gene i, and a function g_{ik} to replace f_{ik}, such as to minimises $\epsilon(A,B) = \mid \mu(A) - \lambda_A \mid + \mid \mu(B) - \lambda_B \mid$, with respect to all contexts, genes, and predictor functions. Note that A and B can be used to represent both desirable as well as undesirable states. While this approach allows changing one predictor function at a time, a generalisation can be made by allowing a number of predictor functions or by adding more constraints on the selected functions, only to give a few examples.

Shmulevich et al. [33] proposed using genetic algorithms to deal with the above optimisation problem. Later, Xiao and Dougherty [94] provided a constructive algorithm for structural intervention and applied it to a WNT5A network. The proposed algorithm focuses on the impact one-bit predictor function perturbations have on state transitions and attractors. Their approach, however, does not directly characterise the steady-state distribution changes that result from (structural) perturbations of a given probability. In order to solve this problem, Qian and Dougherty [95] derived a formal characterisation of optimal structural intervention, based on the general perturbation theory in finite Markov chains. Specifically, they gave an analytical solution for computing the perturbed steady-state distribution by looking at function perturbations. Their work mainly focused on one-bit function (or rank-1 matrix) perturbations, implying that for more general perturbations, one needs to consider an iterative approach. The associated complexity of such an approach is of $O(2^{3n})$, where n is the number of genes in the network. Their results have been applied to a WNT5A network and a mammalian cell cycle related network, respectively. More recently, Qian et al. [96] extended their previous result in [95] to a more efficient solution that uses the Sherman-Morrison-Woodury (SMW) formula [97] to deal with rank-k matrix perturbations. Thus, they managed to reduce the computational complexity of the approach from $O(2^{3n})$ to $O(k^3)$, where $k \ll 2^n$ (k is much smaller than 2^n). The application of the derived structural intervention method to a mutated mammalian cell cycle network shows that the intervention strategy can identify the main targets to stop uncontrolled cell growth in the network.

Qian and Dougherty [98] also looked at how long-run sensitivity analysis can be used in PBNs, in terms of difference between steady-state distributions before and after perturbation, and with respect to different elements of the network, e.g., probabilistic parameters, regulatory functions, etc.

External control

While structural intervention focuses on a permanent change in the network dynamics, *external control* relies on Markov decision processes theory for driving a network out of an undesired state, i.e. as to reach a more desirable one [8,18].

The first approach to deal with PBNs was proposed by Shmulevich et al. [18]. They studied the impact of random *gene perturbations* [g] on the long-run behavior of a

network. The main idea of Shmulevich et al. [18] is to construct a formulation of the state-transition probability that relies on the probability of a gene perturbation and on Boolean functions for finding bounds for the steady-state probability. Their particularly interesting finding is that these states (which in terms of mean first-passage times (MFPT) are easy to reach from other states) are more stable with respect to random gene perturbations. In gene regulatory networks, it is important to identify what genes are more likely to lead the network into a desirable state when perturbed. MFPT naturally captures this idea – a few other methods developed by Shmulevich et al. [18] work, for example, by maximising the probability to enter some particular state in some fixed maximum amount of time, or by minimising the time needed to reach that state.

Gene perturbation works by single flips of a gene's state, providing a natural platform for external intervention control via auxiliary input variables. It makes sense from a biological perspective, for example, to model auxiliary treatments in cancer such as radiation. The value of these variables can be thus chosen such as to make the probabilistic distribution vector of the PBN evolve in some desired manner.

More formally, given a PBN with n genes and k control inputs, $u_1, u_2, \ldots, u_k$, the vector $u(t) = (u_1(t), u_2(t), \ldots, u_k(t))$ is used to denote the values of all control inputs at a given time step t. Let P denote the transition probability matrix of the PBN, evolving according to $w(t+1) = w(t) \cdot P(u(t))$. It is obvious to see that, at each time step t, P depends not only on the initial probability distribution vector, but also on the values of the control inputs. External control is essentially about making the network evolve in some desired manner by choosing, at each time step, input control values. The sequence of control inputs, referred to as a *control policy* or *strategy*, can be associated to a cost function which has to be minimised over the entire class of allowed policies. Such functions capture the cost and benefit of using interventions, and are normally application dependent. For the sake of simplicity, we use $J_\omega(z(0))$ to denote the cost with respect to a control policy ω and an initial state $z(0)$. Then, an *optimal PBN control problem* can be defined as a search for a control policy ω that minimises the cost $J_\omega(z(0))$. External control in PBNs can be classified into the following two groups.

Finite-horizon external control

The *finite-horizon external control* problem is about modifying over a transient period of time the network dynamics of some given PBN, without changing its steady-state distribution. In other words, external control is only applied over a finite number of M time steps, using policies of the form $\omega = (\mu_0, \mu_1, \ldots, \mu_{M-1})$. The first optimal finite control formulation in PBNs, and a solution based on Dynamic Programming [99], were given by Datta et al. [100]. Working assumptions implied known transition probabilities and horizon length, later removed in [101] by making use of measurements, thought to be related to the underlying Markov chain states of the PBN. Pal et al. [17] extended the results of Datta et al. [100,101] to context-sensitive PBNs with perturbation. The results have been used to devise a control strategy that reduces the WNT5A gene's action in affecting biological regulation.

Optimal finite-horizon dynamic programming based control, assuming a fixed number of time steps M and a fixed number of controls k, has a computational complexity of $O(2^{2^n})$, where n is the number of genes in the network. Namely, the problem is limited by the size of the network as one needs to compute the transition probability matrix. In particular, Akutsu et al. [102] proved that the problem is NP-hard.[h] Chen and Ching [103] used dynamic programming in conjunction with state reduction techniques [104,105] to find an optimal control policy for large PBNs. They managed to reduce the computation complexity to $O(|R|)$, where $|R|$ is the number of states after state reduction.

Kobayashi and Hiraishi [106] proposed an integer programming based approach that avoids computing the probability matrix in optimal finite-horizon control. Later, they extended their work to context-sensitive PBNs [107,108], focusing on the lower and upper bounds of the cost function. Furthermore, Kobayashi and Hiraishi [109] proposed a polynomial optimisation approach where a PBN is first transformed into a polynomial system, subsequently allowing to reduce the optimal control to a polynomial optimisation problem. In the above papers, only small examples are used to illustrate the proposed approaches.

Ching et al. [110] looked at hard constraints for an upper bound on the number of controls, and proposed a novel approach that requires minimising the distance between terminal and desirable states. They also gave a method to reduce the computational cost of the problem by using an approximation technique [12]. Cong et al. [111] made one step further by considering the case of multiple hard constraints, i.e., the maximum numbers of times each control method can be applied, developing an algorithm capable of finding all optimal control policies. A heuristic approach was developed by the same authors in order to deal with large size networks [111]. A different and more efficient algorithm, using integer linear programming with hard constraints, was presented later by Chen et al. [112]. The WNT5A network is a typical example used in [111,112].

Instead of minimising the cost, Liu et al. [113] investigated the problem of how control can be used to reach

desirable network states, with maximal probability and within a certain time. Later, Liu [19] imposed another new criterion for the optimal design of PBN control policies, namely the expected average time required to transform undesired states into desirable ones. In both papers, the optimal control problem can be solved by minimising the MFPT of discrete-time Markov decision processes.

The controllability problem of PBNs was studied by Li and Sun [114]. A semi-tensor product of matrices, as described in their work, allows to convert a probabilistic Boolean control network into a discrete time system. They provided some conditions for the controllability of PBNs via either open or closed loop control.

Infinite-horizon external control

Infinite-horizon external control implies working with external auxiliary variables, over an infinite period of time, the steady-state distribution being also changed. Policies in this case have the form of $\omega = (\mu_0, \mu_1, \ldots)$.

In the finite-horizon case, the optimal control policy is calculated by (essentially) using a backward dynamic programming algorithm, ending once the initial state is reached. However, this approach cannot be applied to infinite-horizon control directly due to the non-existence of a termination state in the finite-horizon case, potentially leading to an infinite total cost. Pal et al. [115] extended the earlier finite-horizon results to the infinite-horizon case for context-sensitive PBNs. They solved the above two problems by using the theory of average expected costs and expected discounted cost criteria in Markov decision processes. For applications, they considered a gene network containing the genes WNT5A, pirin, S100P, RET1, MART1, HADHB, and STC2.

A robust control policy can be found in Pal et al. [116], devised via a minimisation of the worst-case cost over the uncertainty set, with uncertainty defined with respect to the entries of the transition probability matrix.

Due to the computational complexity of $O(2^{2^n})$, several greedy algorithms have been proposed in the literature. Vahedi et al. [117] developed a greedy control policy that uses MFPT. Their main idea is to reduce the risk of entering undesirable states by increasing (or decreasing) the time needed to enter such a state (or, respectively a desirable state). Performance of the MFPT-based algorithm was studied on a few synthetic PBNs and a PBN obtained from a melanoma gene-expression dataset, where the abundance of messenger RNA for the gene WNT5A was found to be highly discriminating between cells with properties associated with high or low metastatic competence. Later, three different greedy control policies were proposed by Qian et al. [118], using the steady-state probability mass. The first one explores the structural information of a basin of attractors in order to reduce the steady-state probability mass for undesirable states, while the remaining two policies regard the shift in the steady-state probability mass of undesirable states as a criterion when applying control. The identified three policies, together with the one based on MFPT [117], were evaluated on a large number (around 1000) of randomly generated networks and a mammalian cell cycle network [119].

Some types of cancer therapies like chemotherapy, are given in cycles with each treatment being followed by a recovery period. Vahedi et al. [120] showed how an optimal cyclic control policy can be devised for PBNs. Yousefi et al. [121] extended the results in [120] to obtain optimal control policies for the class of cyclic therapeutic methods where interventions have a fixed-length duration of effectiveness. Both of the two approaches [120,121] were applied to derive optimal cyclic policies to control the behavior of regulatory models of the mammalian cell cycle network [119]. While the goal of control policies is to reduce the steady-state probability mass of undesirable states, in practice it is also important to limit collateral damage, to consider when designing control policies. Based on this observation, Qian and Dougherty [122] developed two new phenotypically-constrained control policies by investigating their effects on the long-run behaviour of the network. The newly proposed policies were examined on a reduced network of 10 nodes. The network was obtained from gene expression data collected for the study of metastatic melanoma (e.g, see [91]).

Relationship between PBNs and other probabilistic graphical models

Probabilistic graphical models, commonly applied in computational biology for network reconstruction, provide the means for representing complex joint distributions. Examples include PBNs, Bayesian networks and their variants, e.g., dynamic and hierarchical Bayesian networks, hidden Markov models, factor graphs, Markov random fields, conditional random fields, Markov logic networks, etc. In this section we discuss the relationship between the two of them which are usually employed to deal with system dynamics: the PBNs and the dynamic Bayesian networks, the latter generalising hidden Markov models.

A Bayesian network is essentially a graphical, compact representation of a joint probability distribution. The Bayesian network consists of two elements. First, a directed acyclic graph (DAG) where the vertices of the graph represent random variables and the directed edges or lack thereof encodes the so-called *Markovian assumption*, which states that each variable is independent of its non-descendants, given its parents [8,123]. Second, a set of local conditional probability distributions for each vertex, given its parents in the graph. By the chain rule of

probabilities, the joint probability distribution on the random variables in the graph can be decomposed into a product of the local conditional probabilities, i.e., if there are n random variables X_i, $i = 1, 2, \ldots, n$ and $\mathrm{Pa}(X_i)$ denotes the parents of X_i in the graph, then the joint probability distribution factors as

$$\Pr(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} \Pr(X_i | \mathrm{Pa}(X_i)). \quad (8)$$

Two different Bayesian networks can encode the same set of independencies. Such networks are said to be *equivalent*. Equivalent networks cannot be distinguished when inferring the network from measurement data. One way to bypass this difficulty is to perform targeted intervention experiments which can narrow the range of possible network architectures.

Dynamic Bayesian networks (DBNs) are extensions of Bayesian networks to the temporal domain and can be used to model stochastic processes [70]. DBNs generalise hidden Markov models and linear dynamical systems by representing the conditional dependencies and independencies between variables over time. Contrary to Bayesian networks, DBNs can be used to model feedback relationships, a ubiquitous element in genetic regulation. In comparison to PBNs, dynamic Bayesian networks support the assignment of quantitative state values, making this modelling approach more flexible to handle various types of data. DBNs are broadly applied to represent biological networks such as gene regulatory networks [124-127], signal transduction networks, e.g., [128-130], metabolic networks [131], as well as networks in physiology and medicine [132-136].

As shown in [137], PBNs and binary-valued DBNs whose initial and transition Bayesian networks are assumed to have only within and between consecutive slice connections, respectively, can represent the same joint probability distribution over their common variables. This is true both for independent as well as dependent variants of PBNs. However, there are many statistically equivalent PBNs that correspond to a DBN. On one hand, the PBN framework can be considered as redundant from the probabilistic point of view. On the other hand, it is richer from the functional point of view because it models the regulatory roles of different gene sets in more detail than the conditional probabilities in DBNs [137]. The conversion algorithms between the two modelling formalism are presented in [137], both for independent and dependent PBNs. Also the extensions of standard PBNs to context-sensitive PBNp is discussed. The perturbations and context switching can be introduced in the DBN formalism by adding additional hidden nodes to the dynamic Bayesian network, as shown in [137].

In terms of applications, it has been shown that both the PBN and the DBN approaches principally have good performance on the inference of gene regulatory networks from microarray data [138]. In addition, the connection between PBNs and DBNs makes it possible to apply the advanced DBNs to PBNs tools and vice versa. For example, an abundant collection of learning theory and algorithms for DBNs already exists and methods for the analysis of temporal behaviour of DBNs are already established. These techniques can be tailored to be applied directly in the context of PBNs. Conversely, the tool for controlling the steady-state behaviour of the networks, tools for network projection, node adjunction, resolution reduction as well as efficient learning schemes can be applied to DBNs.

As presented in [139], PBNs and dynamic Bayesian networks can be viewed as consisting of a probabilistic (Markov chain) and of a (Boolean) logic component. In the case of a dynamic Bayesian network, the probabilistic component is defined by a conditional probability chain rule and a Markov chain while the logic component is given by propositional logic with structural requirements. As shown in [139], Bayesian networks, with their hierarchical and dynamic variants, as well as probabilistic Boolean networks, are all generalised by Markov logic networks. The same separation of components applies. For a Markov logic network, the probabilistic component is a Markov random field and the logic component is the first order logic. We refer to [139] for more details on this framework, its applications in biology and medicine as well as the relationship with Bayesian networks.

PBN applications in biological and biomedical studies

PBN models for the representation of biological networks

Even though a significant part of the research on PBNs is theoretical, a large number of applied studies on the use of PBNs for various biological systems can be found in the literature. This is particularly the case with inference of models for molecular and physiological networks (from prior knowledge or data), with subsequent model analysis that leads to novel knowledge in biology and medicine.

PBNs as models of gene regulatory networks

PBNs were originally developed as models for Gene Regulatory Networks (GRNs) [3,8]. As stated in [32], PBNs 1) incorporate rule-based dependencies between genes; 2) allow the systematic study of global network dynamics; 3) are able to cope with uncertainty, both in the data and model selection; and 4) permit the quantification of the relative influence and sensitivity of genes in their interactions with other genes. In the PBN modelling framework, gene expression is quantised to two levels: ON and OFF.

The dynamical behaviour of PBNs can be used to model many biologically meaningful phenomena, such as cellular state dynamics possessing switch-like behaviour, hysteresis, stability, and etc. [32,140]. Often, the attractor cycles are interpreted as functional states on physiological time scales or as cellular phenotypes on developmental timescales [7,8]. This interpretation is fairly reasonable as most cell types are characterised by stable recurrent patterns of gene expression [31].

In the past years, there were several studies which successfully applied PBNs for the construction of GRNs from high-throughput gene expression microarray experiment data. In 2006, Yu et al. inferred a GRN of the interferon pathway in macrophages using time-course gene expression data [22]. The optimal network was identified applying the CoD approach. It was shown that the respective selection probabilities are varying for different biological conditions, e.g., after interferon treatment or after viral infection on macrophage, while the structure of the constituent network, i.e., predictor functions, remains stable. With a similar approach, Nguyen et al. inferred a GRN of hepatocellular carcinoma from microarray data and compared it to a network derived from control non-cancerous samples [141]. They indicated that certain genes in tumour samples show activity in steady-state periods while there is no activity for these genes in the control (non-cancerous) samples. This allowed to distinguish different gene regulatory processes being realized with the same set of genes.

Hashimoto et al. modelled the cell cycle of budding yeast by using context-sensitive PBNs [23]. They showed that the switching behaviour from stationary G1 phase to excited G1 phase in the PBN model is more frequent, when compared to the stochastic model of Zhang et al. [142]. Recently, Todd et al. identified the ergodic sets of states in PBNs that correspond to each phase of the budding yeast cell cycle, which in turn correspond to the cellular phenotypes [44]. The analysis of the dynamical behaviour gave additional insights on yeast cell cycle regulation, e.g., the yeast cell cycle network showed robustness both to external variable environments and to certain perturbations such as nitrogen deprivation, where yeast cells proceeded through one round of division and arrest at G1 phase without appreciable growth.

In 2011, Flöttmann et al. modelled the regulatory processes that govern the production of induced Pluripotent Stem (iPS) cells by considering the interplay between gene expression, chromatin modification, and DNA methylation [24]. As there is no clear guideline on how to assign Boolean functions to represent the interactions of each gene, their PBN model was designed to work by representing uncertainty via two assignments. First, a number of possible functions were assigned to the corresponding nodes with different probabilities. Second, the influences of certain nodes were split into separated Boolean functions with varied selection probabilities. A flexibility was thus allowed for choosing Boolean functions that fit the experimental data. With their PBN model, an extensive analysis was performed, allowing to demonstrate epigenetic landscape changes from differentiated cells to iPS cells as a function of time step. In addition, by looking at model variants of the core iPS regulation, it was shown that an increased chromatin modification rate could improve reprogramming efficiency while faster changes in DNA methylation could provide an enhanced rate though at the price of trading-off efficiency.

PBN within signal transduction network and metabolic network modelling

To date, there is no study which specifically applied PBN as a stand-alone framework for modelling signal transduction or metabolic networks. Nevertheless, PBN was combined with other algorithms or modelling frameworks. Fertig et al. presented *GESSA*, Graphically Extended Stochastic Simulation Algorithm, a mechanistic hybrid model which integrates the network model of cell signalling with pooled PBN to a differential equation-based model of transcription and translation computed by a stochastic simulation algorithm [25]. The cell signalling PBN model is generated by simulating individual protein copies with the corresponding state transitions updated according to the rules in the PBN. The sum of the resulting molecular states across copies, i.e., of each individual species, is compared to the initial state, the difference being afterwards returned and the cellular state being updated. *GESSA* was applied to the study of the cell fate decision of vulval precursor cells in *C. elegans*, where model predictions matched the experimental results even for minimal parameterisations of the PBN. It was thus shown that PBN could be an essential component when flexibility is needed in multi-level data integration and model construction.

In metabolic modelling, Chandrasekaran et al. presented an automated algorithm for the Probabilistic Regulation of Metabolism (*PROM*), allowing to reconstruct a probabilistic GRN integrated with a metabolic network from high-throughput data[26]. *PROM* makes use of conditional probabilities to model transcriptional regulation, similar to the CoD concept in PBN inference. This formalism permits the strength of transcription factor (TF)-gene regulation as well as gene states to be represented in terms of probabilities. *PROM* was used to generate a genome-scale integrative transcriptomic and metabolomic network of *Escherichia coli*, where *PROM* surpassed the state-of-the-art methods such as the regulatory flux balance analysis. *PROM* was also used to generate an integrative model of *Mycobacterium tuberculosis*.

The results from the model analysis offered additional details on known regulatory mechanisms and also helped to uncover the function of less studied genes on metabolic regulation.

Apart from these two studies, several other works also made use of a probabilistic framework for analysing signal transduction and metabolic networks. Kaderali et al., for instance, developed an algorithm that reconstructs signalling pathways from gene knockdown data (RNAi data) [143]. In this work, pathway topologies are inferred by using Bayesian networks with probabilistic Boolean threshold functions. The algorithm was used to study the Janus Kinase and Signal Transducers and Activators of Transcription (JAK/STAT) pathway, correctly reconstructing the core topology of the pathway along with model variants. Similarly, Sauer et al. [144] used probabilistic equations to determine flux ratios, allowing to express the relative contribution of certain metabolites or pathways as modulators in the network. This assignment is more realistic than using flux absolute integer numbers, given that the flux of each source can relatively contribute to the production of certain metabolites.

PBN applications in the context of physiology

PBNs were also used in the recent years for studying networks in physiology, with a close link to medicine. Tay et al. described a dengue hemorrhagic fever (DHF) infection model which contains the interplay between dengue virus and different cytokines which are cross-regulated in T-helper 1 (Th1) and Th2 cells [9]. In their work, a single probabilistic Karnaugh-Map is generated, modelling the inducement probability of each cell as to define the overall influence of inducing nodes. Simulation results matched clinical data for both synchronous and asynchronous updating, with respect to the form and the average duration-based attractors, respectively. In addition, by applying a genetic algorithm [145] to modulate the DHF attractor basins to dengue fever (DF) basins (a less severe form of DHF), Tay et al. also identified the tumour growth factor beta (TGFβ), interleukine-8 (IL-8) and IL-13, as sensitive intervention points.

Another example in this field can be found in the study of Ma et al., where, based on functional Magnetic Resonance Imaging (fMRI) data, the authors developed a brain connectivity network model for Parkinson's disease [10]. A method similar to the one of Yu et al. [22] was used for probability inference selection, i.e., the calculation of CoDs. Then the CoDs were subsequently used to generate an influence matrix representation of the brain signal connectivity among brain components. The obtained results showed that a significant difference in connectivity exists for many paired brain-components comparing between normal, Parkinson's disease with drug, and Parkinson's disease with drug withdrawal conditions, and this difference was expressed in terms of estimated range of coefficient mean activity. This particular information may allow to construct a new screening procedure for Parkinson's disease diagnose and to determine drug trial responsiveness based on a non-invasive, fMRI-based investigation in the future.

A certain number of the previously described (applied research) articles on PBN have applications not only in molecular biology, but also in physiology or medicine. Only to name a few examples, being able to distinguish among the regulatory networks of cancer and healthy cells, as presented by Nguyen et al., could contribute to an early detection of cancerous genes in susceptible populations [141]. A better understanding of dynamic processes and the control of somatic cell programming, as proposed by Fertig et al., may lead to a future use of iPS cells in cell or tissue replacement therapies [25]. Last but not least, the *PROM* algorithm, as introduced by Chandrasekaran et al., is capable of predicting transcription factor drug targets which are major hubs in the cellular network of pathogenic organism such as *Mycobacterium tuberculosis* [26]. A further development of drugs in this direction may help in the treatment of different infectious diseases. This new line of treatment could have a strong impact for third-world countries where infectious diseases still remain a major cause of death.

PBN for Systems Biology and Systems Biomedicine?

As previously discussed, the PBN framework is a topic of intensive and continuous theoretical research with successful applications in the biomedical area. To describe and extend a vision on future PBNs' applications, we summarise additional arguments to support why this modelling approach is suitable for future research in Systems Biology and Systems Biomedicine.

Data integration

Different types of biological and clinical investigation datasets, ranging from qualitative to high-throughput quantitative experimental data, were successfully applied in PBN inference and analysis. Yu et al. [22] and Nyugen et al. [141], for instance, inferred GRNs of macrophages and hepatocellular carcinoma using microarray gene expression data. Flöttmann et al. [24] built a comprehensive epigenetic regulatory network of iPS cells based on gene expression, chromatin modification and DNA methylation data generated from multiple high-throughput experiments. Ma et al. applied voxel selection on fMRI clinical data to capture the activities of each brain's compartment as the inputs for learning a functional brain connectivity network [10].

We have recently shown that the normalised activity of signalling proteins from quantitative western blot experiment can be compared to the steady-state probability

of certain molecule to be ON in instantaneously-random PBNs. In an ergodic model, the activities of signalling proteins, usually given by their phosphorylated forms normalised to the maximal signal, could be correlated with the steady-state probability distribution on the state space of the PBN model. With this regard, PBN could support the integration of semi-quantitative experimental data. Apart from quantitative western blot data, the profiles of signalling proteins from alternative experiments such as enzyme-link immunosorbent assay (ELISA) and high-throughput protein array data are also compatible with this framework (publication submitted).

The PBN framework also allows for the description and analysis of large-scale models, for instance as in the case of a Boolean model of apoptosis of Schlatter et al. [146]. Therein, a PBN model was derived from the original literature-based BN consisting of 86 nodes and 125 Boolean interactions. Quantitative experimental data in this study were normalised to the maximal signals across experiments and were used as input data for the PBN model. We analysed the strengths of canonical pathways and crosstalk interactions between different signalling components among apoptotic and related signalling pathways through the identification of selection probability. It was possible to obtain these via optimisation. Thereby a curated signal transduction network topology was derived. The resulting PBN demonstrates the correlation between UVB irradiation, NFκB, caspase 3, and apoptotic activities in a semi-quantitative manner which could not be demonstrated by the original BN. The analysis pointed at an inconsistent caspase 3 measurement, which shows no activity for high UVB irradiation while significant apoptosis is measured (see Figure 6, publication submitted).

Furthermore, the PBN framework has a good potential to describe cellular dynamics at multiple levels. Hybrid PBN-related models could be applied, as previously described, e.g., in the studies of Fertig et al. and Chandrasekaran et al. [25,26]. As reviewed in detail by Gonçalves et al. [147], bridging layers towards an integration of signal transduction, regulation and metabolism into mathematical models still posts many challenges as each of the biological layer has their own distinct characteristics and therefore is suitable for only a subset of modelling approaches. To address such challenges, an integrative hybrid model for flux balance analysis was proposed, combining BN modelling for the gene regulatory part, ODE modelling for the signal transduction part and

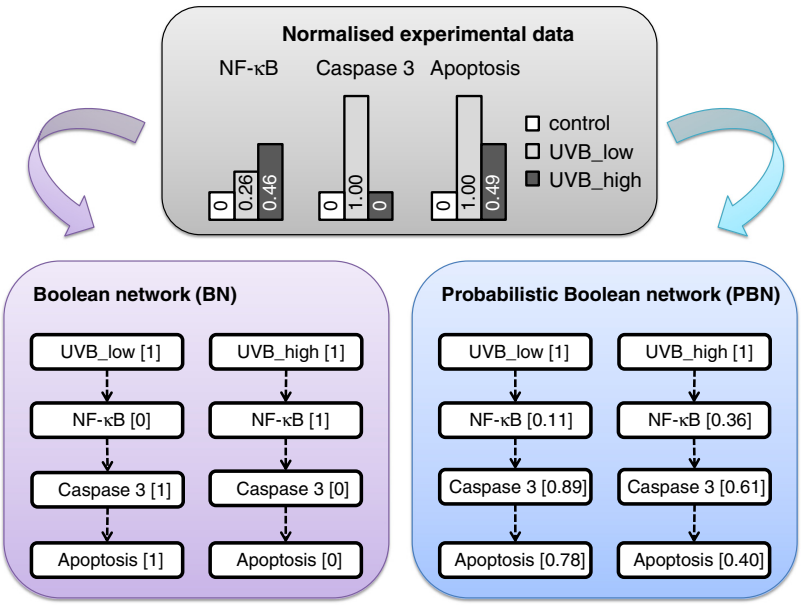


Figure 6 A comparative study of apoptotic signalling in the context of Boolean and probabilistic Boolean networks. Background subtracted and normalised experimental data derived from Schlatter et al. is shown in the top box. The experimental data compare the activities of downstream signalling molecules and apoptotic activity in the control setting (no stimulation) versus two intensities of UVB irradiation (UVB_low, 300 J/m^2 and UVB_high, 600 J/m^2). The activities of caspase 3 refer to the high caspase 3 activities of the original publication. The steady-state values from the original BN and the steady-state probability of the molecules to be ON from the optimised PBN of one exemplifying run are shown in brackets. (Note: The interactions between each node in the actual network are much more complex than the simplified diagram as shown.).

flux balance analysis for the metabolic part. With this regard, PBN could also be integrated as part of such a hybrid model to describe GRNs and/or signalling networks to provide more details on modelling analysis and interpretation comparing to traditional BNs.

Computational tools for PBN modelling and analysis

Several PBN modelling and analysis tools were continuously developed over the past recent years. The *BN/PBN* MATLAB-based toolbox, introduced by Lähdesmäki and Shmulevich in 2003 [27], deals with the simulation, analysis (network statistics, state transitions and distributions), visualisation and intervention analysis of both BN and PBN models. The toolbox was specifically designed for GRN inference and it makes use of CoD calculations. State transition probabilities and influence values (the indicators for interactive effect for each pair of genes) are subsequently calculated based on these calculated CoDs. Ma et al. successfully applied the *BN/PBN* toolbox to infer and analyse the brain connectivity network of Parkinson's disease patients, as previously described in Section 'PBN applications in the context of physiology'.

Hinkelmann et al. introduced *ADAM* (Analysis of Discrete Models of biological systems using computer Algebra) [28], a web-based tool for rapid steady-state identification in various discrete model types. The tool automatically converts discrete models into polynomial dynamic systems, allowing to run computer-based algebra analysis. For probabilistic networks, *ADAM* generates a graph of all possible (local rule) updates, thus being capable to build an enumeration of all steady states. *Boolnet*, as introduced by Müssel et al., is an R-package for the generation, modelling, reconstruction and analysis of both synchronous and asynchronous BNs or PBNs [29]. The toolbox features time-series (experimental data) based network inference, e.g., making use of Markov chain simulations for attractor identification with subsequent visualisation and robustness analysis via network perturbation or heuristic search and random walks. We have recently developed *optPBN*, a MATLAB-based toolbox for PBN optimisation based on the *BN/PBN* toolbox. PBNs can easily be constructed from Boolean rule-based models. The toolbox also provides a flexible platform for data integration (e.g., to integrate data from multiple experiments). Different algorithms can be used to address the resulting optimisation problem. Thus, based on normalised protein activity at steady-state data, one can identify a curated model structure from different candidate models. Subsequent analysis on the curated PBN can be performed in the *BN/PBN* toolbox (publication submitted).

We also discuss a few different algorithms and tools which are not specifically designed for PBN but with a high potential for the analysis of PBNs. *PROM*, for example, offers a mean to calculate the flux activities of a metabolic network in a probabilistic manner based on gene expression data [26]. Specifically, this gives rise to the applicability of the PBN framework for metabolic models. Recently, Terfve et al. introduced *CellNOptR*, a flexible toolkit for training protein signalling networks based on a multiple logic formalism [148]. *CellNOptR* offers support for optimisation with respect to multiple modelling frameworks, ranging from logical to ODE (logic rule derived) models. Extending *CellNOptR* towards a probabilistic modelling framework is also foreseen for future work.

A perspective on potential applications of PBNs in a clinical setting

It has been a decade since the completion of the Human Genome Project in 2003 that initiated the era of biological and medical investigation in omic scales [149]. Due to technology advancements, the costs of genome sequencing and high throughput biomedical investigations are exponentially decreased and they might become part of the routine medical investigations in a foreseeable time frame [150]. Datasets from omics experiments usually consist of large lists of numbers that represent genes, transcripts, proteins, or metabolites depending on the method applied. In the near future all these methodologies might be applied together routinely, even in time series examinations. The major problem with such data is their high complexity and the need to make them interpretable by the medical staff. Therefore, there is a strong demand for reasonable computational approaches to integrate multidimensional "big data" [151]. In addition, given the rich sets of information from individual patients that physicians will acquire, smart approaches are mandatory to translate and simplify these large-scale biomedical data. Such approaches should facilitate a physician's decision-making process to provide more accurate diagnosis and optimal treatment.

For these fields we identify the PBN framework as a powerful tool. Recent applications of PBN modelling of gene regulatory and signalling networks have been described in the previous section. As previously summarised, PBNs allow an effective visualisation of GRN models [9,10], allowing to represent gene function and activity [152]. These efforts foster the understanding of gene-gene interactions, consequences of aberrant gene function and targeted perturbations of such networks, as well as finding out the least adverse effects of perturbations [9,153]. PBNs allow for the integration of information from large data sets and for inferring logical relationships between genes/networks. This feature is of particular benefit as many relationships and structural connections among genes are not known. Unknown

relationships between transcripts and proteins can also be assessed. In a therapeutic perspective PBNs could be used in a disease-relevant context because many, foremost chronic diseases, share probably common underlying mechanisms that are not elucidated so far [154]. Using PBNs in the study of disease-related networks could enable us to take genetic interactions into account and associations could be generated to identify comorbidities sharing common causative factors. Skahanenko et al. for instance have applied Markov logic networks, a probabilistic logic modelling approach in the same category as PBN, to explore gene-phenotype associations. Whereas traditional statistical methods are employed to identify the marker that associates the most with an observed phenotype, Markov logic networks can be used to identify a subset of markers that predicts the phenotype. Within this method, the relationship between the genetic markers and phenotype(s) can be hypothesised and modelled. All models can then be tested and their respective probability can be derived [139].

In the context of a single, yet complex disease, the study on brain connectivity in Parkinson's disease by Ma et al. [10] is a good example showing how a probabilistic model such as PBN could translate large-scale biomedical data into a potential application in clinic. fMRI principally measures blood oxygen level-dependent signals that are correlated to the blood flow into different regions of the brain, which in turn give physicians information on the functional activity of specific areas of the brain [155]. For some neurological disorders, such as Parkinson's disease, the lesions mainly affect a specific area of the brain such as basal ganglia, but have consequences on the overall integrity of brain connectivity, especially on the dopaminergic pathway-dependent motor and cognitive control [156]. Therefore, considering the aetiology and disease progression from only conventional MRI data which demonstrate only structural information is certainly insufficient to yield a comprehensive understanding on the course of disease. Considering diseases as network perturbations [157], the PBN model from Ma et al. demonstrated differences of brain connectivity networks comparing healthy population and diseased cases with and without medication. Such observations could possibly be further developed towards clinical biomarkers which could then be added to physicians' portfolio and in turn facilitate diagnostic process, treatment design, and follow-up strategy.

Generally, the incorporation of tentativeness and probability could be evolved into a valid concept in a clinical setting, as routine medical investigation often provides no conclusive data. Together with a comprehensive reduction and translation of large-scale and complex biomedical data, the PBN framework might serve as a mean to develop simplistic terms like a probability score for certain condition, e.g., for having a disease or of being responsive to treatment. Such a probabilistic score could serve as a simple but powerful additional input for physicians in order to improve their healthcare management. As a whole, healthcare systems would benefit from reducing costs related to unnecessary diagnostic investigations and treatment failures.

Conclusion

Even though the concept of PBN for the modelling of biological systems is still young compared to other modelling approaches, a broad area of research activities on this modelling approach such as network inference and network control have been well-established and are continuously developed. For a meaningful comparison of different inference algorithms in the future, it is necessary to quantify their performance. The prospective research in the area of network inference is to develop a formal framework for validation of network inference procedures. Moreover, there is a demand for establishing the properties of network inference procedures under various conditions, e.g., model class, distance function, etc. The current trend in structural intervention and external control is to develop new methods to reduce their computational complexity and to define the optimal control problems and find the corresponding optimal policies for specific therapies. With its flexibility for data integration and the availability of supporting algorithms and computational tools, PBN is one of the most suitable modelling frameworks to describe and analyse complex biological systems from molecular to physiological levels with possible future application at clinical level.

Endnotes

[a]In general, $\gamma_1, \gamma_2, \ldots, \gamma_n$ need not be independent and identically distributed random variables, but for the simplicity of presentation are assumed so.

[b]A state in a Markov chain is said to be ergodic if returns to the state can occur at irregular times and the state is positive recurrent. If all states in an (irreducible) Markov chain are ergodic, then the chain itself is said to be ergodic.

[c]In a generalised PBN framework a network variable can have any value in $\{0, 1, \ldots, d-1\}$, where $d > 2$.

[d]In the graph-theoretical terminology the notion of an ergodic set of states in a Markov chain corresponds to the notion of a bottom strongly connected component in a graph.

[e]In computer science, the complexity of a function or an algorithm is expressed or characterised using the big *O* notation, namely, how the function or algorithm responds to changes in its input size.

[f]Optimisation deals with a broad range of problems, relying on, for example, convex programming, optimal

control, combinatorial optimisation or evolutionary computation paradigms; examples and additional information can be found by referring to [158-165]

[g]A one-time gene perturbation changes the value of one or more genes without modifying the rules or probabilistic parameters of the network.

[h]In computational complexity theory, NP-hard is a class of problems that are at least as hard as the hardest problems in NP (nondeterministic polynomial time).

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
PT wrote background, PBNs for the representation of biological networks, PBNs for multi-level Systems Biology, computational tools, partly of a perspective on potential applications of PBNs in clinic, and co-ordinate overall writing. AM wrote theoretical and mathematics sections on introduction, PBNs dynamic, inference and technical comparison between PBN to other probabilistic graphical models. JP wrote theoretical and mathematics sections on intervention and control of PBNs. AT wrote a part on the optimisation of PBNs. JS shared a medical perspective on the application of PBNs and wrote a perspective on potential applications of PBNs in clinic. TS supervised and integrated the overall writing together with PT and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements
We acknowledge with thanks financial support from the Fonds National de la Recherche (FNR) Luxembourg (grant 1233900). Andrzej Mizera is on leave of absence from the Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland.

Author details
[1] Life Sciences Research Unit,University of Luxembourg, Luxembourg. [2] Computer Science and Communications Research Unit, University of Luxembourg, Luxembourg. [3] Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg. [4]Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, Luxembourg. [5]Saarland University Medical Center, Department of Internal Medicine II, Homburg, Saarland, Germany.

Received: 29 March 2013 Accepted: 22 June 2013
Published: 01 July 2013

References

1. Raue A, Kreutz C, Maiwald T, Klingmueller U, Timmer J: **Addressing parameter identifiability by model-based experimentation.** *IET Syst Biol* 2010, **5**(2):120–130.
2. Jack J, Wambaugh J, Shah I: **Simulating quantitative cellular responses using asynchronous threshold Boolean network ensembles.** *BMC Syst Biol* 2011, **5**(1):1–13.
3. Shmulevich I, Dougherty ER, Zhang W: **From Boolean to probabilistic Boolean networks as models of genetic regulatory networks.** *Proc of the IEEE* 2002, **90**(11):1778–1792.
4. Kauffman SA: **Metabolic stability and epigenesis in randomly constructed genetic nets.** *J Theor Biol* 1969, **22**(3):437–467.
5. Kauffman SA: **Homeostasis and differentiation in random genetic control networks.** *Nature* 1969, **224:**177–178.
6. Kauffman SA: **The large scale structure and dynamics of gene control circuits: an ensemble approach.** *J Theor Biol* 1974, **44:**167–190.
7. Kauffman SA: *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford: Oxford University Press; 1993.
8. Shmulevich I, Dougherty ER: *Probabilistic Boolean Networks: The Modeling and Control of Gene Regulatory Networks*. Philadelphia PA: SIAM Press; 2010.
9. Tay J, Tan P: **Finding intervention points in the pathogenesis of dengue viral infection.** In *Proc. 28th IEEE EMBS Annual International Conference: 30 August - 3 September; New York City, NY, USA*: IEEE Computer Society; 2006:5315–5321.
10. Ma Z, Wang Z, McKeown M: **Probabilistic Boolean network analysis of brain connectivity in Parkinsons disease.** *IEEE J Selected Topics in Signal Process* 2008, **2**(6):975–985.
11. Shmulevich I, Gluhovsky I, Hashimoto RF, Dougherty ER, Zhang W: **Steady-state analysis of genetic regulatory networks modelled by probabilistic Boolean networks.** *Comp and Funct Genomics* 2003, **4:**601–608.
12. Ching W, Zhang S, Ng M, Akutsu T: **An approximation method for solving the steady-state probability distribution of probabilistic Boolean networks.** *Bioinformatics* 2007, **23**(12):1511–1518.
13. Li W, Cui L, Ng MK: **On computation of the steady-state probability distribution of probabilistic Boolean networks with gene perturbation.** *J Comput Appl Math* 2012, **236:**4067–4081.
14. Hashimoto R, Kim S, Shmulevich I, Zhang W, Bittner M, Dougherty E: **Growing genetic regulatory networks from seed genes.** *Bioinformatics* 2004, **20**(8):1241–1247.
15. Ching W, Chen X, Tsing N: **Generating probabilistic Boolean networks from a prescribed transition probability matrix.** *IET Syst Biol* 2009, **3**(6):453–464.
16. Vahedi G, Ivanov I, Dougherty E: **Inference of Boolean networks under constraint on bidirectional gene relationships.** *IET Syst Biol* 2009, **3**(3):191–202.
17. Pal R, Datta A, Bittner ML, Dougherty ER: **Intervention in context-sensitive probabilistic Boolean networks.** *Bioinformatics* 2005, **21**(7):1211–1218.
18. Shmulevich I, Dougherty ER, Zhang W: **Gene perturbation and intervention in probabilistic Boolean networks.** *Bioinformatics* 2002, **18**(10):1319–1331.
19. Liu Q: **An optimal control approach to probabilistic Boolean networks.** *Physica A* 2012, **391:**6682–6689.
20. Kobayashi K, Hiraishi K: **Reachability analysis of probabilistic Boolean networks using model checking.** In *Proc. SICE Annual Conference: 18-21 August; Taipei, Taiwan*: IEEE Computer Society; 2010:829–832.
21. Qian X, Dougherty E: **Comparative study on sensitivities of Boolean networks.** In *Proc. IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS): 10-12 November; Cold Spring Harbor, NY, USA*: IEEE Computer Society; 2010:1–4.
22. Yu L, Marshall S, Forster T, Ghazal P: **Modelling of macrophage gene expression in the interferon pathway.** In *Proc. IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS): 28-30 May; College Station, Texas, USA*: IEEE Computer Society; 2006:45–46.
23. Hashimoto R, Stagni H, Higa C: **Budding yeast cell cycle modeled by context-sensitive probabilistic Boolean network.** *Proc. IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS): 17-21 May; Minneapolis, MN, USA* 2009:1–4.
24. Flöttmann M, Scharp T, Klipp E: **A stochastic model of epigenetic dynamics in somatic cell reprogramming.** *Front Physiol* 2012, **3**(216):1–19.
25. Fertig E, Danilova L, Favorov A, Ochs M: **Hybrid modeling of cell signaling and transcriptional reprogramming and its application in C.elegans development.** *Front Genet* 2011, **2**(77):1–9.
26. Chandrasekaran S, Price N: **Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis.** *Proc Nat Acad Sci* 2010, **107**(41):17845–17850.
27. **BN/PBN Toolbox.** [http://code.google.com/p/pbn-matlab-toolbox]
28. Hinkelmann F, Brandon M, Guang B, McNeill R, Blekherman G, Veliz-Cuba A, Laubenbacher R: **ADAM: analysis of discrete models of biological systems using computer algebra.** *BMC Bioinformatics* 2011, **12:**295
29. Müssel C, Hopfensitz M, Kestler H: **Boolnet: an R package for generation, reconstruction and analysis of Boolean networks.** *Bioinformatics* 2010, **20**(10):1378–1380.
30. Norris JR: *Markov Chains*. Cambridge UK: Cambridge University Press; 1998.
31. Brun M, Dougherty ER, Shmulevich I: **Steady-state probabilities for attractors in probabilistic Boolean networks.** *Signal Process* 2005, **85:**1993–2013.

32. Shmulevich I, Dougherty ER, Kim S, Zhang W: **Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks.** *Bioinformatics* 2002, **18**(2):261–274.
33. Shmulevich I, Dougherty ER, Zhang W: **Control Of stationary behavior in probabilistic Boolean networks by means Of structural intervention.** *J Biol Syst* 2002, **10**(4):431–445.
34. Choudhary A, Datta A, Bittner ML, Dougherty ER: **Intervention in a family of Boolean networks.** *Bioinformatics* 2006, **22**(2):226–232.
35. Stewart WJ: *Introduction to the Numerical Solution of Markov Chains*. Princeton NJ: Princeton University Press; 1994.
36. Zhang S, Ching W, Ng MK, Akutsu T: **Simulation study in probabilistic Boolean network models for genetic regulatory networks.** *Int J Data Mining Bioinf* 2007, **1**(3):217–240.
37. Xu W, Ching W, Zhang S, Li W, Xiao-ShanChen: **A matrix perturbation method for computing the steady-state probability distributions of probabilistic Boolean networks with gene perturbations.** *J Comput Appl Math* 2011, **235:**2242–2251.
38. Berg BA: *Markov Chain Monte Carlo Simulations and Their Statistical Analysis*. Singapore: World Scientific Publishing; 2004.
39. Rosenthal JS: **Minorization Conditions and Convergence Rates for Markov Chain Monte Carlo.** *J Am Stat Assoc* 1995, **90**(430):558–566.
40. Robert CP: **Convergence control techniques for Markov chain Monte Carlo algorithms.** *Stat Sci* 1995, **10**(3):231–253.
41. Cowles MK, Carlin BP: **Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review.** *J Am Stat Assoc* 1996, **91**(434):883–904.
42. Raftery AE, Lewis S: **How Many Iterations in the Gibbs Sampler?** In *Bayesian Statistics, Volume 4*. Edited by Bernardo JM, Berger JO, Dawid AP, Smith AFM. Oxford UK: Oxford University Press; 1992:763–773.
43. Ribeiro AS, Kauffman SA: **Noisy attractors and ergodic sets in models of gene regulatory networks.** *J Theor Biol* 2007, **247**(4): 743–755.
44. Todd R, Helikar T: **Ergodic Sets as cell phenotype of budding yeast cell cycle.** *PLOS One* 2012, **7**(10):e45780.
45. Pal R: **Analyzing steady state probability distributions of context-sensitive probabilistic Boolean networks.** In *Proc. IEEE, International Workshop on Genomic Signal Processing and Statistics (GENSIPS): 17-21 May; Minneapolis, MN, USA*: IEEE Computer Society; 2009:1–4.
46. Pal R: **Context-Sensitive Probabilistic Boolean networks: steady-state properties, reduction, and steady-state approximation.** *IEEE Trans Signal Process* 2010, **58**(2):879–890.
47. Hayashida M, Tamura T, Akutsu T, Ching W, Cong Y: **Distribution and enumeration of attractors in probabilistic Boolean networks.** *IET Syst Biol* 2009, **3**(6):465–474.
48. Li F, Sun J: **Stability and Stabilization Issue of Probabilistic Boolean Network.** In *Proc. 30th Chinese Control Conference: 22-24 July; Yantai, China*: IEEE Computer Society; 2011:6381–6385.
49. Avi nó M A: **Homomorphisms of probabilistic gene regulatory networks.** In *Proc. IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS): 28-30 May; College Station, Texas, USA*: IEEE Computer Society; 2006:85–86.
50. Pal R, Ivanov I, Datta A, Bittner ML, Dougherty ER: **Generating Boolean networks with a prescribed attractor structure.** *Bioinformatics* 2005, **21**(21):4021–4025.
51. Pfahringer B: **Compression-based discretization of continuous attributes.** In *Proc. 12th International Conference on Machine Learning: 9-12 July; Tahoe City, CA, USA*: Morgan Kaufmann Publishers; 1995:456–463.
52. Dougherty J, Kohavi R, Sahami M: **Supervised and unsupervised discretization of continuous features.** In *Proc. 12th International Conference on Machine Learning: 9-12 July; Tahoe City, CA, USA*: Morgan Kaufmann Publishers; 1995:194–202.
53. Mircean C, Tabus I, Astola J, Kobayashi T, Shiku H, Yamaguchi M, B Y, Shmulevich I, Zhang W: **Quantization and similarity measure selection for discrimination of lymphoma subtypes under k-nearest neighbor classification.** In *Microarrays and Combinatorial Techniques: Design, Fabrication, and Analysis, Volume 5328 of Proceedings of the SPIE: January; San Jose, CA, USA*; 2004:6–17. BiOS.
54. Snoussi EH: **Qualitative dynamics of piecewise-linear differential equations: a discrete mapping approach.** *Dyn and Stability Syst* 1989, **4**(3–4):189–207.
55. Gersho A, Gray RM: *Vector Quantization and Signal Compression*: Boston Kluwer Academic Publishers; 1992.
56. Shmulevich I, Zhang W: **Binary analysis and optimization-based normalization of gene expression data.** *Bioinformatics* 2002, **18**(4):555–565.
57. Zhou X, Wang X, Dougherty ER: **Binarization of microarray data on the basis of a mixture model.** *Mol Cancer Ther* 2003, **2**(7):679–684.
58. Hopfensitz M, Müssel C, Wawra C, Maucher M, Kühl M, Neumann H, Kestler HA: **Multiscale binarization of gene expression data for reconstructing Boolean networks.** *IEEE/ACM Trans Comput Biol Bioinf* 2012, **9**(2):487–498.
59. Duan R, Man H, Jiang W, Liu W: **Activation detection on fMRI time series using hidden Markov model.** In *Proc. 2nd International IEEE EMBS Conference on Neural Engineering: 16-19 March; Arlington, VA, USA*: IEEE Computer Society; 2005:510–513.
60. Pitt L, Valiant LG: **Computational limitations on learning from examples.** *J ACM* 1988, **35:**965–984.
61. Boros E, Ibaraki T, Makino K: **Error-free and best-fit extensions of partially defined Boolean functions.** *Inf Comput* 1998, **140**(2):254–283.
62. Lähdesmäki H, Shmulevich I, Yli-Harja O: **On learning gene regulatory networks under the Boolean network model.** *Mach Learn* 2003, **52:**147–167.
63. Böck M, Ogishima S, Tanaka H, Kramer S, Kaderali L: **Hub-centered gene network reconstruction using automatic relevance determination.** *PloS ONE* 2012, **7**(5):e35077.
64. Liu W, Lähdesmäki H, Dougherty ER, Shmulevich I: **Inference of Boolean networks using sensitivity regularization.** *EURASIP J Bioinf Syst Biol* 2008, **2008:**780541.
65. Arnone MI, Davidson EH: **The hardwiring of development: organization and function of genomic regulatory systems.** *Development* 1997, **124**(10):1851–1864.
66. Chen T, He HL, Church GM: **Modeling gene expression with differential equations.** In *Proceedings of the Pacific Symposium on Biocomputing, Volume 4: 4-9 January; Hawaii, USA*; 1999:29–40. World Scientific Press.
67. Someren EPV, Wessels LFA, Reinders MJT, Backer E: **Searching for limited connectivity in genetic network models.** In *Proc. 2nd International Conference on Systems Biology: 4-7 November; Pasadena, CA, USA*: Omnipress; 2001:222–230.
68. Guthke R, Möller U, Hoffmann M, Thies F, Töpfer S: **Dynamic network reconstruction from gene expression data applied to immune response during bacterial infection.** *Bioinformatics* 2005, **21**(8):1626–1634.
69. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási A: **The large-scale organization of metabolic networks.** *Nature* 2000, **407:**651–654.
70. Murphy K, Mian S: *Modelling Gene expression Data Using Dynamic Bayesian Networks*. Tech. Rep., Berkeley CA: University of California Press; 1999.
71. Perrin B, Ralaivola L, Mazurie A, Bottani S, Mallet J, d'Alché-Buc F: **Gene networks inference using dynamic Bayesian networks.** *Bioinformatics* 2003, **19**(Suppl. 2):ii138–ii148.
72. Styczynski MP, Stephanopoulos G: **Overview of computational methods for the inference of gene regulatory networks.** *Comput Chem Eng* 2005, **29:**519–534.
73. Lindlöf A, Olsson B: **Could correlation-based methods be used to derive genetic association networks?** *Inf Sci* 2002, **146**(1-4):103–113.
74. Maucher M, Kracher B, M Kühl M, Kestler HA: **Inferring Boolean network structure via correlation.** *Bioinformatics* 2011, **27**(11):1529–1536.
75. Xiao Y, Dougherty E: **Optimizing consistency-based design of context-sensitive gene regulatory networks.** *IEEE Trans Circuits and Syst* 2006, **53**(11):2431–2437.
76. Dougherty ER, Kim S, Chen Y: **Coeffcient of determination in nonlinear signal processing.** *Signal Process* 2000, **80:**2219–2235.
77. Ching W, Ng MM, Fung ES, Akutsu T: **On construction of stochastic genetic networks based on gene expression sequences.** *Int J Neural Syst* 2005, **15**(4):297–310.

78. Ching W, Chen X, Tsing N, Leung H: **A heuristic method for generating probabilistic Boolean networks from a Prescribed Transition Probability Matrix.** In *Proc. 2nd Symposium on Optimization and Systems Biology (OSB): 31 October - 3 November; Ligiang, China*: World Publishing Corporation; 2008:271–278.
79. Dougherty ER, Xiao Y: **Design of probabilistic Boolean networks under the requirement of contextual data consistency.** *IEEE Trans Signal Process* 2006, **54**(9):3603–3613.
80. Zhou X, Wang X, Pal R, Ivanov I, Bittner M, Dougherty ER: **A Bayesian connectivity-based approach to constructing probabilistic gene regulatory networks.** *Bioinformatics* 2004, **20**(17):2918–2927.
81. Rissanen J: **Modelling by shortest data description.** *Automatica* 1978, **14:**465–471.
82. Zhao W, Serpedin E, Dougherty ER: **Inferring gene regulatory networks from time series data using the minimum description length principle.** *Bioinformatics* 2006, **22**(17):2129–2135.
83. Dougherty J, Tabus I, Astola J: **Inference of gene regulatory networks based on a universal minimum description length.** *EURASIP J Bioinf Syst Biol* 2008, **2008:**482090.
84. Marshall S, Yu L, Xiao Y, Dougherty ER: **Temporal inference of probabilistic Boolean networks.** In *Proc. IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS): 28-30 May; College Station, Texas, USA*: IEEE Computer Society; 2006:71–72.
85. Marshall S, Yu L, Xiao Y, Dougherty ER: **Inference of a probabilistic Boolean network from a single observed temporal sequence.** *EURASIP J Bioinf Syst Biol* 2007, **2007:**32454.
86. Braga-Neto U: **Optimal state estimation for Boolean dynamical systems.** In *Proc. 45th Annual Asilomar Conference on Signals, Systems, and Computers: 6-9 November; Pacific Grove, CA, USA*: IEEE Computer Society; 2011:MP8a4-5.
87. Kim S, Li H, Dougherty ER, Cao N, Chen Y, Bittner M, Suh EB: **Can Markov chain models mimic biological regulation?** *J Biol Syst* 2002, **10**(4):337–357.
88. Shmulevich I, Dougherty ER: **Modeling genetic regulatory networks with probabilistic Boolean networks.** In *Genomic Signal Processing and Statistics*. Cairo Egypt: Hindawi Publishing Corporation; 2005.
89. Zhao W, Serpedin E, Dougherty ER: **Inferring connectivity of genetic regulatory networks using information-theoretic criteria.** *IEEE/ACM Trans Comput Biol Bioinf* 2008, **5**(2):262–274.
90. Zhou X, Wang X, Dougherty ER: **Construction of genomic networks using mutual-information clustering and reversible-jump Markov-chain-Monte-Carlo predictor design.** *Signal Process* 2003, **83:**745–761.
91. Bittner M, Meltzer P, Chen Y, Jiang Y, Sefftor E, Hendrix M, Radmacher M, Simon R, Yakhini Z, Ben-Dor A: **Molecular classification of cutaneous malignant melanoma by gene expression profiling.** *Nature* 2000, **406**(6795):536–540.
92. Zhang S, Ching W, Chen X, Tsing NK: **Generating probabilistic Boolean networks from a prescribed stationary distribution.** *Inf Sci* 2010, **180:**2560–2570.
93. Li W, Ching W, Cui L: **A modified Newton's method for inverse problem of probabilistic Boolean networks with gene perturbations.** In *Proc. IEEE International Conference on Systems Biology (ISB): 28 August - 1 September; Heidelberg-Mannheim, Germany*: Curran Associates, Inc; 2011:167–172.
94. Xiao Y, Dougherty ER: **The impact of function perturbations in Boolean networks.** *Bioinformatics* 2007, **23**(10):1265–1273.
95. Qian X, Dougherty ER: **Effect of function perturbation on the steady-state distribution of genetic regulatory networks: Optimal structural intervention.** *IEEE Trans Signal Process* 2008, **56**(10-1):4966–4975.
96. Qian X, Yoon B, Dougherty ER: **Structural intervention of gene regulatory networks by general rank-k matrix perturbation.** In *Proc. 2012 IEEE International Conference on Acoustics, Speech and Signal Processing: 25-30 March 2013 Kyoto, Japan*: IEEE Computer Society; 2012:729–732.
97. Sherman J, Morrison WJ: **Adjustment of an inverse matrix corresponding to a change in one element of a given matrix.** *Ann Math Stat* 1950, **21:**124–127.
98. Qian X, Dougherty ER: **On the long-run sensitivity of probabilistic Boolean networks.** *J Theor Biol* 2009, **257**(4):560–577.
99. Bellman R: *Dynamic Programming*. Mineola NY: Dover Publications; 2004.
100. Datta A, Choudhary A, Bittner ML, Dougherty ER: **External control in Markovian genetic regulatory networks.** *Mach Learn* 2003, **52**(1-2):169–191.
101. Datta A, Choudhary A, Bittner ML, Dougherty ER: **External control in Markovian genetic regulatory networks: the imperfect information case.** *Bioinformatics* 2004, **20**(6):924–930.
102. Akutsu T, Hayashida M, Ching W, Ng M: **Control of Boolean networks Haridness results and algorithms for tree structured networks.** *J Theor Biol* 2007, **244**(4):670–679.
103. Chen X, Ching W: **Finding optimal control policy by dynamic programming in conjunction with state reduction.** In *Proc. IEEE International Conference on Systems Biology: 2-4 September 2011; Zhuhai, China*: IEEE Computer Society; 2011:274–278.
104. Ghaffari N, Ivanov I, Qian X, Dougherty ER: **A CoD-based reduction algorithm for designing stationary control policies on Boolean networks.** *Bioinformatics* 2010, **26**(12):1556–1563.
105. Qian X, Ghaffari N, Ivanov I, Dougherty ER: **State reduction for network intervention in probabilistic Boolean networks.** *Bioinformatics* 2010, **26**(24):3098–3104.
106. Kobayashi K, Hiraishi K: **An integer programming approach to control problems in probabilistic Boolean networks.** In *Proc. American Control Conference (ACC): 30 June - 2 July; Baltimore, MD, USA*: American Automatic Control Council; 2010:6710–6715.
107. Kobayashi K, Hiraishi K: **Optimal control of context-sensitive probabilistic Boolean networks using integer programming.** In *Proc. 49th IEEE Conference on Decision and Control (CDC): 15-17 December; Atlanta, Georgia, USA*: IEEE; 2010:7507–7512.
108. Kobayashi K, Hiraishi K: **An integer programming approach to optimal control problems in context-sensitive probabilistic Boolean networks.** *Automatica* 2011, **47**(6):1260–1264.
109. Kobayashi K, Hiraishi K: **Optimal control of probabilistic Boolean networks using polynomial optimization.** *IEICE Trans* 2012, **95-A**(9):1512–1517.
110. Ching W, Zhang S, Jiao Y, Akutsu T, Tsing N, Wong A: **Optimal control policy for probabilistic Boolean networks with hard constraints.** *IET Syst Biol* 2009, **3**(2):90–99.
111. Cong Y, Ching W, Tsing N, Leung H: **On finite-horizon control of genetic regulatory networks with multiple hard-constraints.** *BMC Syst Biol* 2010, **4**(Suppl 2). article S14.
112. Chen X, Akutsu T, Tamura T, Ching W: **Proc. IEEE International Conference on Bioinformatics and Biomedicine: 18-21 December; Hong Kong, China.** IEEE Computer Society; 2010:240–246.
113. Liu Q, Guo X, Zhou T: **Optimal control for probabilistic Boolean networks.** *IET Syst Biol* 2010, **4**(2):99–107.
114. Li F, Sun J: **Controllability of probabilistic Boolean control networks.** *Automatica* 2011, **47**(12):2765–2771.
115. Pal R, Datta A, Dougherty ER: **Optimal infinite horizon control for probabilistic Boolean networks.** *IEEE Trans Signal Process* 2006, **54**(6):2375–2387.
116. Pal R, Datta A, Dougherty ER: **Robust intervention in probabilistic Boolean networks.** *IEEE Trans Signal Process* 2008, **56**(3): 1280–1294.
117. Vahedi G, Faryabi B, Chamberland J, Datta A, Dougherty ER: **Intervention in gene regulatory networks via a stationary mean-first-passage-time control policy.** *IEEE Trans Biomed Eng* 2008, **55**(10):2319–2331.
118. Qian X, Ivanov I, Ghaffari N, Dougherty ER: **Intervention in gene regulatory networks via greedy control polices based on long-run behavior.** *BMC Syst Biol* 2009, **3:** article 61.
119. Faure A, Naldi A, Chaouiya C, Theiffry D: **Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle.** *Bioinformatics* 2006, **22**(14):124–131.
120. Vahedi G, Faryabi B, Chamberland J, Datta A, Dougherty ER: **Optimal intervention strategies for cyclic therapeutic methods.** *IEEE Trans Biomed Eng* 2009, **56**(2):281–291.
121. Yousefi MR, Datta A, Dougherty ER: **Optimal intervention strategies for therapeutic methods with fixed-length duration of drug effectiveness.** *IEEE Trans Signal Process* 2012, **60**(9): 4930–4944.

122. Qian X, Dougherty ER: **Intervention in gene regulatory networks via phenotypically constrained control policies based on long-run behavior.** *IEEE/ACM Trans Comput Biol Bioinf* 2012, **9:**123–136.
123. Friedman N, Linial M, Nachman I, Peér D: **Using Bayesian networks to analyze expression data.** *J Comput Biol* 2000, **7**(3–4):601–620.
124. Dojer N, Gambin A, Mizera A, Wilczyński B, Tiuryn J: **Applying dynamic Bayesian networks to perturbed gene expression data.** *BMC Bioinformatics* 2006, **7:**249.
125. Needham C, Manfield I, Bulpitt A, Gilmartin P, Westhead D: **From gene expression to gene regulatory networks in Arabidopsis thaliana.** *BMC Syst Biol* 2009, **3**(85):1–17.
126. Ferrazzi F, Engel F, Wu E, Moseman A, Kohane I, Bellazzi R, Ramoni M: **Inferring cell cycle feedback regulation from gene expression data.** *J Biomed Inf* 2011, **44:**565–575.
127. Yu H, Zhu S, Zhou B, Xue H, Han J: **Inferring causal relationships among different histone modifications and gene expression.** *Genome Res* 2008, **18:**1314–1324.
128. Ciaccio M, Wagner J, Chuu C, Lauffenburger D, Jones R: **Systems analysis of EGF receptor signaling dynamics with Micro-Western arrays.** *Nature Methods* 2010, **7**(2):148–155.
129. Cursons J, Hurley D, Angel C, Dunbar R, Crampin E, Jacobs M: **Inference of an in situ epidermal intracellular signaling cascade.** In *Proc. 32nd Annual International Conference of the IEEE EMBS: 1-4 September; Buenos Aires, Argentina*: IEEE Computer Society; 2010:799–802.
130. Yoeruek E, Ochs M, Geman D, Younes L: **A Comprehensive statistical model for cell signaling.** *IEEE/ACM Trans Comput Biol Bioinf* 2011, **8**(3):592–606.
131. Elvitigala T, Singh A, Pakrasi H, Ghosh B: **Bayesian Network approach to understand regulation of biological processes in Cyanobacteria.** In *Proc. Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference: 15-18 December; Shanghai, China*: IEEE Computer Society; 2009:3739–3744.
132. Peelen L, De Keizer N, De Jonge E, Bosman R, Abu-Hanna A, Peek N: **Using hierarchical dynamic Bayesian networks to investigate dynamics of organ failure in patients in the Intensive Care Unit.** *J Biomed Inf* 2010, **43:**272–286.
133. Himes B, Dai Y, Kohane I, Weiss S, Ramoni M: **Prediction of Chronic Obstructive Pulmonary Disease (COPD) in Asthma Patients using electronic medical records.** *J Am Med Inf Assoc* 2009, **16**(3):371–379.
134. Estabragh Z, Kashani M, Moghaddam F, Sari S, Oskooyee K: **Bayesian network model for diagnosis of social anxiety disorder.** In *Proc. IEEE International Conference on Bioinformatics and Biomedicine Workshops: 12-15 Noverber; Atlanta, GA, USA*: IEEE Computer Society; 2011:639–640.
135. Adams L, Khare S, Lawhon S, Rossetti C, Lewin H, Lipton M, Turse J, Wylie D, Bai Y, Drake K: **Enhancing the role of veterinary vaccines reducing zoonotic diseases of humans: Linking systems biology with vaccine development.** *Vaccine* 2011, **29:**7197–7206.
136. Si S, Liu G, Cai Z, Xia P: **Using Bayesian networks and importance measures to indentify tumour markers for breast cancer.** In *Proc. IEEE Conference on Industrial Engineering and Engineering Management (IEEM): 6-9 December; Singapore*: IEEE Computer Society; 2011:1826–1830.
137. Lähdesmäki H, Hautaniemi S, Shmulevich I, Yli-Harja O: **Relationships between probabilistic Boolean networks and dynamic Bayesian networks as models of gene regulatory networks.** *Signal Process* 2006, **86:**814–834.
138. Li P, Zhang C, Perkins E, Gong P, Deng Y: **Comparison of probabilistic Boolean network and dynamic Bayesian network approaches for inferring gene regulatory networks.** *BMC Bioinformatics* 2007, **8** (Suppl 7):S13.
139. Sakhanenko N, Galas D: **Probabilistic logic method and some applications to biology and medicine.** *J Comput Biol* 2012, **19**(3):316–336.
140. Huang S: **Gene expression profiling, genetic networks, and cellular states: an integrating concept for tumorigenesis and drug discovery.** *J Mol Med* 1999, **77**(6):469–480.
141. Nguyen D, Azadivar F: **Early detection of cancer by reguression analysis and computer simulation of gene regulatory rules.** In *Proc. IEEE International Conference on Biomedical Engineering (ICoBE): 27-28 February; Penang, Malaysia*: IEEE Computer Society; 2012:144–148.
142. Zhang Y, Quian M, Ouyang Q, Deng M, Li F, Tang C: **Stochastic model of yeast cell-cycle network.** *Physica D* 2006, **209:**35–39.
143. Kaderali L, Dazert E, Zeuge U, Frese M, Bartenschlager R: **Reconstructing signaling pathways from RNAi data using Probabilistic Boolean threshold networks.** *Bioinformatics* 2009, **25**(17):2229–2235.
144. Sauer U: **Metabolic networks in motion: 13C-based flux analysis.** *Mol Syst Biol* 2006, **62:**1–10.
145. Holland J: *Adaptation in Natural and Artificial Systems*. Cambridge MA: MIT Press; 1992.
146. Schlatter R, Schmich K, Vizcarra IA, Scheurich R, Sauter R, Borner C, Ederer M, Merfort I, Sawodny O: **ON/OFF and beyond - a Boolean model of Apoptosis.** *PLOS Comput Biol* 2009, **5**(12):e1000595.
147. Gonçalves E, Bucher J, Ryll A, Niklas J, Mauch K, Klamt S, Rocha M, Saez-Rodriguez J: **Bridging the layers: towards integration of signal transduction, regulation and metabolism into mathematical models.** *Mol Biosyst* 2013, **9**(7):1576-83.
148. Terfve C, Cokelaer T, Henriques D, MacNamara A, Goncalves E, Morris M, Van Lersel M, Lauffenburger D, Saez-Rodriguez J: **BMC Syst Biol.** *Volume 6*; 2012:133.
149. Collins F, Morgan M, Patrinos A: **The human genome project: lesson from large-scale biology.** *Science* 2003, **300**(5617):286–290.
150. Dewey F, Pan S, Wheeler M, Quake S, Ashley E: **DNA sequencing - clinical applications of new DNA sequencing technologies.** *Circulation* 2012, **125:**931–944.
151. Sethi P, Theodos K: **Translational bioinformatics and healthcare informatics: computational and ethical challenges.** *Perspect in Health Inf Manage* 2009, **6:**. 1h.
152. Rzhetsky A, Koike T, Kalachikov S, Gomez SM, Krauthammer M, Kaplan SH, Kra P, Russo JJ, Friedman C: **A knowledge model for analysis and simulation of regulatory networks.** *Bioinformatics* 2000, **16**(12):1120–1128.
153. Faryabi B, Vahedi G, Chamberland J, Datta A, Dougherty E: **Optimal constrained stationary intervention in gene regulatory networks.** *EURASIP J Bioinf Syst Biol* 2008, **620767:**1–10.
154. Goh K, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL: **The human disease network.** *Proc Nat Acad Sci* 2007, **104**(21):8685–8690.
155. Gore J: **Principles and practice of function MRI of the human brain.** *J Clin Invest* 2003, **112**(1):4–9.
156. Davie C: **A review of Parkinson's disease.** *British Med Bull* 2008, **86**(1):109–127.
157. Del Sol A, Balling R, Hood L, Galas D: **Diseases as network perturbations.** *Curr Opin Biotechnol* 2010, **21:**566–571.
158. Del Moral P: *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Berlin Germany: Springer; 2004.
159. Moral PD, Kallel L, Rowe J: **Modeling genetic algorithms with interacting particle systems.** In *Theoretical Aspects of Evolutionary Computing*. Berlin Germany: Springer; 2001:10–67.
160. Moral P, Miclo L: **Branching and interacting particle systems approximations of feynman-kac formulae with applications to non-linear filtering.** In *Seminaire de Probability XXXIV, Volume 1729 of Lecture Notes in Mathematics*. Berlin Germany: Springer; 2000:1–145.
161. Back T, Fogel DB: *Handbook of Evolutionary Computation* (Michalewicz Z, ed.) London UK: IOP Publishing Ltd.; 1997.
162. Jong KAD: *Evolutionary Computation: A Unified Approach*. Cambridge MA: MIT Press; 2006.
163. Eiben AE, Smith JE: *Introduction to Evolutionary Computing*. Berlin Germany: Springer; 2003.
164. Ashlock DA: *Evolutionary Computation for Modeling and Optimization*. Berlin Germany: Springer; 2006.
165. Spears W, Jong K, Bäck T, Fogel D, Garis H: **An overview of evolutionary computation.** In *Machine Learning: ECML-93, Volume 667 of Lecture Notes in Computer Science*. Berlin Germany: Springer; 1993:442–459.

doi:10.1186/1478-811X-11-46
Cite this article as: Trairatphisan *et al.*: Recent development and biomedical applications of probabilistic Boolean networks. *Cell Communication and Signaling* 2013 **11**:46.

1.4.7. Research activities on probabilistic Boolean networks in the past two years

In the course of two years after our review article was published (from mid-2013 to mid-2015), research activities on the PBN approach still continuously progressed. On the dynamics of PBNs, several novel PBN variants have been proposed. To mention a few, a semi-tensor product approach (a.k.a. probability structure matrix for simplicity) was applied to represent the state transitions in PBNs instead of applying transitional probability matrix as in the traditional approach (Cheng et al., 2014). Context-sensitive stochastic Boolean networks which apply the formulation of instantaneously random PBNs were shown to decrease computational complexity on simulations compared to the ones performed on the whole collection of context-sensitive PBNs (Zhu, Liang, & Han, 2014). In another study, stochastic multiple-valued networks were extended from multiple-valued PBNs to enhance the computational speed on the state transition matrix calculation. Two small applications on TP53-Mdm2 interaction and WNT5A network were also demonstrated in this article (Zhu & Han, 2014). On top of the advancement on computational frameworks, several properties of PBNs' dynamics such as partial stability upon disturbance of the initial states (Z. Li, Song, & Yang, 2014) and the fluctuations of steady-state distributions according to internal noise (Gao, Xu, Wang, & Liu, 2013) were also explored. The outcome of these studies revealed additional properties of the networks that are essential to be considered for the inference and control of PBNs.

In mid-2013, we commented in our review article that it is necessary to quantify the performance of the multiple existing BN and PBN inference algorithms. Following this perspective, Qian and Dougherty proposed a method to evaluate the inference algorithms for gene regulatory networks based on controllability which is considered to be a better objected-based measure compared to the assessment on the goodness-of-fit alone (Qian & Dougherty, 2013). This study compared several inference algorithms such as REVEAL, BIC, MDL and Best-Fit where it was shown that certain algorithm with poorer performance on network inference can inversely have a better performance in controllability. On another perspective, we stated in the conclusion of our review article that novel approaches with lesser computational complexity are required for designing the optimal control of PBNs. This outlook was realised by several studies using new computational pipelines such as the application of hard constraints integrated with integer and dynamic programming (X. C. X. Chen, Akutsu, Tamura, & Ching, 2013), the introduction of control-fixed point designed based on reachability and stability analyses (Zhao & Cheng, 2014), and the application of the semi-tensor product technique combined with optimisation

applying genetic algorithms (Yang, Li, & Chu, 2014). Furthermore, several existing computational software and methods such as PRISM and polynomial optimisation were also shown to be applicable for the verification and control of PBNs (Kobayashi & Hiraishi, 2014). These advancements allow scientists to design and apply optimal control on PBNs in a more effective and efficient way.

Browsing through the applications of PBNs and related frameworks in biomedicine, only a few studies were reported. For example, exact likelihood computation in Boolean networks with probabilistic exponential time delays was applied to infer regulatory signals in the gene network during embryonic stem cell development based on RNA interference dataset (Dümcke et al., 2014). In another study, single-cell RNA profiling data and the structure of the cell-lineage tree were applied to infer cell-fate controlling gene regulatory networks during mouse early development in the PBN framework (H. Chen et al., 2014). On the study related to signalling networks, Kiani and Kaderali introduced dynamic probabilistic threshold networks which utilise dynamic Bayesian network with probabilistic Boolean threshold function to describe the activated state of signalling proteins (Kiani & Kaderali, 2014). Together with the integration of RNA interference data and prior knowledge introduced as posterior distribution, a series of candidate signalling networks with certain probability distributions were inferred. This approach was successfully applied to study the ErbB signalling network where a large portion of known interactions were recovered and novel regulatory mechanisms between signalling molecules were derived.

To summarise, the research activities on the PBN approach in the past two years were still mainly performed on the theoretical surface with very few proof-of-concept examples. Biomedical applications of PBNs, even if a few existed, mainly focused on the inference and analysis of gene regulatory networks. Only the applications from the group of Kaderali (Kaderali, Dazert, Zeuge, Frese, & Bartenschlager, 2009; Kiani & Kaderali, 2014) with probabilistic Boolean threshold functions were applied to signal transduction networks. Nevertheless, the original concept of PBN as initially introduced by the group of Shmulevich (Shmulevich et al., 2002) has not been applied to study the properties within this kind of network so far even if it has a high applicability to be applied to study and analyse many biological and physiological networks as shown in (Trairatphisan et al., 2013). Hence, we take this opportunity to explore the application of the original PBN concept on signal transduction networks in order to unravel their complexity as presented in this dissertation.

Chapter 2

SCOPE AND AIMS OF THESIS

A large array of modelling approaches in systems biology has been applied to study and analyse signal transduction networks in order to gain improved mechanistic insights. Each modelling approach has its inherent advantages and disadvantages. For instance, qualitative models such as Boolean networks allow the description of large-scale networks but they fail to provide quantitative insights. On the other hand, quantitative approaches such as ODE-based models are able to depict the quantitative changes of molecular species over continuous time. However, they require detailed prior knowledge on molecular interactions as well as an extensive set of experimental data to infer kinetic parameters. Such demanding requirements limit the scalability of these approaches to only small to medium-size networks.

Given the trade-off between different modelling approaches, one always has to balance between the simplicity of the frameworks versus their expressiveness. To choose an appropriate modelling approach, several factors should therefore be considered, e.g. the research objective, the scope of the model, and the type and availability of data. For the modelling of signal transduction networks where a quantitative mechanistic analysis on a large-scale biological network needs to be derived based on a limited amount of dataset, e.g. steady-state data, we identify probabilistic Boolean networks (PBNs), a class of quantitative mechanistic modelling approaches with minimal parameterisation, as a promising approach due to their capability in describing quantitative changes of signalling molecules at steady-state in large-scale models.

In this dissertation, our global aim is to apply the PBN approach for the study of signal transduction networks with steady-state measurements. To perform the analysis of signalling networks in the PBN framework, a specific tool for building, optimising and analysing the PBN models based on the integration of prior knowledge and contextualised experimental data was so far not yet available. In addition, the application of probabilistic Boolean networks to signal transduction networks was not investigated. Furthermore, the potential applications of the PBN approach in a pharmaceutical setting which link biological sciences to clinical medicine were not yet explored either.

In order to account for these challenges in applying the PBN approach for the study of signal transduction networks, we formulate three specific aims of this dissertation as follows:

1. To develop a computational tool for the optimisation and analysis of PBNs.

2. To apply the PBN approach to analyse cellular signal transduction networks.

3. To explore potential applications of PBNs in a pharmaceutical setting.

Chapter 3

MATERIALS AND METHODS

Summarised workflow of the dissertation

In this dissertation, we address the three specific aims as presented in the scope and aims of the thesis in a successive order. These aims include the development of a novel computational tool for PBNs, the application of PBNs for the analysis of cellular signal transduction networks, and the exploration of potential PBNs' applications in a pharmaceutical setting, respectively. A brief summary of the applied methods can be found in Chapter 3.1 to 3.4 and a full description is contained in the respective publications and manuscripts in the results section (see Chapter 4).

To account for the first specific aim which is to develop a novel computational tool for PBNs, we present *optPBN* which allows for the building of PBN models based on literature knowledge and for the optimisation of PBNs against steady-state experimental data. Once the *optPBN* pipeline is established, we account for the second specific aim which is to apply PBNs for the analysis of cellular signal transduction networks by subsequently applying *optPBN* to build and analyse a signal transduction network in a physiological condition. For this task, we analyse a large-scale apoptotic network of primary hepatocytes as a case study. We also consider the optimised selection probabilities in PBNs as weights of interaction together with their distributions which allows to assess the relevance of interactions within the context of the study. A detailed representation of *optPBN* and the case study on the apoptotic network are presented in Chapter 4.1.

To further address the second specific aim, we also apply the PBN framework to study two deregulated signal transduction networks in a pathological context of cancer. In the first study, we apply *optPBN* to analyse a cancer-specific deregulated PDGF signalling network in gastrointestinal stromal tumour (GIST) presented in Chapter 4.2. Initially, we fit a literature-derived PDGF signalling network to single perturbation experiments including genetic point mutations or signalling inhibitors. Then, we analyse the importance of crosstalk interactions between intracellular signalling pathways and we perform an additional set of experiments to validate the modelling results. Subsequently, we revise the model structure based on the new findings from the additional experiments and we

apply the revised model to predict the signalling profiles of combined perturbation experiments to evaluate its predictive power. This study shows that *optPBN* can be applied to identify relevant crosstalk interactions in the context of GIST and the revised PBN model has an excellent performance to predict the signalling profiles of combinatorial treatments.

In addition to the PBN study on deregulated PDGF signalling in GIST, we also apply an improved version of the *optPBN* toolbox to analyse the deregulated L-plastin signalling via the ERK/MAPK pathway in four breast cancer cell lines presented in Chapter 4.3. We firstly fit a literature-derived L-plastin signalling network to a large set of experimental data which include the profiles of signalling molecules after applying various activators and inhibitors. Then, we apply a knock-out study by removing specific interactions from the fitted model and re-perform the optimisation to observe new fitting costs which in turn allows to evaluate the relevance of the respective interactions in the network. In addition, we perform a bootstrapping study by optimising the fitted model against 100 newly generated experimental data derived from the distributions of the original dataset in order to assess the robustness of optimised parameters. This study shows that only 2 out of 5 investigated interactions are relevant within the context of breast cancer and the optimised parameters, which are considered as weights of interaction, are robust against variability within the observed dataset.

Once we could demonstrate the applicability of the PBN framework for the analysis of signal transduction networks in physiological and pathological contexts, we explored if the PBN framework can also be applied in a pharmaceutical setting as stated in the third specific aim of this dissertation. Prior to this exploration, I present a part of my work during a 6-month internship at Merrimack Pharmaceuticals (Boston, MA, USA) in Chapter 4.4.1 where I applied a mechanistic ODE-based modelling approach to analyse a deregulated c-MET signalling network in solid tumours. I fitted the ODE-based model to the time-course data of downstream signalling molecules and I applied the fitted model to assess the effectiveness of a therapeutic agent *OA-5D5*. From the study, I learned that detailed mechanistic models such as ODE-based or PDE-based models are commonly used in a pharmaceutical setting and they often require high computational power to perform the optimisation and analysis.

As we previously demonstrated that the PBN framework can be applied to assess the relevance of network interactions which allows for network contextualisation, we foresee that we could also apply PBNs to pre-select a suitable network topology prior to

the modelling of mechanistic models which in turn helps to reduce the computational demand. We explored this potential application in Chapter 4.4.2 by applying *optPBN* to pre-select a suitable model structure from a small example model with 10 uncertain interactions within the network. In parallel, we systematically built a total of 1024 ODE-based model variants based on the same network topology and evaluated their fitting costs and optimisation time. Then, we compared the results and subsequently found that the model topology pre-selected by *optPBN* matched the minimal model topology of ODE-based model still fitting well to the data.

To summarise, we address the three specific aims of this dissertation by 1) developing *optPBN* as a novel optimisation and analytical tool for PBNs, 2) demonstrating the applications of the PBN framework on the study of cellular signal transduction networks both in physiological and pathological contexts, and 3) demonstrating a proof-of-concept example showing that the pre-selected network topology by the PBN approach can be used as a promising model structure for the mechanistic modelling.

3.1. Summarised materials and methods for Chapter 4.1

Manuscript title: optPBN: an Optimisation Toolbox for Probabilistic Boolean Networks

In this software article, we developed the Matlab-based computational toolbox *optPBN* for the optimisation and analysis of probabilistic Boolean networks. The *optPBN* toolbox was written based on the *BN/PBN* toolbox by H. Lähdesmäki and I. Shmulevich (https://code.google.com/p/pbn-matlab-toolbox) which allows for the simulation and inference of BN and PBN models. We extended the functionalities of the toolbox by integrating a set of computational scripts which allows *optPBN* to determine selection probabilities in PBNs based on experimental data and prior knowledge and also to run on parallel computing infrastructure on a grid-based pipeline for faster computation.

To build and perform the optimisation of PBNs, we initially coded literature-derived model structures as a set of Boolean rules with corresponding selection probabilities. Then, steady-state measurements from multiple experimental conditions were integrated with the model into a single integrated optimisation problem which was subsequently solved by global optimisation algorithms. During the optimisation, we evaluated the quality of model fitting by comparing marginal stationary distributions of molecular states in PBNs to steady-state measurement data. The stationary distributions were approximated by applying the Two-State Markov-Chain (TSMC) approach following the methodology in (Shmulevich, Gluhovsky, Hashimoto, Dougherty, & Zhang, 2003) while the ergodicity of PBN's underlying Markov chain was ensured by the introduction of a small perturbation parameter as introduced in (Miranda & Parga, 2007). Subsequent analysis on the distribution of optimised selection probabilities can be performed to derive novel biological insights. The summarised optimisation pipeline is shown in Figure 3 inside the article.

In parallel, we built a grid-based pipeline of the *optPBN* toolbox by compiling Matlab scripts into an executable C++ package and integrating it to the parallel computing optimisation platform *ParadisEO* (http://paradiseo.gforge.inria.fr). Two optimisation algorithms, i.e. differential evolution (DE) and evolutionary algorithm (EA) were applied to identify the best optimised parameters from the parallelised cluster on Grid'5000 (https://www.grid5000.fr) in an effective and efficient manner.

3.2. Summarised materials and methods for Chapter 4.2

Manuscript title: A Probabilistic Boolean Network Approach for the Analysis of Cancer-specific Signalling: a Case Study of Deregulated PDGF Signalling in GIST

In this research article, we applied our optimisation toolbox *optPBN* (Trairatphisan, Mizera, Pang, Tantar, & Sauter, 2014) to perform model-based data integration which combines literature-based knowledge on the structure of PDGF signalling and steady-state proteomic data in the context of gastrointestinal stromal tumour (GIST).

On the biological part, three HEK cell-based containing PDGFRα mutants with D842V mutation found in GIST (kindly provided by Prof. S. Haan) treated with four signalling molecule inhibitors were investigated by Western blot experiments (partly performed by Dr. C. Bahlawane). Two D842V-PDGFRα mutants additionally contain Y720F or YY731/742FF mutations to abrogate the recruitment sites of SHP2 or PI3K in the MAPK or PI3K/AKT/mTOR pathways, respectively. The four signalling molecule inhibitors includes PI3K/PKC inhibitor Wortmannin, MEK1,2 inhibitor U0126, pan-PKC inhibitor GF109203X, and PKCα/β1-specific inhibitor Gö6976. After the cells were cultured under serum reduced (1%) condition for 14 hours, signalling inhibitors were applied (if applicable) under serum free (0%) condition for 3 hours. Then, the cells were lysed to obtain protein lysates which were subjected to SDS-PAGE and Western blotting. Seven downstream signalling molecules were probed, i.e. PDGFRα, pPDGFRα, pSTAT5, pPLCγ, pPKC substrates, pERK1,2, and pAKT together with alpha-tubulin as loading controls. The Western blot images were then quantified by Image Studio v.4.0. Quantified data after passing quality control were normalised against tubulin, calibrated against calibrator, and normalised to maximum signals across three biological replicates.

On the modelling part, a literature-derived PDGF signalling network (reviewed by Dr. M. Wiesinger) was converted into a PBN model and combined with the quantified Western Blot data to generate integrated optimisation problems. The grid-based version of the *optPBN* toolbox was applied to perform the optimisation on Grid'5000 with 80 cores where the best results from DE algorithm were chosen for further analysis. A single knock-in approach was used to investigate the influence of each crosstalk interaction on model fitting quality after integrating them into the core model topology without crosstalk. The core PBN model and variants were initially trained against a set of single perturbation conditions comprising PDGFRα mutants or signalling inhibitors to obtain optimised selection probabilities. The final refined PBN model was then evaluated with a validation dataset comprising combined perturbation conditions for its predictive power.

3.3. Summarised materials and methods for Chapter 4.3 (modelling part)

Manuscript title: L-plastin Ser5 phosphorylation in breast cancer cells and in vitro is mediated by RSKs, the downstream effectors of the MAPK-ERK pathway

In this research article, we applied an improved version of the *optPBN* toolbox to build, optimise and analyse the signalling profile of L-plastin signalling via the ERK-MAPK pathway. This work was performed in collaboration with M. Lommel and Dr. E. Schaffner-Reckinger from the Laboratory of Cytoskeleton and Cell Plasticity at the University of Luxembourg.

On the experimental part, a large panel of experimental conditions combining different treatments with activators and inhibitors of the L-plastin signalling pathway were investigated in three biological replicates (experiments performed by Maiti Lommel). Four breast cancer cell lines, i.e. the non-invasive MCF7 and SK-BR-3 cell lines and the invasive BT-20 and MDA-MB-435S cell lines were included in the study. The activators of the L-plastin signalling pathway includes PKC-activator phorbol 12-myristate 13-acetate (PMA), PKA-activator 8-Bromo-cAMP, and EGFR-activator EGF ligand, while the group of inhibitors consists of pan-PKC/RSK-inhibitor GF109203X (GF), PKA/RSK-inhibitor H89, Src-inhibitor PP2, MEK1,2-inhibitor PD98059 and RSK-inhibitor BI-D1870 (BI). Two downstream signalling molecules which are phosphorylated PKA (pPKA) substrates and phosphorylated L-plastin (pLPL) were detected where total L-plastin (tLPL) was used as a loading control. To prepare a normalised dataset for modelling, signals were normalised to a loading control to account for well-to-well differences and then to the mean per gel to account for gel-to-gel variability. Then, the normalised signals were calibrated to the signals from PMA conditions which serve as a calibrator condition from different sets of experiment. Subsequently, the calibrated data were pooled to generate the final mean and standard deviation values where were finally normalised against maximal mean value.

On the modelling part, the model topology of L-plastin signalling was derived from literature and was integrated together with the experimental data up to 28 conditions from four cell lines applied as inputs (see Figure 5 in the article). The optimisation was performed on a stand-alone machine (Intel CPU Xeon @3.50GHz, 16GB Ram) for the four different model variants based on random initial conditions. Bootstrapping was performed by randomly sampling 100 artificial data sets based on mean and standard deviation as acquired from the wet-lab experiments. Optimisation was subsequently performed 100 times to identify the distribution of the identified selection probabilities.

3.4. Summarised materials and methods for Chapter 4.4

Chapter title: Applications of quantitative modelling in pharmaceutical industry

Chapter 4.4.1: Detailed mechanistic models and applications in pharmaceutical industry

From January to June 2013, I participated in an internship program at Merrimack Pharmaceuticals (Boston, MA, USA, http://www.merrimackpharma.com) as an intern modeller. I applied several data integration techniques which combined quantitative modelling approaches in systems biology and in-house experimental data to analyse the properties of therapeutic targets. In this dissertation, I present a part of my work which demonstrates the construction of a dynamical c-MET signalling model in ODE-based format that was in turn used for assessing the effectiveness of a therapeutic agent *OA-5D5* (formerly under development by Roche/Genentech as *MetMab*).

On the experimental part, two sets of in-house experiment were performed, one to generate a training dataset and another to generate a validation dataset which were used for building and validating the c-MET signalling model, respectively. The training dataset was generated by Emily Pace. In this set, ACHN (renal cancer) cell line was treated with hepatocyte growth factor (HGF) at concentrations from 0.037nM to 9nM over 2 hours where the phosphorylated form of 3 signalling molecules were measured as read-outs, i.e. phosphorylated c-MET (pMET), phosphorylated ERK1,2 (pERK1,2/pMAPK1,2), and phosphorylated AKT (pAKT). In parallel, the validation dataset was generated by Stephen Su. In this set, 3 cancer cell lines, i.e. A549 (adenocarcinomic alveolar epithelial cells), H2170 (squamous cell carcinoma of lung) and SW-620 (metastasised colorectal adenocarcinoma) were pre-treated with c-MET monoclonal antibody (mAb) *OA-5D5* at concentrations from 1pM to 1μM for 2 hours followed by the treatment of HGF at 1nM for 10 minutes where pAKT was measured as a read-out.

On the modelling part, a c-MET signalling model was built based on the network topology as proposed in literature (Schoeberl et al., 2009, 2002). The model was initially fitted to the training dataset and to additional soft constraints derived from literature in order to obtain optimised kinetic parameters. The optimisation was performed by applying a modified version of the particle swarm algorithms *(confidential)* on an in-house computational cluster. The inhibitory mechanism and kinetic parameters of the mAb *OA-5D5* were incorporated into the optimised c-Met signalling model and the integrated model was subsequently evaluated with the validation dataset. Once the integrated model fitted well to the validation data, it was used to perform *in silico* simulations to predict the

changes of molecular states of downstream signalling molecules within the c-MET signalling pathway in order to derive novel biological knowledge.

Chapter 4.4.2: A PBN-based approach for network pre-selection

To demonstrate that the PBN framework can be applied to efficiently pre-select the network topology prior to the building of mechanistic ODE-based models which are often used in a pharmaceutical setting (see Chapter 4.4.1), we investigated a small example model with an artificial set of measurement data (see Figure 9). The model comprises 11 nodes and 15 interactions where 5 interactions are assigned to be in the canonical pathway (core interactions) while the remaining 10 interactions are yet to be determined if they are relevant in the context of the study. We first applied *optPBN* to identify the weights and to determine the relevance of the interactions. The interactions with optimised weights close to zero are not essential to explain the experimental data and could therefore be removed.

In parallel, we built multiple variants of ODE-based models from the same network topology. Starting from the initial model variant with only 5 canonical interactions, we added combinations of 0 to 10 interactions selected from the pool of all uncertain interactions. We derived 1024 ODE-based model variants applying the law of mass action. We then applied the Simplex algorithm in the Systems Biology Toolbox 2, *SBtoolbox2,* (Schmidt & Jirstrand, 2006) to perform the optimisation of the ODE-based models and recorded the fitting cost and the optimisation time for each model variant. Subsequently, we compared the results from *optPBN* and from *SBtoolbox2* to assess whether the contextualised network topology identified by *optPBN* matches the minimal topology of the ODE-based model that fits well to the same set of measurement data.

Chapter 4
RESULTS

Summarised main findings from all studies

In Chapter 4.1, we built *optPBN* as an optimisation toolbox for probabilistic Boolean networks. In terms of functionality, *optPBN* is the first toolbox which explicitly allows the integration of candidate model structures and normalised experimental data in the PBN framework. Also, the *optPBN* pipeline can be operated on a large-scale computational resource such as a cluster-based or a grid-based infrastructure to optimise large-scale networks in a reasonable timeframe. We applied *optPBN* to analyse a large-scale apoptotic network of (Schlatter et al., 2009) where we demonstrated that the apoptotic model in the PBN framework fitted much better to the normalised experimental data compared to the original study in the BN formalism. The optimisation time for such a large model is also very low on a cluster with 160 parallel computing cores, i.e. less than 5 minutes. In addition, we successfully applied the fitted PBN model to evaluate the relevance of a missing interaction in the network which connects the activation of the apoptotic cascade by Fas ligand to the activation of NFκB. Furthermore, the optimised selection probabilities also provide an initial clue on the weight of interactions within the network and allow the comparison of weights for different interactions in a quantitative scale.

In Chapter 4.2, we applied *optPBN* to analyse a deregulated PDGF signalling network in GIST. We successfully built a literature-derived PDGF signalling network which fitted well to the training set of experimental data comprising single perturbation experiments. Then, we applied the fitted model to analyse the relevance of 9 crosstalk interactions proposed in literature and identified that only 3 crosstalk interactions from PI3K to Ras or to MEK1,2 in the MAPK pathway and from PKC to MEK1,2 are the most relevant interactions in the context of GIST. We subsequently performed an additional set of experiments to validate the modelling results and unexpectedly found that one of the assumptions in our model, i.e. the activation of PKC being depended on the activation of PDGFRα via PLCγ, is wrong. Therefore, we revised the model structure and re-performed the optimisation and analysis and learned that only the on-going crosstalk interactions from PI3K to the MAPK pathways are relevant. Lastly, we evaluated the predictive power of the revised model by predicting the signalling profiles of the combined perturbation

experiments in the validation dataset where our PBN showed an excellent performance to correctly predict 19 out of 20 data points for five downstream signalling molecules.

In Chapter 4.3, we applied an improved version of *optPBN* to analyse the deregulated L-plastin signalling in breast cancer cell lines. We managed to build a literature-derived L-plastin signalling network with a single model topology that fitted well to a broad set of steady-state protein data of four breast cancer cell lines. The results from a knock-out study revealed that fitting costs after removing the interactions from RSK to L-plastin or from PKC to PKA are much higher compared to the fitted model. This finding implied that these two interactions are necessary to be included in the context of the study. In contrast, the removal of other investigated interactions, i.e. from PKC or PKA to L-plastin, and from PKA to PKC did not affect the fitting cost and can practically be removed from the model. These findings also correlate well to a subsequent bootstrapping study where we fitted the model to 100 newly generated datasets derived from the distribution of the original experimental data (see also Chapter 7.3). The identified weights of the two essential interactions (RSK -> L-plastin and PKC -> PKA) are much higher than the other 3 investigated interactions. Also, the distributions of these identified parameters are low. This implied that the identified parameters are robust against variation within the observed dataset.

Besides the main results of the dissertation in Chapter 4.1, 4.2 and 4.3, there are also some interesting findings which are shared in common in the three PBN studies. First, even if the signalling networks that we investigated have several complex features such as feedback loop and crosstalk interactions, we found that all of them can be modelled within the PBN framework in a simplified manner. Second, our literature-derived PBN models can fit well to various sets of steady-state protein data. This finding applied to the models in physiological condition such as the apoptotic model in primary hepatocytes, as well as to the models in pathological contexts including the model of deregulated PDGF signalling in GIST and the model of deregulated L-plastin signalling in breast cancer cell lines. Lastly, the fitted PBN models can be used to assess the relevance of network interactions in the context of the study based on a given set of steady-state measurement. Practically, various approaches can be applied to indicate the relevance of interactions such as the knock-in of interactions (Chapter 4.1 and 4.2) or the knock-out of interactions (Chapter 4.3) with subsequent evaluation of the fitting costs. Also, the values of optimised selection probabilities in PBNs (considered as weights of interaction) can be used as indicators for relevance, i.e. the lower the value is, the less relevant the respective interaction becomes. In addition, the distributions of these parameters indicate

if the identified parameter values are robust against variations in the initial conditions (Chapter 4.1 and 4.2) and/or within the measured dataset (Chapter 4.3).

In Chapter 4.4.1, I demonstrated that detailed mechanistic models such as ODE-based models are often used in the pipeline of pharmaceutical development. I presented a large and detailed ODE-based model of a literature-derived c-MET signalling network where I successfully fitted the model to the signalling profiles of 3 downstream signalling molecules over the time-course of 2 hours. Then, I integrated the inhibitory mechanism of a c-MET antibody *OA-5D5* to the fitted c-MET signalling model and showed that the extended model predicted the changes of pAKT profile at different *OA-5D5* concentrations correctly against wet-lab experimental data. Subsequently, I performed an *in silico* prediction by simulating the inhibitory effect of *OA-5D5* on the signalling profiles of 3 downstream signalling molecules in artificial cell lines where the number of c-MET receptors is varied from 10,000 to 1,000,000 molecules per cell. I compared the predictions from two experimental settings: pre-incubation of *OA-5D5* prior to HGF stimulation versus co-incubation where I found that the inhibitory profiles on pMET and pAKT differ substantially. These results showed that detailed mechanistic models can be applied to identify the best experimental condition to demonstrate the effectiveness of the therapeutic agents of choice and also to compare the inhibitory profiles of multiple therapeutic agents. One remark to be mentioned is that the optimisation and analysis of this large-scale ODE-based model required large computational resources which are not always available and this limitation subsequently impeded the progression of pharmaceutical development to some extent.

As we demonstrated that the PBN framework can be applied to assess the relevance of network interactions, it can therefore be applied to perform a pre-selection of the network topology prior to detailed mechanistic modelling in order to lessen computational demands. In Chapter 4.4.2, we showed in a small example model with 10 uncertain interactions that the network pre-selected by *optPBN* matched the minimal network topology of the ODE-based model still fitting well to experimental data. Such results demonstrated that PBN is a promising modelling framework to be further explored for contextualising network structures of mechanistic models in an actual pharmaceutical setting.

4.1. *optPBN*: an optimisation toolbox for probabilistic Boolean networks (Trairatphisan et al., 2014)

Preface: A number of computational tools for the inference and analysis of biological networks in Boolean formalism were proposed but only a few of them extended their functionalities into a quantitative scale. In parallel, the computational time of the tools which allows for quantitative optimisation does not usually scale which poses a challenge for the analysis of large-scale biological models.

To overcome such limitations, we developed a novel computational tool in the probabilistic Boolean network (PBN) framework which allows for capturing the change of molecular states quantitatively. Here we present *optPBN*, an optimisation and analytical toolbox for probabilistic Boolean networks that allows for determining selection probabilities in PBNs based on experimental data and prior knowledge. Subsequent analysis on optimised selection probabilities of PBNs in turn reveal new insights into biological systems such as the relevance of network interactions within the context of the study. Optimisation of large-scale networks can be performed on parallel computing infrastructures applying the grid-based version of the *optPBN* pipeline where the computational time can be scaled according to the available computational resources.

In this article, we present several toy models to illustrate the functionality of the *optPBN* toolbox including the selection of suitable Boolean rules in Boolean networks (BNs) and the identification of selection probabilities in PBNs. We compared the results from our tool to a Boolean network optimiser *CellNOptimiser* (http://www.cellnopt.org) where *optPBN* could identify the same network topology of BNs and it additionally offers quantitative optimisation. Lastly, we applied the grid-based version of the *optPBN* toolbox to optimise a large-scale Boolean network of apoptosis by (Schlatter et al., 2009) where it requires only several minutes to complete the task. A following analysis on the optimised parameters revealed that a novel interaction which was not included in the original network, i.e. the activation of NFκB by Caspase 8 (Imamura et al., 2004) is essential to be integrated in order to generate a contextualised apoptotic model for primary hepatocytes.

OPEN ACCESS Freely available online

PLOS ONE

optPBN: An Optimisation Toolbox for Probabilistic Boolean Networks

Panuwat Trairatphisan[1], Andrzej Mizera[2], Jun Pang[2,3], Alexandru Adrian Tantar[3], Thomas Sauter[1]*

1 Systems Biology Group, Life Sciences Research Unit, University of Luxembourg, Luxembourg, Luxembourg, **2** Computer Science and Communications Research Unit, University of Luxembourg, Luxembourg, Luxembourg, **3** Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, Luxembourg, Luxembourg

Abstract

Background: There exist several computational tools which allow for the optimisation and inference of biological networks using a Boolean formalism. Nevertheless, the results from such tools yield only limited quantitative insights into the complexity of biological systems because of the inherited qualitative nature of Boolean networks.

Results: We introduce *optPBN*, a Matlab-based toolbox for the optimisation of probabilistic Boolean networks (PBN) which operates under the framework of the *BN/PBN toolbox*. *optPBN* offers an easy generation of probabilistic Boolean networks from rule-based Boolean model specification and it allows for flexible measurement data integration from multiple experiments. Subsequently, *optPBN* generates integrated optimisation problems which can be solved by various optimisers. In term of functionalities, *optPBN* allows for the construction of a probabilistic Boolean network from a given set of potential constitutive Boolean networks by optimising the selection probabilities for these networks so that the resulting PBN fits experimental data. Furthermore, the *optPBN* pipeline can also be operated on large-scale computational platforms to solve complex optimisation problems. Apart from exemplary case studies which we correctly inferred the original network, we also successfully applied *optPBN* to study a large-scale Boolean model of apoptosis where it allows identifying the inverse correlation between UVB irradiation, NFκB and Caspase 3 activations, and apoptosis in primary hepatocytes quantitatively. Also, the results from *optPBN* help elucidating the relevancy of crosstalk interactions in the apoptotic network.

Summary: The *optPBN* toolbox provides a simple yet comprehensive pipeline for integrated optimisation problem generation in the PBN formalism that can readily be solved by various optimisers on local or grid-based computational platforms. *optPBN* can be further applied to various biological studies such as the inference of gene regulatory networks or the identification of the interaction's relevancy in signal transduction networks.

Citation: Trairatphisan P, Mizera A, Pang J, Tantar AA, Sauter T (2014) optPBN: An Optimisation Toolbox for Probabilistic Boolean Networks. PLoS ONE 9(7): e98001. doi:10.1371/journal.pone.0098001

Editor: Lars Kaderali, Technische Universität Dresden, Medical Faculty, Germany

Received October 29, 2013; **Accepted** April 27, 2014; **Published** July 1, 2014

Funding: We acknowledge with thanks financial support from the Fonds National de la Recherche (FNR) Luxembourg (grant 1233900), http://www.fnr.lu. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: thomas.sauter@uni.lu

Introduction

The Boolean network (BN) modelling framework was first introduced by Kauffmann in 1969 for the study of gene regulatory networks [1]. It has been widely applied to analyse the dynamics of different biological systems such as the gene regulatory network of the yeast cell cycle [2], T-cell signalling [3], signal transduction in the apoptotic pathway [4] and many more. For an overview see, e.g., [5,6]. Despite its simplicity, the framework has been shown to be capable of modelling large-scale biological networks and providing meaningful biological interpretations, e.g., the attractors can be correlated to different cellular states [7]. Nevertheless, BNs only provide a very limited quantitative insight into biological systems due to their inherent qualitative nature of state and time.

In 2002, the probabilistic Boolean network (PBN) modelling framework was introduced by Ilya Shmulevich and colleagues for the modelling of gene regulatory networks [8]. PBNs combine the rule-based modelling of Boolean networks with uncertainty principles as described by Markov chains [8,9]. The PBN formalism allows multiple Boolean functions to be assigned to a certain node with corresponding selection probabilities. This assignment forms a collection of Boolean networks (so-called *constituent networks*) that are being randomly chosen in accordance with their selection probabilities throughout the course of a simulation of the PBN. A constituent Boolean network determining the state transition of the PBN is randomly chosen at each epoch in an *instantaneously random PBN*, while the transition determining constituent Boolean network remains constant for a period of time until a binary random variable asks for a switch in a *context-sensitive PBN* [10]. Modelling with PBNs provides a quantitative understanding of biological systems. For example, interactive effects (so called *influences*) between certain genes [8] or average activities of certain genes given by steady-state probabilities [7] can be computed and expressed in quantitative terms.

Over the past years, PBNs have been widely applied to study various biological systems. For instance, Yu *et al.* inferred a gene regulatory network of the interferon pathway in macrophages from time-course gene expression data via the calculation of

Coefficient of Determination (CoD) to determine the selection probability of each predictor function [11]. Using a similar approach, Ma *et al.* inferred a brain connectivity network from functional Magnetic Resonance Imaging (fMRI) data where the influence of each brain compartment in patients with Parkinson's disease could be determined [12]. In recent years, Flöttmann *et al.* modelled the regulatory processes taking place during the production of induced pluripotent stem cells by combining the interplay between gene expression, chromatin modification, and DNA methylation [13]. An extensive analysis on the PBN model of Flöttmann *et al.* suggests possible interventions on gene regulation which might be further developed into clinical applications. For more examples, see [6,14,15,16], where, among others, PBN models for the pathogenesis of dengue viral infection and the transcriptional programming during *C. elegans* development, are discussed.

To facilitate the building of BN and PBN models as a representation of biological systems, several computational tools have been developed which can be applied for the optimisation and inference of models in a Boolean formalism. For instance, *CellNetOptimiser (CellNOpt a.k.a. CNO)* by Saez-Rodriguez *et al.* [17] was used for the inference of a signal transduction network from high-throughput sandwich immunoassay data [18]. *Dynamic Deterministic Effects Propagation Networks (DDEPN)* by Bender *et al.* allows for the reconstruction of signalling networks based on time-course experimental data [19]. Lähdesmäki and Shmulevich introduced *BN/PBN toolbox* [20], a Matlab-based toolbox which allows for the simulation, visualisation and analysis of BN and PBN models. *BN/PBN* toolbox also provides a pipeline for network inference in both the BN and PBN formalisms based on experimental measurements such as microarray data. The network inference process is performed via the calculation of the CoD by exploring the error size of a given Boolean function (or so-called *predictor*) compared to data. The state transition probabilities and the influences that determine the interactive effect for each pair of molecules (such as genes) are subsequently calculated. The *BN/PBN toolbox* was initially designed for the inference and analysis of gene regulatory networks [20]. However, it was also applied for the study of different biological systems such as the brain connectivity network as previously mentioned [12].

Based on the existing functionalities of the *BN/PBN toolbox*, we introduce *optPBN*, a Matlab-based optimisation toolbox for probabilistic Boolean networks. *optPBN* allows for a simple generation of PBN models from rule-based Boolean modelling. Prior biological knowledge such as known interactions in the network, which was not considered in the original *BN/PBN* toolbox, can additionally be integrated as inputs in terms of Boolean rules. *optPBN* facilitates the incorporation of experimental data to BN/PBN models in order to generate an integrated optimisation problem which can subsequently be solved by various optimisation solvers. In comparison to the *BN/PBN* toolbox, *optPBN* extends the functionality by allowing the identification of suitable Boolean rules in BNs and the determination of selection probabilities in PBNs based on experimental data and prior knowledge. Even though the optimisation pipeline of *optPBN* is rather simple and straight-forward, the results generated from *optPBN* retain meaningful qualitative and quantitative biological interpretations which are in accordance with the observed biological phenomena as captured in the experimental data.

In terms of functionality, *optPBN* can handle optimisation problems of networks characterised by various complexities. We offer a stand-alone version of *optPBN* toolbox which is suitable for solving simple optimisation problems, e.g., for small networks. For solving optimisation problems of complex biological networks such as the Boolean model of apoptosis from Schlatter *et al.* which comprises 86 nodes and 125 interactions [4], we also offer a grid-based optimisation pipeline of *optPBN* toolbox that operates on a large-scale computational platform such as the Grid'5000 [21]. Based on the results obtained from the optimisation of Schlatter's model in the PBN format, we quantitatively identified an inverse correlation between UVB irradiation, nuclear factor kappa-B (NFκB) and Caspase 3 activations, and apoptotic activity which could not be demonstrated in the original article due to the qualitative limitation of the Boolean network framework. In addition, we were able to estimate the relevancy of a newly introduced molecular interaction, i.e., the activation of NFκB by Caspase 8, by considering the value of fitted parameter sets and the sensitivity of parameters as indicated by parameter distributions.

Method and Implementation

Probabilistic Boolean networks

A probabilistic Boolean network (PBN) is a collection of Boolean networks in which a constituent network governs the state (activity) of a node (molecule) for a random period of time before another randomly chosen constituent network takes over [7]. Formally, a probabilistic Boolean network $G(V,F)$ is defined by a set of binary-valued nodes $V=\{x_1,x_2,\ldots,x_n\}$ and a family of sets $F=\{F_1,F_2,\ldots,F_n\}$. For each $i=1,2,\ldots,n$ the set F_i is $F_i=\{f_1^{(i)},f_2^{(i)},\ldots,f_{l(i)}^{(i)}\}$ where $f_j^{(i)}$ $(1\le j\le l(i))$ is a possible Boolean predictor function for the node x_i and $l(i)$ is the number of such predictor functions. A *realisation* of the PBN at a given instant of time is determined by a vector of predictor functions, where the ith element of that vector contains the function selected at that time point for x_i. For a PBN with N realisations there are N possible network transition functions $\boldsymbol{f_1},\boldsymbol{f_2},\ldots,\boldsymbol{f_N}$ of the form $\boldsymbol{f_l}=\{f_{l_1}^{(1)},f_{l_2}^{(2)},\ldots,f_{l_n}^{(n)}\}$, $l=1,2,\ldots,N$, $1\le l_j\le l(j)$, $f_{l_j}^{(i)}\in F_j$ and $j=1,2,\ldots,n$. Each network transition function $\boldsymbol{f_l}$ defines a constituent Boolean network of the PBN. In this way the realisations of the PBN can be identified with the constituent Boolean networks.

Let $c_j^{(i)}$ be the probability that the predictor $f_j^{(i)}$, $1\le j\le l(i)$, which is selected to determine the value of x_i at the next time instance. It follows that $\sum_{j=1}^{l(i)} c_j^{(i)}=1$. The PBN is said to be *independent* if the predictors for all nodes are selected independently of each other. Assuming independence, there are $N=\Pi_{i=1}^{n} l(i)$ constituent Boolean networks of the PBN and the probability governing the selection of a particular network is given by $\Pr(f_l)=c_l=\Pi_{i=1}^{n} c_{l_i}^{(i)}$ for all $1\le l\le N$. Two selection schemes are possible: the selection of the constituent Boolean network takes place at each consecutive time step (instantaneously random PBN) or there is a random variable which governs whether the PBN is updated in accordance with the current Boolean network or a newly selected one (context-sensitive PBN). In both cases the constituent network is chosen according to the selection probabilities c_j, $1\le j\le N$. For further details on PBN, we refer to [6,7] which give a comprehensive overview on probabilistic Boolean networks. An example of a PBN with three nodes is given in Figure 1.

The example model consists of three nodes V = (N1,N2,N3) and the functional classes F1 = $\{f_1^{(N1)}\}$, F2 = $\{f_1^{(N2)}\}$, and F3 = $\{f_1^{(N3)},f_2^{(N3)}\}$. N1 and N2 are inputs, where N1 activates N3 while N2 partly inhibits N3 (40%). The respective truth table is shown in Figure 1C. Once both N1 and N2 are activated (taking a state value of 1), node N3 could either solely be under the

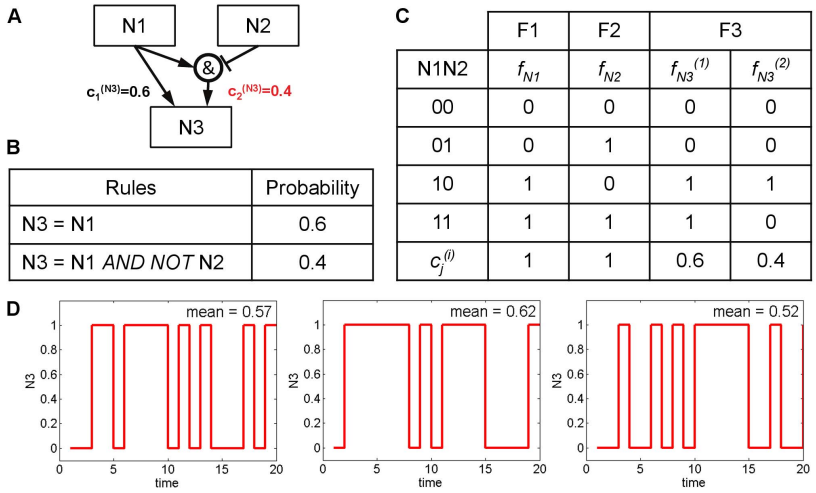


Figure 1. An example model with the corresponding Boolean rules, truth table and model simulation results. [A] The example model consists of 3 nodes with one activation edge and one partial inhibition edge. The weights of both edges are expressed as selection probability next to the arrow. [B] Two representative Boolean rules were assigned with the corresponding selection probabilities ($c_j^{(i)}$) to represent the example model in PBN format. [C] The truth table of the example model demonstrates the state values according to different inputs. Once both inputs (N1 and N2) are active, the output (N3) has a probability of being ON at 0.6 and of being OFF at 0.4 according to the selection probability of Boolean rules. [D] Three separated Monte-Carlo simulations were performed on an instantaneously random PBN of the example model in Figure 1. The state values of N3 are shown on the y-axis as a function of time. The mean of the N3 state values over 20 time steps is given on the upper right corner of each run.
doi:10.1371/journal.pone.0098001.g001

influence of N1 with a probability of 0.6, resulting in the activation of N3 that will take a state value of 1. Or, node N3 could also be under the influence of both N1 and N2 with a probability of 0.4, resulting in the inhibition of N3 that will take a state value of 0. The probabilistic terms that correspond to the selection probabilities ($c_j^{(i)}$) for the Boolean predictor functions are indicated in the truth table. We study this example in the context of instantaneously random PBNs and show three exemplary model simulations in Figure 1D.

There are two constituent Boolean networks of the example model given by the two different Boolean rules for node N3 shown in Figure 1B. These two constituent networks are randomly chosen at each time step of a simulation which for N1 = N2 = 1 results in flips of the state value of N3 between 0 and 1 as shown in Figure 1D.

The dynamics of the PBN is governed by a Markov chain which structure is presented in Figure 2. The nodes represent the states of the system and the possible transitions between the states are labelled with the respective transition probabilities. The graph of the Markov chain consists of four disjoint parts referred to as A, B, C, and D, respectively. There are four bottom strongly connected components of the graph that correspond to four irreducible subchains of the Markov chain: 000 (part A), 010 (part B), 101 (part C) and 110, 111 (part D). It follows that the Markov chain is not ergodic.

With N1 and N2 set to 1, the dynamics of the resulting PBN is given by part D of the Markov chain in Figure 2D which in fact is an ergodic two-state Markov chain. The steady-state probability for N3 to be active is 0.6. This value can be estimated by taking the mean activity over a Monte-Carlo run as shown in Figure 1D. The respective values obtained for 3 independent runs are 0.57, 0.62 and 0.52. In general, longer runs would result in a better estimation of the steady-state probability value.

From a biological point of view, the steady-state probability of a node being active can be interpreted as mean activity of the respective molecule in a cell population normalised to the maximal observed value. Let us assume that some *a priori* knowledge on the model structure is given in the form of a set of constituent Boolean networks, but the selection probabilities are unknown. The above biological interpretation provides basis for considering inferring the selection probabilities from measurement data from different biological conditions (e.g., different ligand stimulations, mutants, or inhibitor treatments). Once selection probabilities are inferred, the relevancy of Boolean interactions can be determined by the values of selection probability and the parameters sensitivity as indicated by their distribution. Furthermore, selection probabilities can be further used to calculate the influences, which reflect the relative importance of parent molecules on the target molecules in the resulting PBN [8].

optPBN pipeline

optPBN is a Matlab-based toolbox which operates under the framework of the *BN/PBN toolbox* by Lähdesmäki and Shmulevich [20], see File S1 for the toolbox and File S2 for *optPBN*'s examples. *optPBN* extends the existing functionalities of the original toolbox by allowing 1) for an easy BN/PBN models generation procedure allowing to incorporate prior knowledge, 2) for improved model fitting to multiple experimental data, i.e., the optimisation of selection probabilities for different experimental settings, 3) for a subsequent statistical analysis of the optimised parameters, and 4) for a fast computation on grid-based platforms. A simplified pipeline of the optimisation process in *optPBN* is shown in Figure 3 and a detailed explanation of the pipeline and computational scripts can be found in File S3.

The *optPBN* pipeline starts with the generation of a BN/PBN model from a preliminary model structure which is usually derived from literature. This step can be easily done by assigning different

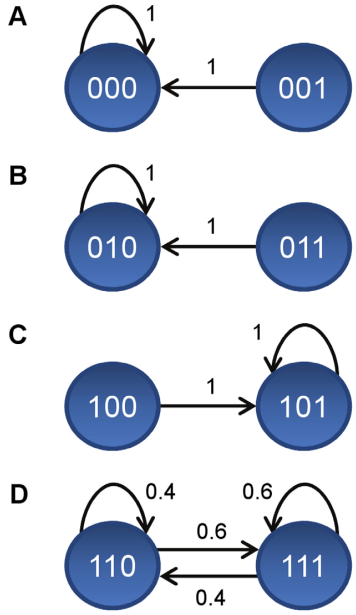


Figure 2. Structure of the dynamics of the example model. The dynamics of the example PBN model presented in Figure 1 is governed by a Markov chain. The graph of the Markov chain consists of four disjoint parts as presented in [A], [B], [C], and [D], respectively. In each graph, the nodes represent the states of the system and the possible transitions between the states are labelled with the respective transition probabilities. Four bottom strongly connected components of the graph that correspond to four irreducible subchains of the Markov chain are shown as follows: 000 [A], 010 [B], 101 [C] and 110, 111 [D].
doi:10.1371/journal.pone.0098001.g002

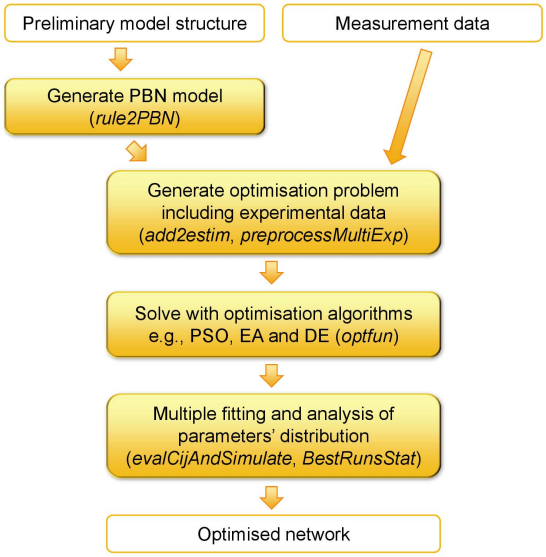


Figure 3. Optimisation pipeline of the *optPBN* toolbox. A preliminary model structure is required as an input for the generation of a PBN model. The generated PBN models from different experimental conditions together with the corresponding measurement data are subsequently combined to generate an integrated optimisation problem which can be solved by various optimisation algorithms. Once the optimisation algorithm(s) generate sufficient amount of good parameter sets, a statistical analysis of the optimised parameter sets (i.e., of PBN's selection probabilities) is performed to indicate the identifiability and the sensitivity of parameters through the consideration on parameters' distribution. The *optPBN* scripts used for each task are given in parentheses.
doi:10.1371/journal.pone.0098001.g003

Boolean interactions in a rule-based Boolean modelling format for each molecule. This means, prior information is considered in terms of a set of possible constituent Boolean functions. For each molecule in a network, single or multiple Boolean functions with the corresponding selection probabilities can be assigned to define how often the respective Boolean function will be present in the chosen constituent network. For unknown or uncertain interaction(s), the selection probabilities of these Boolean rules can be inferred later by optimisation to normalised experimental data.

In the next step, an optimisation problem is generated based on the integration of the preliminary BN/PBN model structure and experimental data. The description of each experimental condition (e.g., different ligand stimulations, mutants, or inhibitor treatments) with its respective measurement data are defined as separate modelling cases. The integration step is simplified by applying the script *rule2PBN* to convert the rules and experimental description of each modelling case into the *BN/PBN* toolbox's internal variables (see the documentation in [20] or File S3) and by subsequently applying the script *add2estim* to collect and combine multiple modelling cases into a single global data structure named *estim*. Following this step, the script *preprocessMultiExp* derives only essential information to generate a final integrated optimisation problem which can subsequently be solved by different optimisation algorithms.

optPBN can be operated in two optimisation modes: 'discrete' and 'continuous'. In the 'discrete' mode each Boolean network from the pool of considered networks is assigned one of two possible values: 0 or 1. Only Boolean networks with value 1 are considered as constitutive Boolean networks of the inferred PBN, each with equal selection probability. In the 'continuous' mode the selection probabilities can be any numbers in the range from 0 to 1 with the only constraint that the sum of selection probabilities of all constitutive Boolean networks of the inferred PBN is 1. Additional details on the two optimisation modes can be found in File S3.

To solve the integrated optimisation problem, two different sets of optimisers are used in *optPBN*: 1) particle swarm optimisation (PSO), *pswarmSB* [22], a global optimisation algorithm as described in the *Systems Biology Toolbox 2 (SBToolbox2)* [23,24], and 2) the evolutionary algorithm (EA) [25] which is integrated in the population-based meta-heuristic optimisation framework *ParadisEO* [26] coupled with a differential evolution algorithm (DE) [27]. We therefore offer two versions of *optPBN*: a stand-alone version which uses *pswarmSB* and a grid-based version which uses the coupled EA and DE algorithms. For additional details on the pipeline of the grid-based version and the algorithms used, please see File S4.

The stand-alone version of the *optPBN* toolbox (PSO-based) was designed for solving simple optimisation problems, e.g., for small networks, while the grid-based version (EA- and DE- based) was customised to be implemented on a large-scale computational platform such as the Grid'5000 [21] for solving complex optimisation problems. The respective objective function of the optimisation process in *optPBN* pipeline is to minimise the sum of squared errors (SSE) between 1) molecular activities as represented

by their steady-state probabilities and 2) measurement data in each experimental condition. The interface for communication between the integrated optimisation problem in *BN/PBN* toolbox's internal format and the optimisers, e.g., the conversion from sampled parameter values to selection probabilities ($c_j^{(i)}$) in PBN, is provided in the script *optfun* with a set of adjustable parameters to customise the optimisation process (see File S3).

During the optimisation process, we approximate the marginal steady-state distribution of the output nodes by applying the two-state Markov chain method as presented in the study of Shmulevich *et al.* [28]. The ergodicity of the PBN's underlying Markov chain is ensured by the introduction of perturbations controlled by a small perturbation parameter (p) as introduced by Miranda and Parga [29]. The two-state Markov chain method is subsequently applied to determine the number of simulation steps to be discarded before reaching steady-state ('burn-in period', m_0) and the minimal number of time steps (N) required to estimate the marginalised steady-state distribution at a pre-defined accuracy. The accuracy of the steady-state approximation can be adapted by adjusting the parameters (ε, r, and s) as described by Raftery and Lewis [30]. Note that the equations used for the calculation of m_0 and N as presented in [28] and [30] contain two small errors. We present the correct derivations of these two-state Markov chain approach formulas in File S5 and we applied the correct formulas in the *optPBN* pipeline accordingly.

Starting from an initial parameters setting (e.g. $m_0 = 0$ and $N = 100$), we iteratively determine a new pair of values for m_0 and N from the estimated transition probabilities between the two meta-states. If the new value for m_0+N is greater than the previous value, the model is simulated further in order to extend the trajectory to the length given by the new value of m_0+N. Then, the transition probabilities are re-estimated from the last N states in the trajectory and used to calculate new values for m_0 and N. This process is repeated until the new value for m_0+N is not greater than its previous value. Finally, the marginalised steady-state probability is estimated with the frequency with which the corresponding state in the two-state Markov chain was sampled within the N last elements of the obtained trajectory.

We observed in the investigated case studies that at least 5,000 iterations of selection probability ($c_j^{(i)}$) parameter sampling by optimisers for small models ($n<10$) and at least 7,500 iterations for large model ($10<n<100$) are sufficient to get a good fit and to obtain representative parameter sets for further statistical analysis. Note that this only holds for the investigated examples and cannot be generalised for other large-scale models. Once the optimisation process is finished, the best parameter set is reported and the model can be re-simulated with the script *evalCijAndSimulate* in order to check the quality of model fitting by comparing simulated steady-state probability to measurement data (see detailed explanation in File S3). Note that under the stochastic events of constituent networks chosen during PBN simulations, the same exact result might not be observed from the re-simulation. Nevertheless, the differences of the results between each simulation are expected to be minimal based on the assumption that the approximation of the steady-state distribution with the two-state Markov chain approach is rather accurate.

After checking model fitting, a representative set of parameters which fit well to measurement data can be chosen for further statistical analysis. The calculation of mean and standard deviation (SD) of the selected set of parameters can be performed by applying the script *BestRunsStat*. The mean of selection probabilities from selected parameter sets indicates to some extent what are the expected selection probabilities for the potential constitutive Boolean networks in PBNs that fit the experimental data. In parallel, the SD value gives an insight on the identifiability for each parameter and parameters' sensitivity can be assessed from parameter distributions. These pieces of information in turn allow for the estimation of the relevancy of Boolean interactions within the context of the study.

Results and Discussion

In this section, four case studies with different levels of complexity are presented to demonstrate the functionalities of *optPBN*. We applied the *optPBN* pipeline, which approximates marginalised steady-state distribution with simulation of ergodic PBNs coupled with the two-state Markov chain method, to generate the results in this section. The parameters for checking steady-state convergence are set as follows: $p = 0.001$, $r = 0.025$, $\varepsilon = 0.01$ and $s = 0.95$.

For each case study, we consider the best 500 parameter sets in terms of the optimal cost to analyse the identifiability of the model parameters and to perform subsequent statistical analyses. The spread of the identified parameters for each case study is shown in Figure 4. The scatter plots show that the obtained parameters are clustered for the first three case studies. However, the parameters in case study 4 are not always clustered. Therefore, we demonstrated the result generated from the best run of each case study which was marked as a red asterisk on Figure 4. A summary of these results can be found in File S6.

Case study 1: optPBN allows for the identification of suitable Boolean rule(s) in Boolean networks

With respect to optimising qualitative Boolean networks, *optPBN* is capable of identifying a set of suitable Boolean rules from a user-defined list of candidate rules based on experimental data. For this task, *optPBN* is operated in the 'discrete mode' for which only 0 and 1 values for the selection probabilities are evaluated (see detail in File S3).

In order to demonstrate the respective functionality of *optPBN*, we use case study 1 (Figure 5) as an example. We pre-define a set of five different Boolean rules to represent the potential influence of PI3K and TNFa on NFkB as follows: connect PI3K and TNFa activation with an OR gate (|), connect PI3K and TNFa activation with an AND gate (&), has only an activation from PI3K (PI), has only an activation from TNFa (TN), and has no interactions from either PI3K or TNFa and output is fixed to 0 (Ø). We consider 4 experimental measurements of NFkB, each with a configuration of the input nodes PI3K and TNFa. For each individual measurement, we applied *optPBN* to determine which of the pre-selected rules(s) is capable of explaining the experimental data. The obtained results show that *optPBN* could identify the correct Boolean rules. Then, we applied *optPBN* to all experimental measurements considered simultaneously. In this case, *optPBN* identified the connection of PI3K and TNFa to NFkB with an AND gate (&) as the only suitable Boolean rule which can explain the complete set of experimental data. This case study shows that *optPBN* can be applied for the inference of biological networks in the Boolean formalism. The obtained results are summarised in Table 1.

Case study 2: optPBN allows for the determination of selection probabilities in probabilistic Boolean networks

In this theoretical case study, we consider the regulation of PIP3 by PI3Kand PTEN. We assume that this process can be modelled with the network presented in Figure 1A, where the nodes N1, N2, and N3 represent PI3K, PTEN, and PIP3, respectively. Nodes N1

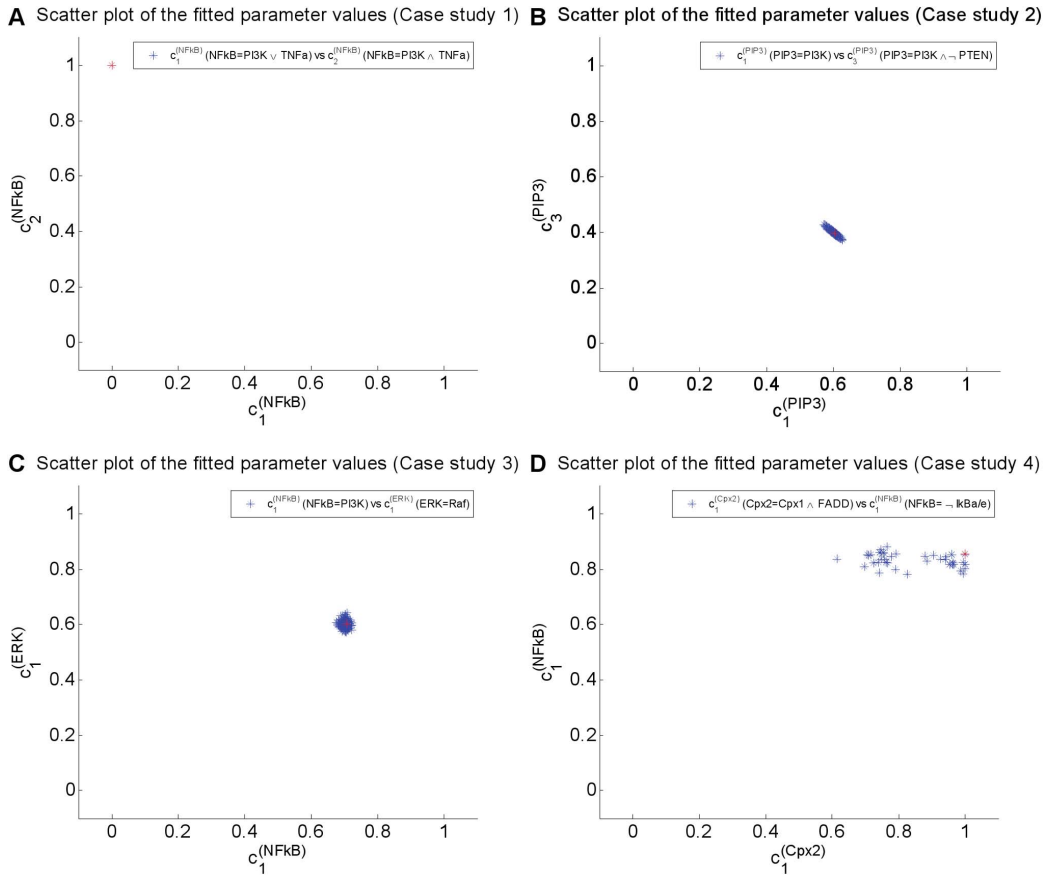
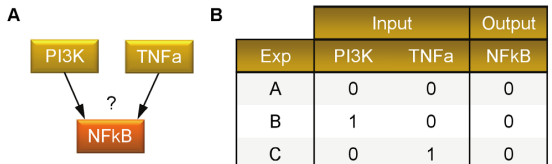


Figure 4. Scatter plots of a set of fitted parameters from all case studies. The distributions of selection probabilities from the best 500 parameter sets in term of optimal cost are shown in [A] for case study 1 (node NFkB), [B] for case study 2 (node PIP3). The dependency among selection probabilities across two nodes are shown in [C] for case study 3 (nodes NFkB and ERK) and [D] for case study 4 (nodes NFkB and complex2). The parameter values for the first 3 case studies form a single cluster which indicates that the respective parameters are identifiable. However, the parameter which influences on NFkB (y-axis) seem to be identifiable in case study 4 but the parameter which influences on complex2 (x-axis) are much sparser. Such observation raises an issue in term of parameters' identifiability. Only the best parameter set from each case study (marked as a red asterisk) was therefore used for further analysis and interpretation.
doi:10.1371/journal.pone.0098001.g004



	Input		Output
Exp	PI3K	TNFa	NFkB
A	0	0	0
B	1	0	0
C	0	1	0
D	1	1	1

Figure 5. Case study 1. [A] Case study 1 deals with a Boolean network that consists of 3 nodes with an unknown Boolean interaction from the two inputs. [B] The table contains artificial experimental data from four different combinations of input states (Experiments 'A', 'B', 'C', and 'D') of case study 1.
doi:10.1371/journal.pone.0098001.g005

and N2 are the so-called input nodes, i.e., they are not influenced by any node in the network and their values are determined by explicit assignment. This makes that the underlying Markov chain consists of four disjoint, non-communicating Markov subchains, one for each of the four different assignments of values to the input nodes.

Let us now assume that the model structure is only partially known, i.e., it is known that PI3K activates PIP3, but there is no certain information on whether PTEN activates or inhibits PIP3 and to what extent. Therefore, as a prior knowledge, we consider four different Boolean rules that encode four potential signal flows from PI3K and PTEN to PIP3 as follows: only activation from PI3K (PI), only activation from PTEN (PT), activation from PI3K and inhibition from PTEN (PI&∼PT), and no interaction from either PI3K or PTEN and output is 0 (Ø). Furthermore, four experiments are performed, where various combinations of values for PI3K and PTEN as the initial conditions are considered. As the measured values of PIP3 we take the theoretical values of the

Table 1. Results from the *optPBN* toolbox for case study 1 compared to the original network.

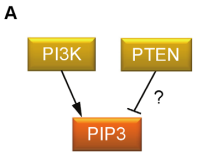
Optimisation results					
Exp|rules	**I**	**&**	**PI**	**TN**	**Ø**
A	✓ (0.2)	✓ (0.2)	✓ (0.2)	✓ (0.2)	✓ (0.2)
B	✗ (0)	✓ (0.33)	✗ (0)	✓ (0.33)	✓ (0.33)
C	✗ (0)	✓ (0.33)	✓ (0.33)	✗ (0)	✓ (0.33)
D	✓ (0.25)	✓ (0.25)	✓ (0.25)	✓ (0.25)	✗ (0)
All (A to D)	**✗ (0)**	**✓ (1)**	**✗ (0)**	**✗ (0)**	**✗ (0)**
Original network					
Correct rule	**✗ (0)**	**✓ (1)**	**✗ (0)**	**✗ (0)**	**✗ (0)**

The table shows the results of optimisation for four different individual datasets (A, B, C and D) and for the combined four datasets (All) compared to the original network. Five different Boolean rules are applied as follows: connect PI3K and TNFa activation with an OR gate (|), connect PI3K and TNFa activation with an AND gate (&), has only an activation from PI3K (PI), has only an activation from TNFa (TN), and has no interactions from either PI3K or TNFa and output is fixed to 0 (Ø). The symbol '✓' indicates that the respective rule can explain the measurement data while the symbol '✗' refers to the contrast observation. The results from the *optPBN* toolbox divide the sum of probabilities, i.e. 1, by the number of correct result(s) in each experiment (given in parentheses) and they all correspond to the correct results.
doi:10.1371/journal.pone.0098001.t001

underlying Markov chain stationary probabilities determined by the initial conditions. The partially known network structure and the experimental data are shown in Figure 6. Now, we applied *optPBN* to perform the optimisation in the 'continuous mode' where an extensive continuous parameter space (the interval from 0 to 1) is explored within the optimisation process to determine the selection probabilities for the four different Boolean rules.

When performing optimisation, the values of the input nodes are fixed to the values specified by the available experimental conditions, one by one. For each experimental dataset only the subchain determined by the experimental condition is considered. In order to make the considered part of the underlying Markov chain ergodic, perturbations are introduced which make the considered subchain irreducible and aperiodic. In this way, the considered part has a unique steady-state probability distribution which can be estimated by the two-state Markov chain approach independently of the choice of the initial state of a simulation. The obtained steady-state probabilities are estimated and the squared difference of the estimated value and the experimental value is calculated. To get the final fit score, the squared differences from all experimental conditions are added.

Two remarks are in place. First, it should be noted that the fit quality of these experiments could be improved by increasing the accuracy for the approximation of steady-state distribution, e.g., by adjusting the parameter 'r'. Details on model fitting's quality in relationship to the accuracy parameter 'r' can be found in File S6.



B

	Input		Output
Exp	PI3K	PTEN	PIP3
A	0	0	0
B	1	0	1
C	0	1	0
D	1	1	0.6

Figure 6. Case study 2. [A] Case study 2 deals with a probabilistic Boolean network that consists of 3 nodes with an unknown type and weight of interaction from PTEN to PIP3. [B] The table contains artificial experimental data from four different combinations of input states (Experiments 'A', 'B', 'C', and 'D') of case study 2.
doi:10.1371/journal.pone.0098001.g006

Second, more importantly, the inference results heavily depend on the experimental data. In this case study the set of experimental data was comprehensive in the sense that it covered possible assignments of values to the input nodes: by considering experiments A, B, C, and D in Figure 6 part [B], all the four non-communicating subchains of the Markov chain in Figure 2 are taken into account. If this is not the case, the inference may result in wrong outcomes. According to the results of this case study shown in Table 2, the selection probabilities inferred from all experiments agree well with the selection probabilities of the original network. However, when only experimental data from experiments A and D were taken into account, the optimisation inferred a PBN consisting of all four constituent Boolean networks given by the rules 'PI', 'PT', 'PI&~PT',and 'Ø', with selection probabilities 0.4602, 0.1344, 0.3607 and 0.0447, respectively.

In summary, this case study demonstrates that the *optPBN* toolbox can be applied to infer selection probabilities from given comprehensive data. Once the selection probabilities are obtained, they can subsequently be used to estimate the relevancy of Boolean interactions. In addition, they can also be used to determine the influence between molecules as presented in [8].

Case study 3: optPBN generates comparable results to an existing tool while having a broader functionality

To date, there are several computational tools which are applicable for the optimisation of biological networks in the Boolean formalism. One of the leading tools is *CellNetOptimizer (CellNOpt* a.k.a. *CNO)* introduced by Saez-Rodriguez *et al.* [17]. CellNOpt was applied for building logic-based models of signal transduction networks in different logic formalisms that are trained against high-throughput proteomics data [18].

To illustrate and prove the functionalities of *CellNOpt*, the tool was used to optimise a toy model based on a set of artificial experimental data. The objective function of *CellNOpt* is based on two components: 1) the mean squared error (MSE) deviation between data and predicted states, and 2) a penalised term for increasing model size ($\dot{E}_s$) which is adjustable by a scaling factor (α). By minimising a combination of these two terms, *CellNOpt* was able to identify the Boolean interactions that correspond to experimental data (see Figure S1).

To benchmark our newly developed toolbox, we applied *optPBN* to optimise the compressed version of the toy model in Boolean

Table 2. Results from the *optPBN* toolbox for case study 2 compared to the original network.

Optimisation results				
Exp|rules	**PI**	**PT**	**PI&∼PT**	**Ø**
A,D	0.4602	0.1344	0.3607	0.0447
All (A to D)	**0.6041**	**0**	**0.3959**	**0**
Original network				
Selection probabilities	**0.6**	**0**	**0.4**	**0**

The table shows the results of optimisation for two datasets: 1) containing measurement data from experiments A and D 2) containing all measurement data from experiments A, B, C and D. Four different Boolean rules are applied as follows: only activation from PI3K (PI), only activation from PTEN (PT), activation from PI3K and inhibition from PTEN (PI&∼PT), and no interaction from either PI3K or PTEN and output is fixed to 0 (Ø). The selection probabilities inferred from all experiments agree well with the selection probabilities of the original network. The dataset consisting of the measurement from only experiments A and D is insufficient to reconstruct the original network.

doi:10.1371/journal.pone.0098001.t002

formalism as presented in the original *CellNOpt* article [18], see also Figure S1. The original toy model comprises 8 nodes. There are 2 input nodes which are TGFa and TNFa with two downstream nodes that can be inhibited by inhibitors (PI3Ki and Rafi). The presence of inhibitor is depicted in the input table with '−' once it is absent and with '+' when it is present. The rest of the nodes are considered as output nodes and are all measured. Here, we applied *optPBN* to optimise the two unknown logic gates for NFkB and ERK.

The original model structure with 6 different experimental conditions were plugged into the *optPBN* pipeline in 'discrete mode' as previously described for case study 1. *optPBN* is capable of acquiring the same results as *CellNOpt*, i.e. to identify 'NFkB = PI3K & TNFa' and 'ERK = Raf' as the correct Boolean rules. In addition, *optPBN* also identified the rule 'ERK = -Raf|NFkB' as an additional solution that can also explain the data from all experimental conditions (see Figure S1). We also verified that both correct Boolean rules are independent as the optimal costs after assigning these two rules one-at-a-time to be the correct rule are highly similar.

Then, we extended the current study by applying *optPBN* for the optimisation of a modified toy model based on a new set of artificial data (case study 3) as shown in Figure 7. In this version, we assumed that the weights of molecular activation and the inhibitors' efficacies are not absolute, resulting in a propagation of signals in a non-discrete (continuous) manner. Once output nodes can be activated by multiple molecules, i.e., NFkB can be activated by PI3K and TNFa while ERK can be activated by Raf and NFkB, we consider disjoint activating signals from both inputs which are sum up to a full activation. When inhibitor is additionally present, the activating signal is reduced proportionally to the inhibitor's efficacy.

Considering e.g. experiment D, both inputs are ON and there are two unknown weights of activation from PI3K and TNFa towards NFkB with the presence of PI3K-inhibitor treatment. If we assume that the activating signal from PI3K is 70% and from TNFa is 30% with the presence of PI3K-inhibitor that inhibits PI3K signal at 70%, the signal for the activation on NFkB in this experiment can be calculated from the sum of the remaining PI3K signal after inhibition (100%−70% = 30%) multiplied by the weight of PI3K's activation (70%), resulted in the signal value of 0.3*0.7 = 0.21. This signal is then combined with the disjoint activating signal from TNFa (30% or 0.3). The sum of activating signals for NFkB node is therefore 0.51 in this experimental setting.

To perform an optimisation study on this modified toy model, *optPBN* was applied in the 'continuous mode' as previously described for case study 2. The optimisation results as shown in Table 3 are in a good agreement with the selection probabilities of the original model.

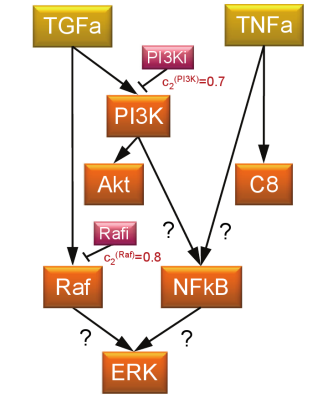
In summary, the results from the two toy model studies demonstrate that the optimised networks generated from *CellNOpt* and *optPBN* are similar when operated in a discrete (qualitative) optimisation mode. At the same time, *optPBN* offers an additional functionality of a continuous (quantitative) optimisation mode to identify selection probabilities which might yield additional insight into the relevancy of interactions within the network.

Case study 4: optPBN allows for the optimisation of an apoptotic network at scalable computational time and for the estimation of interactions' relevancy in a context-specific manner

Optimisation of an apoptotic network in the PBN framework. Schlatter *et al.* introduced a large-scale Boolean network of apoptosis in hepatocytes that consists of 86 nodes and 125 interactions as shown in Figure 8. [4] The assigned Boolean interaction for each molecule was derived from literature. After the Boolean model was built, it was subsequently validated by experimental data which were categorised into three discrete values: no activity '0', low activity '1', and high activity '2'. The analysis of Schlatter's model was conducted in *CellNetAnalyzer (CNA)*, a Boolean network and constraint-based models analyser which allows for the calculation of logical steady-states [31]. As the original model structure comprises many feedback loops, 13 interactions were removed from the model in order to generate a new model variant which delivers fixed point steady-states and thus is compatible to be analysed in CNA. The analysis revealed the effects from different cytokines stimulations and UVB irradiations towards apoptosis in hepatocytes, but only in a limited qualitative manner [4].

Based on the original study, we applied *optPBN* to optimise Schlatter's model in the PBN formalism. We converted the multi-value Boolean model of apoptosis into a binary PBN model which comprises 96 nodes and 106 interaction functions ('initial apoptosis model'). We used the selected set of Boolean interactions as described in the original article with minimal modifications on a few Boolean rules to make them suitable for modelling in the PBN format (see File S2). Our initial aim is to optimise selection probabilities of our PBN model in order to return the steady-state probabilities of 3 output nodes, i.e., Apoptosis, Caspase 3 and

A



B

	Input			
Exp	TGFa	TNFa	PI3Ki	Rafi
A	1	0	-	-
B	0	1	-	-
C	1	1	-	-
D	1	1	+	-
E	1	1	-	+
F	1	1	+	+

C

	Output			
Exp	NFkB	ERK	C8	Akt
A	0.7	0.88	0	1
B	0.3	0.12	1	0
C	1	1	1	1
D	0.51	0.804	1	0.3
E	1	0.52	1	1
F	0.51	0.324	1	0.3

Figure 7. Modified toy model of Saez-Rodriguez *et al.* and corresponding artificial experimental data (case study 3). [A] The modified toy model from Saez-Rodriguez *et al.* [18] is a probabilistic Boolean network that consists of 8 nodes with two unknown weights of Boolean interactions for NFkB and ERK. [B] The table describes the states of inputs and inhibitor treatments for 6 experimental conditions. [C] The corresponding normalised artificial experimental data of the experimental conditions as described in [B]. The 6 experimental conditions based on the combination of stimulus and inhibitor treatments yield different readouts on four downstream molecules.
doi:10.1371/journal.pone.0098001.g007

NFkB, which match the measurement data. We optimised the selection probabilities of Boolean rules for 7 target nodes: IKK, IkBa, IkBe, complex2, caspase8 and caspase3 (both at low and high activities), which are connected to the 3 output nodes. This results in the optimisation of 17 selection probabilities.

Next, pre-processing of the original measurement data on hepatocytes for the three output nodes: Apoptosis, Caspase 3 and NFkB, was performed by background subtraction and normalisation to the maximal value. Saturation of Caspase 3's signals was assumed in our study (see File S7). Then, the normalised experimental data and the PBN model description were combined into an integrated optimisation problem which was subsequently solved with *optPBN* in 'continuous mode'.

For this case study six different experimental conditions which were experimentally validated are given. During optimisation six subchains of the underlying Markov chain are considered, each determined by fixing an input node's value in accordance with the experimental condition. The fixed value of the input node is not perturbed, but all the other nodes can be perturbed, which makes the subchain ergodic and in consequence having a unique steady-state distribution. It should be noted that performing the optimisation with use of a more rich set of experimental data, i.e., for conditions which correspond to setting the input nodes to different combinations of values, could provide more insight into the network interactions.

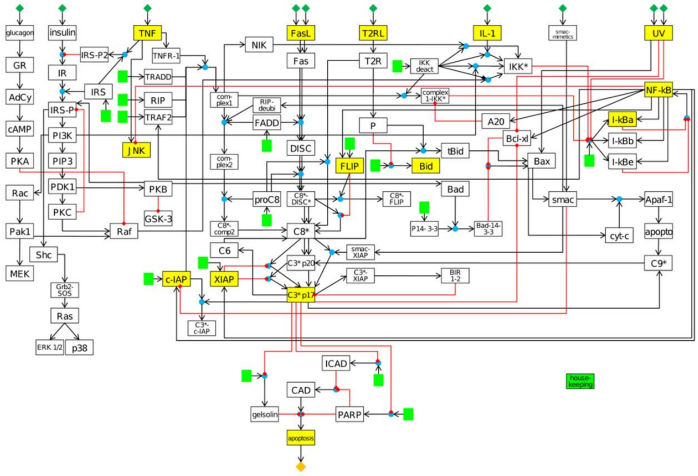
Optimisation of a complex network at scalable computational time. To evaluate the fitting cost during the optimisation process, the marginalised steady-state probabilities for activity of one molecule at a time for a set of output nodes (Apoptosis, Caspase 3 and NFkB) needed to be estimated. As previously presented in the description of the *optPBN* pipeline, this was achieved by applying the two-state Markov chain approach with the accuracy set to 0.025 (r = 0.025). Although the size of the underlying Markov chain of the PBN is huge, i.e., 2^{96} states, it

Table 3. Results from the *optPBN* toolbox for case study 3 compared to the original network.

Optimisation results				
Outputs	**NFkB**		**ERK**	
Exp|rules	**PI**	**TN**	**RA**	**NF**
All (A to F)	**0.708**	**0.292**	**0.603**	**0.397**
Original network				
Selection probabilities	**0.7**	**0.3**	**0.6**	**0.4**

The table shows the optimisation results for the dataset which combines all six experiments from A to F (All) in case study 3. Two different Boolean rules are applied for NFkB and another two rules for ERK as follows: activation signal from PI3K (PI) or TNFa (TN) to NFkB, and activation signal from Raf (RA) or NFkB (NF) to ERK. The results from the optimisation of modified toy model from Saez-Rodriguez *et al.* are corresponded with the correct results.
doi:10.1371/journal.pone.0098001.t003

A



B

Input	Output		
Stimuli	Apoptosis	C3ap17	NFkB
Mock	0.00	0.00	0.00
FasL (2)	0.63	1.00	0.32
IL-1	0.00	0.03	1.00
TNF	0.10	0.39	0.98
UV (1)	1.00	1.00	0.26
UV (2)	0.49	0.00	0.46

Figure 8. Boolean model of apoptosis by Schlatter *et al.* and normalised experimental data (case study 4). [A] The Boolean model of apoptosis by Schlatter *et al.* [4] consists of 86 nodes with 125 Boolean interactions. The model was analysed with *CellNetAnalyzer (CNA)* to study the correlations between 10 different inputs from different cytokines stimulations and UVB irradiations towards apoptosis. [B] The normalised experimental dataset were generated based on the experimental data as presented in the original article (see details in File S7). The inverse correlation between UVB irradiation, NFκB and P17 form of activated Caspase 3 (C3ap17) activations, and apoptotic activity is quantitatively observed in the original measurement data.
doi:10.1371/journal.pone.0098001.g008

Table 4. Run-time analysis of grid-based *optPBN* pipeline.

Model	Iterations	Stand-alone version (1 core)	Grid-based version (80 cores)	Improvement (folds)
1	1000	305s (5m5s)	8s	38.1
	5000	1093s (18m13s)	16s	68.3
2	1000	321s (5m21s)	9s	35.7
	5000	1205s (20m5s)	16s	75.3
3 (modified model)	1000	1302s (21m42s)	18s	72.3
	5000	6151s (1h42m31s)	39s	157.7
4 (extended structure)	1000	16783s (4h39m43s)	99s (1m39s)	169.5
	5000	45503s (12h38m23s)	259s (4m19s)	175.7

The table shows the computational time of the optimisation process required by the grid-based version of *optPBN* pipeline operated on 80 cluster cores in comparison to the one required by the stand-alone version running on a single local machine. The run-time analysis was performed on the four case studies for 1,000 and 5,000 parameter samplings to approximate the steady-state probability distribution of output nodes. The results reveal a remarkable reduction of the computational time from 35 to 175 folds. Abbreviations: s = seconds, m = minutes, and h = hours.
doi:10.1371/journal.pone.0098001.t004

Table 5. Original and *optPBN* results for the Boolean model of apoptosis (case study 4).

	Apoptosis				C3ap17				NFkB			
Exp	Orig.	Init.	Ext.	Meas.	Orig.	Init.	Ext.	Meas.	Orig.	Init.	Ext.	Meas.
Mock	0	0.0000	0.0054	0.00	0	0.0168	0.0082	0.00	0	0.0168	0.0027	0.00
FasL (2)	1	0.9208	0.7170	0.63	1	0.9505	0.8491	1.00	0	0.0198	0.1501	0.32
IL-1	0	0.0028	0.0063	0.00	0	0.0042	0.0013	0.03	1	0.8873	0.8481	1.00
TNF	0	0.0020	0.0018	0.10	0	0.4073	0.4049	0.39	1	0.7889	0.8723	0.98
UV1	1	0.9920	0.9868	1.00	1	0.9966	0.9901	1.00	0	0.0023	0.0033	0.26
UV2	0	0.4681	0.4914	0.49	0	0.0016	0.0012	0.00	1	0.3083	0.2932	0.46

The results from logical steady-state analysis in *CellNetAnalyzer* of the original Boolean model from Schlatter *et al.* [4] (Orig.) are shown in comparison to the *optPBN* results generated from the 'initial apoptosis model' (Init.), a probabilistic Boolean network (PBN) model which was modified from the original Boolean model, and from the 'extended apoptosis model' (Ext.), a PBN model variant with an additional literature-derived Boolean interaction from Caspase 8 to NFκB. The sum of optimal costs based on the calculation of sum of square error between simulated steady-state probabilities and measurement data (Meas.) are 1.002, 0.328, and 0.199, respectively. The inverse correlation between UVB irradiation (UV1 = 300 £/m^2, UV2 = 600 £/m^2), NFκB and P17 form of activated Caspase 3 (C3ap17) activations, and apoptosis are observed in a limited qualitative manner in the original model in Boolean Network format; while this correlation is observed quantitatively in the initial and extended apoptosis models in PBN format. Note that the model fitting of NFκB in the case of Fas ligand activation, 'FasL (2)', in the extended apoptosis model is improved as a result from the additional molecular interaction derived from literature.

doi:10.1371/journal.pone.0098001.t005

turned out that in all three cases (Apoptosis, Caspase 3 and NFκB) the two-state Markov chains were well-mixing: there were frequent transitions between the two meta-states of the two-state Markov chains that were considered on top of the underlying Markov chain of 2^{96} states. In consequence, the marginalised probabilities could be estimated from trajectories of length less than 4000 (see File S6). Given this, we were able to perform the optimisation task in a feasible amount of computational time.

To confirm the accuracy of our results, we performed additional analyses by fixing one set of selection probabilities (randomly generated) and started the simulation from random initial conditions as well as from the extreme cases where initial conditions for all nodes are either zero or one. We found that the variations between these 3 cases are minimal, all less than 0.01 (1%). Also, we generated another set of results with higher accuracy (r = 0.01) where the number of required time steps is increased approximately 6 folds. We found that the steady-state distributions of output nodes which were estimated at the two levels of accuracy (r = 0.025 and r = 0.01) are almost the same, i.e., they differ by less than 0.01 (1%) in all case. This indicates that the chosen parameters lead to a good estimate of the steady-state distribution of the considered model. We also performed the same analysis multiple times for different parameter sets where we obtained similar results. We present a comprehensive datasheet of the analysis performed for one parameter set in File S8.

Due to the large size of the apoptotic network (Figure 8), the stand-alone version of *optPBN* pipeline which runs on a standard local computer (1 CPU Intel@2.99 GHz, 2 cores/CPU, 3.25 GB RAM) is not suitable to solve the optimisation problem. Since the optimisation required 12 hours of computational time to evaluate approximately 5,000 parameter samples on a standard computer, we applied the grid-based version of *optPBN* to solve this optimisation problem on the Grid'5000 using 10 server machines (1 server machine comprises 2 CPUs Intel@1.995 GHz, 4 cores/CPU, 15 GB RAM, see detailed documentation on the installation and execution of the grid-based pipeline in File S4 and File S9). The optimisation process was operated in the 'continuous mode' on 80 parallel processing cores where the results were delivered in a timely manner (approximately 10 minutes to evaluate 15,000 parameter samples). The run-time analysis of four case studies reveals a reduction of computational time from 35 to 175 folds when running the same optimisation tasks on Grid'5000 with 80 cores. More details on this analysis are shown in Table 4.

Approximation of steady-state distributions by optPBN reveals more quantitative insight into biological data. For comparison, the original results from Schlatter's study (Orig.) and the set of results from the optimisation of the 'initial apoptosis model' by *optPBN* (Init.) are shown in Table 5. We were able to identify a PBN model structure with a set of selection probabilities that could match relatively well to the measurement data quantitatively, while the results from the original model are only limited to 0 and 1 value. The fitting costs of the original model and of the 'initial apoptosis model' based on the calculation of SSE are 1.002 and 0.328, respectively. This indicates that the apoptosis model in the PBN formalism fits the measurement data better.

In addition, the inverse correlation between the intensity of UVB irradiation, the activations of NFκB and Caspase 3, and the apoptotic activity were also identified in a quantitative manner. Namely, a stronger UVB irradiation (i.e., UV2 = 600 £/m^2> UV1 = 300 £/m^2) resulted in a stronger NFκB pathway activation (0.3083 against 0.0023) but a weaker Caspase 3 activation (0.0016 against 0.9966) and a weaker apoptotic activity (0.4681 against 0.9920). In contrast, the original study of Schlatter *et al.* could only identify this relationship in a limited qualitative manner [4].

Table 6. The distributions of fitted selection probabilities on the extended apoptosis model.

Molecule	Functions	Mean	SD
NFkB	=(~I_kBa|~I_kBe)	0.8354	0.0209
	=(~I_kBa|~I_kBe)|(C8a|C8a_2)	0.1646	0.0209
complex2	=complex1 & FADD	0.8480	0.1105
	=complex1 & FADD & RIP_deubi	0.1520	0.1105

The Boolean rules of two target molecules in the 'extended apoptosis model' and the parameter distributions of the top 500 fitted selection probabilities in terms of optimal cost are shown. A low standard deviation (SD) value comparing to the corresponding mean value indicates that model is sensitive to the parameter of the respective Boolean interaction. This information highlights the relevancy of the respective interaction within the context of study e.g., the case of nuclear factor kappa-B (NκFB) activation by Caspase 8 (C8a or C8a_2) together with the absence of nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor alpha and epsilon (IκBa and IκBe) inhibitions. In contrast, a high value of standard deviation comparing to the mean value as seen in the case of TNF receptor-1 signalling complex 2 (complex2) activation by deubiquitinated form of receptor associated receptor kinase 1 (RIP_deubi) suggests that the model might not be sensitive to the respective parameter in this study. The relevancy of this interaction within the context of apoptotic signalling in primary hepatocytes is therefore in question.
doi:10.1371/journal.pone.0098001.t006

optPBN allows for an estimation of interactions' relevancy in a context-specific manner. As can be seen in Table 5, the fitted 'initial apoptosis model' (Init.) failed to explain some of the experimental data. For instance, the fitting of NFκB in a condition with a high concentration of Fas ligand stimulation (FasL (2)) is not in a good agreement with the experimental measurement (0.0198 against 0.32). This observation raises a question whether the set of considered molecular interactions was sufficient to model the context-specific apoptotic signalling in the hepatocytes.

In fact, NFκB can also be activated through Caspase 8 with a mechanism distinct from that of tumour necrosis factor alpha (TNFα) for cytokine production as described in [32,33]. Therefore, we modified the Boolean rules to take this information into account and thereby derived a new model called 'extended apoptosis model' (see detail in File S2). The results from the optimisation on the 'extended apoptosis model' (Ext.) are shown in Table 5. We found that the 'extended apoptosis model' fits the experimental data better (cost of 0.199) than the 'initial apoptosis model' (cost of 0.328) and the inverse correlation between UVB intensity, NFκB and Caspase 3 activations, and apoptotic activity is still preserved. In addition, we also found that the discrepancy between the simulated model state and the measurement data of NFκB in the FasL (2) experiment is decreased after the addition of the new molecular interaction derived from literature.

In Figure 4, we presented the best 500 values for selection probabilities $c_1^{(NFkB)}$ and $c_1^{(complex2)}$ in term of fitting cost. By taking 1-$c_1^{(NFkB)}$ and 1-$c_1^{(complex2)}$, we could determine the values for the selection probabilities $c_2^{(NFkB)}$ and $c_2^{(complex2)}$ respectively. The mean and the SD values for these selection probabilities are given in Table 6. These statistics were confirmed in 3 independent optimisation runs with 500 best parameter sets considered in each run (see File S2).

We found that the selection probability $c_2^{(NFkB)}$ for the Boolean rule which represents the co-influence of Caspase 8 and nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor alpha and epsilon (IκBα/IκBε) degradations on the activation of NFκB obtained a mean value of 16% with SD of 2%. The non-zero mean value indicates that this interaction is important to explain the experimental data in the context of our study. Also, the narrow distribution of the values of selection probability for this Boolean rule suggests that the model is sensitive to this parameter. Such information therefore highlights the relevancy of this newly introduced interaction. In contrast, the distribution of the values of the selection probability $c_2^{(complex2)}$ which describes the influence of deubiquitinated form of receptor associated kinase 1 (RIPdeubi) on the activation of TNF receptor-1 signalling complex 2 (complex2) is more spread. This might suggest that the model is less sensitive to this parameter. Therefore, the relevancy of this interaction in the context of our study is still in question.

We tried to further investigate the influence of RIPdeubi on complex2 by following the methodology as presented in [8]. However, we observed that, the required length of the trajectory to estimate the joint distribution on the parent nodes of complex2 is very long and therefore it is practically infeasible to perform this analysis in a reasonable amount of time with the current implementation of the *optPBN* toolbox. Thus, we further estimate the relevancy of this interaction by considering the model topology within the context of our study instead.

Considering the model topology as presented in [4], the Boolean interaction in question represents positive signals on the activation of complex2 by RIPdeubi via two sources (see Figure 8): 1) A positive feedback loop from activated Caspase 8 (C8*) + unknown proteins in type 2 apoptosis (P)→truncated Bid (tBid)→Bax→S-mac→RIPdeubi→complex2→C8*-complex2→C8*, and 2) The positive signal from UVB radiation (UV)→Bax→smac→RIPdeubi→complex2. According to the original study, the activation of type 2 receptor (T2R) was only considered and experimentally validated in the context of Jurkat cells, but not hepatocytes [4]. Therefore, the node P which is activated by T2R will never be activated and the feedback loop as shown in 1) is not active in our PBN model (which only fits to the hepatocyte data). In parallel, the activation of complex2 by UVB (via RIPdeubi) requires the presence of TNF receptor-1 signalling complex 1 (complex1). This means the positive interaction in 2) will only be valid when TNF and UVB co-stimulate. Nevertheless, this condition was not validated in the original study [4] nor was constrained in the PBN model either. Thus, the interactions on complex2 activation from 1) and 2) via RIPdeubi as described are not relevant in the context of primary hepatocyte. These interactions could therefore be removed from the PBN model in our study.

We demonstrated that *optPBN* can be applied to determine the selection probabilities in PBN which can subsequently be used to estimate the relevancy of molecular interactions in the context of study. Such information in turn leads to the generation of computational models which could represent the dynamics of biological networks in a context-specific manner.

Conclusions

We present *optPBN*, a novel optimisation toolbox which provides a simple yet comprehensive pipeline for the generation of integrated optimisation problems in the PBN formalism which can readily be solved by various optimisers on local or grid computational platforms. The *optPBN* toolbox offers two modes of optimisation, discrete and continuous, for the selection of appropriate constituent Boolean networks for the PBN from the pool of available choices and for the determination of selection probabilities from experimental data, respectively. The two modes can be applied for the optimisation and/or inference of the networks in both qualitative and quantitative manner.

optPBN was tested and compared against existing optimisation tools for the Boolean formalisms where *optPBN* delivers similar results and it also offers quantitative optimisation. The updated version of *CellNOpt*, *CellNOptR*, is also capable of handling quantitative optimisation, but the respective Boolean models have to be converted into fuzzy logic or ordinary differential equation (ODE) modelling frameworks beforehand [17,34]. We also demonstrated that *optPBN* allows for the optimisation of an apoptotic network, leading to the generation of an optimised PBN model with the corresponding selection probabilities that fit experimental data. Such results do not only yield a better quantitative insight into biological networks, they can also be further used to evaluate interactions' relevancy within the network in a context-specific manner. Lastly, the computational time which is a major limitation when dealing with complex optimisation problems can be better handled by applying the grid-based implementation of the *optPBN* optimisation pipeline.

Limitations

Even though *optPBN* offers many simple-to-use functionalities, there are some limitations that come along with the simplicity of the toolbox. First, our approach always requires prior knowledge on the possible interactions between molecules which are given in the form of potential constituent Boolean networks. There exist approaches using Bayesian network with Boolean variables for reconstructing biological networks which do not have this limitation, see e.g., [35]. Second, the formulation of the rule-based modelling has to be in a specific order when there is a combination between parameter and constant value in the assignment of different Boolean rules. For instance, given an output which can be activated by an input (a parameter is assigned to the rule) while it can also be inhibited by an inhibitor (the constant value '0' is assigned to the rule), the Boolean rule which represents the activation by an input has to come first (see more examples in the help section of the script *rule2PBN* in the *optPBN* toolbox). Lastly, *optPBN* uses only one global data structure (estim) to store and process all information of the network so that the integrated optimisation problem can be solved simultaneously for all experimental cases. This setting might not be applicable to solve the optimisation problem where different parameter values are expected for each experimental case (i.e., the optimisation of local parameters). The optimisation of such local parameters is not yet available in the current version of the toolbox.

Outlook

First, we aim to improve *optPBN* to be capable of optimising local parameters. In addition, we plan to introduce the concept of penalisation for increasing model size ($\grave{E}_s$ and α) as implemented in *CellNOpt* and *CellNOptR* as a part of our objective function in order to generate better results. Second, we foresee that other global optimisation techniques, e.g., Simulated Annealing, Pattern Search methods, or Mode Hopping Metropolis sampling could be integrated into the *optPBN* pipeline. Third, the ambiguity of Boolean rules formulation to properly represent biochemical reactions is yet to be addressed. Lastly, we envisage many useful applications from implementing the *optPBN* toolbox to study biological systems such as the inference of gene regulatory networks from microarray data and the identification of crosstalk signalling's relevancy in mammalian signal transduction networks based on experimental data in a context-specific manner.

Software Availability and Requirements

- **Project name:** optPBN
- **Project home page:** http://sourceforge.net/projects/optpbn
- **Operating system(s):** Platform independent
- **Programming language**: Matlab (and C++ for grid-based version)
- **Other requirements**: BN/PBN Toolbox, Systems Biology Toolbox 2 (with Message Passing Interface (MPI) and ParadisEO for grid-based version)
- **License**: GNU GPL v3.0 (with CeCILL license required for ParadisEO framework)
- **Restrictions**: no restrictions except for commercial use

Supporting Information

Figure S1 Compared results from *optPBN* and *CellNOpt* on the original toy model of Saez-Rodriguez *et al.* [A] The model structure of the original toy model of Saez-Rodriguez *et al.* [18] is shown on the left panel. The experimental descriptions and the corresponding artificial measurement data are shown in the left and right tables respectively. [B] The results from *CellNOpt* under a defined set of size penalty ($0 \leq \alpha < 0.23$ and $\grave{E}_s = 0.58$) identifies the AND (&) gate for the connection between PI3K and TNFa to NFkB. The interaction from Raf was identified as the only factor that activates ERK. [C] The results from the optimisation with *optPBN* toolbox in discrete mode are in a good agreement with *CellNOpt* for NFkB. Furthermore, *optPBN* also discovered the OR (|) gate for the connection between Raf and NFkB to ERK as an additional solution.
(PDF)

File S1 Stand-alone version of *optPBN* toolbox. The compressed zip file contains all scripts of *optPBN* toolbox in stand-alone version. Also, it contains three scripts of *optPBN* toolbox in grid-based version with the tag '_G5K.m' for the analysis of optimisation results obtained from the grid-based pipeline which runs on a local computer. To install the stand-alone version, a wrap-up script in Matlab (install.m) is provided for the ease of installation. The installations of the *BN/PBN* toolbox [20] and the optimisation toolbox of Systems Biology Toolbox 2 (SBtoolbox2) [24], which are required for the running of *optPBN* toolbox, are also included in the wrap-up script. The grid-based version of the *optPBN* toolbox is provided separately and can be downloaded at https://sourceforge.net/projects/optpbn. It comprises all packages needed for running the algorithm on a cluster or on a grid-based infrastructure. This includes the ParadisEO 1.1 framework (evolutionary algorithms and parallelisation support), MPICH2, LibXML2, GSL, MCR and the *optPBN* grid-based implementation in itself. A detailed description on the installation and execution of the grid-based pipeline can be found in File S9.
(ZIP)

File S2 Computational scripts of all examples and the corresponding original result files. The compressed zip file contains the *optPBN* optimisation pipeline in the form of Matlab scripts (.m) for all examples used in this study (case study 1, case study 2, toy models of Saez-Rodriguez *et al.*, i.e. case study 3, and Boolean model of apoptosis of Schlatter *et al.*, i.e. case study 4). For the results presented in this article, the integrated optimisation problems of the first 3 case studies were optimised by using the stand-alone version of the *optPBN* pipeline applying particle swarm optimisation as the optimiser on a single local machine. The last case study (Schlatter's model, i.e. case study 4), due to a complex optimisation problem, was optimised by using the grid-based version of the *optPBN* pipeline applying a combined differential evolution and evolutionary algorithms as the optimisers on the Grid'5000. The corresponding results from the optimisations of each model presented in the article are included in a matrix format (.mat and .log) for further analysis on the distributions of optimised parameters. A complete set of result files for four case studies and the results from additional analysis as presented in File S8 can be downloaded at https://sourceforge.net/projects/optpbn.
(ZIP)

File S3 Manual of the *optPBN* toolbox. The PDF manual provides a detailed description of the *optPBN* optimisation pipeline. A step-by-step guideline on how to use the *optPBN* toolbox together with the explanation of the core idea for each computational script is provided in the document.
(PDF)

File S4 Grid-based pipeline of *optPBN* toolbox. The PDF documentation provides a description on the grid-based pipeline of the *optPBN* toolbox. It also describes a strategy to combine two optimisation algorithms, evolutionary algorithm (EA) and differential evolution algorithm (DE), as a single optimiser. An optimisation run on Grid'5000 is demonstrated as an example.
(PDF)

File S5 Derivations of the two-state Markov chain approach's formulas. The PDF documentation provides the corrected derivations of the two-state Markov chain approach's formulas in relation to the derivations presented in the original work of Raftery and Lewis [30]. The term λ which refers to $1-\alpha-\beta$ in the calculation of m_0 should be replaced with $|\lambda|$ (i.e., $|1-\alpha-\beta|$). In addition, Φ which denotes the standard normal cumulative distribution function in the calculation of N should be substituted with the inverse of its function, Φ^{-1}, after derivations.
(PDF)

File S6 Results of *optPBN* for four case studies. The spread sheet provides a complete set of results generated from the *optPBN* pipeline for the four case studies. Different numbers of (selection probability) parameter samplings from optimisers and a range of accuracy parameter 'r' from 0.01 to 0.05 are explored. Computation time and quality of model fitting were reported for both stand-alone and grid-based versions.
(XLSX)

File S7 Normalised and justified experimental data for the study of apoptosis model from Schlatter *et al.* The spread sheet provides a detailed description of the measurement data that were used for the optimisation of the Boolean and probabilistic Boolean models of apoptosis from Schlatter *et al.* The pipeline on background subtraction and normalisation of experimental data with the justification on the saturation of the signal for p17 form of activated Caspase 3 is described in detail.
(XLS)

File S8 Analysis of the approximation of steady-state distributions with different initial conditions and accuracies. The spread sheet presents the steady-state probability of output nodes in case study 4 which are approximated by *optPBN* applying a fixed set of selection probabilities (randomly generated) with different sets of initial conditions. Three sets of initial conditions (random, all zeros, and all ones) together with two levels of accuracies ($r = 0.025$ and $r = 0.01$) were explored. Steady-state probabilities of output nodes and parameters of two-state Markov chain approach are shown. The differences of these values across different sets of initial conditions and across two levels of accuracies are also presented.
(XLS)

File S9 Installation guide for the grid-based pipeline of *optPBN* and an execution example on Grid'5000. The PDF documentation provides a list of commands to set-up the grid-based version of the *optPBN* toolbox on a cluster or on a grid-based infrastructure. An example set of commands to reserve resources and to execute an optimisation task on Grid'5000 is also included.
(PDF)

Acknowledgments

— The authors would like to thank Fréderic Fourré for fruitful discussions on qualitative and semi-quantitative modelling.
— Experiments presented in this paper were partly carried out using the Grid'5000 testbed, supported by a scientific interest group hosted by Inria and including CNRS, RENATER and several Universities as well as other organizations (see https://www.grid5000.fr).
— Andrzej Mizera is on leave of absence from the Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland.

Author Contributions

Carried out the development and implementation of the toolbox: PT TS. Revised the optimisation pipelines: AM JP. Set up the grid-based optimisation platform on Grid'5000 and wrote supporting documentations: AT. Carried out the case studies and validated the obtained results: PT AM TS. Drafted the manuscript: PT AM TS. Revised and approved the final manuscript: PT AM JP AT TS.

References

1. Kauffman SA (1969) Metabolic stability and epigenesist in randomly constructed genetic nets. Journal of Theoretical Biology 22: 437–467.
2. Li F, Long T, Lu Y, Ouyang Q, Tang C (2004) The yeast cell-cycle network is robustly designed. PNAS 101(14): 4781–4786.
3. Saez-Rodriguez J, Simeoni L, Lindquist JA, Hemenway R, Bommhardt U, et al. (2007) A Logical Model Provides Insights into T Cell Receptor Signaling. PLoS Computational Biology 3(8): e163.
4. Schlatter R, Schmich K, Vizcarra IA, Scheurich P, Sauter T, et al. (2009) ON/OFF and Beyond – A Boolean Model of Apoptosis. PLoS Computational Biology 5(12): e1000595.
5. Bornholdt S (2008) Boolean network models of cellular regulation: prospects and limitations. Journal of the royal society 5: S85–S94.
6. Trairatphisan P, Mizera A, Pang J, Tantar AA, Schneider J, et al. (2013) Recent development and biomedical applications of probabilistic Boolean networks. Cell communication and signaling 4(6): 1–25.
7. Shmulevich I, Dougherty ER (2010) Probabilistic Boolean Networks, The Modeling and Control of Gene Regulatory Networks. Philadelphia: Society for Industrial and Applied Mathematics. 281 p.
8. Shmulevich I, Dougherty ER, Zhang W (2002) From Boolean to Probabilistic Boolean Networks as Models of Genetic Regulatory Networks. Proceedings of the IEEE 90(11): 1778–1792.
9. Shmulevich I, Dougherty ER, Kim S, Zhang W (2002) Probabilistic Boolean Network: a rule-based uncertainty model for gene regulatory networks. Bioinformatics 18(2): 261–274.

10. Pal R, Datta A, Bittner ML, Dougherty ER (2005) Intervention in context-sensitive probabilistic Boolean networks. Bioinformatics 21(7): 1211–1218.
11. Yu L, Marshall S, Forster T, Ghazal P (2006) Modelling of macrophage gene expression in the interferon pathway. In IEEE International Workshop on Genomic Signal Processing and Statistics 2006 (GENSIPS '06): 28–30 May 2006; Texas, U.S.A. Edited by Edward R. . Dougherty: 45–46.
12. Ma Z, Wang ZJ, McKeown MJ (2008) Probabilistic Boolean Network Analysis of Brain Connectivity in Parkinson's Disease. IEEE Journal of selected topics in signal processing 2(6): 975–985.
13. Flöttmann M, Scharp T, Klipp E (2012) A stochastic model of epigenetic dynamics in somatic cell reprogramming. Frontiers in physiology 3(216): 1–19.
14. Tay JC, Tan P (2006) Finding Intervention Points in the Pathogenesis of Dengue Viral Infection. In Proceedings of the 28th IEEE EMBS Annual International Conference: 30 August–3 September 2006; New York City, U.S.A. Edited by Atam P. . Dhawan: 5315–5321.
15. Hashimoto RF, Stagni H, Higa CHA (2009) Budding yeast cell cycle modeled by context-sensitive probabilistic Boolean network. In IEEE International Workshop on Genomic Signal Processing and Statistics 2009 (GENSIPS '09): 17–21 May 2009; Minneapolis, Minnesota. Edited by Ahmed Tewfik: 1–4.
16. Fertig EJ, Danilova LV, Favorov AV, Ochs MF (2011) Hybrid modeling of cell signaling and transcriptional reprogramming and its application in *C. elegans* development. Frontiers in Genetics 2(77): 1–9.
17. CellNetOptimizer software website. Available: http://www.ebi.ac.uk/saezrodriguez/cno. Accessed 2014 May 22.
18. Saez-Rodriguez J, Alexopoulos LG, Epperlein J, Samaga R, Lauffenburger DA, et al. (2009) Discrete logic modeling as a means to link protein signaling networks with functional analysis of mammalian signal transduction. Molecular Systems Biology 5(331): 1–19.
19. Bender C, Henjes F, Fröhlich H, Wiemann S, Korf U, et al. (2012) Dynamic deterministic effects propagation networks: learning signalling pathways from longitudinal protein array data. Bioinformatics 26: i596–i602.
20. BN/PBN toolbox website. Available: http://code.google.com/p/pbn-matlab-toolbox. Accessed 2014 May 22.
21. Grid'5000 website. Available: https://www.grid5000.fr. Accessed 2014 May 22.
22. Vaz AIF, Vicente LN (2007) A particle swarm pattern search method for bound constrained global optimization. Journal of Global Optimization 39(2): 197–219.
23. Schmidt H, Jirstrand M (2006) Systems Biology Toolbox for MATLAB: A computational platform for research in Systems Biology. Bioinformatics 22(4): 514–515.
24. Systems Biology Toolbox 2 website. Available: http://www.sbtoolbox2.org. Accessed 2014 May 22.
25. Holland JH (1992) Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. Cambridge MA: MIT Press (A Bradford Book). 211 p.
26. Cahon S, Melab N, Talbi EG (2004) ParadisEO: A Framework for the Reusable Design of Parallel and Distributed Metaheuristics. Journal of Heuristics 10: 357–280.
27. Storn R, Price K (1997) Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. Journal of Global Optimization 11: 341–359.
28. Shmulevich I, Gluhovsky I, Hashimoto RF, Dougherty ER, Zhang W (2003) Steady-state analysis of genetic regulatory networks modelled by probabilistic Boolean networks. Comparative and Functional Genomics 4: 601–608.
29. Miranda EN, Parga N (1989) Noise effects in the Kauffman model. Europhysics Letters 10:293–298.
30. Raftery AE, Lewis S (1991) How many iterations in the Gibbs sampler?. In: Bayesian Statistic 4: Proceedings of the Fourth Valencia International Meeting. Berger JO, et al., editors. Oxford: Oxford University Press. pp. 763–773.
31. Klamt S, Saez-Rodriguez J, Lindquist JA, Simeoni L, Gilles ED (2006) A methodology for the structural and functional analysis of signaling and regulatory networks. BMC Bioinformatics 7(56): 1–26.
32. Imamura R, Konaka K, Matsumoto N, Hasegawa M, Fukui M, et al. (2004) Fas ligand induces cell-autonomous NF-kB activation and interleukin-8 production by a mechanism distinct from that of tumor necrosis factor-α. Journal of Biological Chemistry 279: 46415–46423.
33. Cullen SP, Henry CM, Kearney CJ, Logue SE, Feoktistova M, et al. (2013) Fas/C95-induced chemokines can serve as "Find-Me" signals for apoptotic cells. Molecular Cell 49:1034–1048.
34. Terfve C, Cokelaer T, Henriques D, MacNamara A, Goncalves E, et al. (2012) CellNOptR: a flexible toolkit to train protein signaling networks to data using multiple logic formalisms. BMC Systems Biology 6(133): 1–14.
35. Kaderali L, Dazert E, Zeuge U, Frese M, Bartenschlager R (2009) Reconstructing signalling pathways from RNAi data using probabilistic Boolean threshold networks. Bioinformatics 25(17): 2229–2235.

4.2. A Probabilistic Boolean Network Approach for the Analysis of Cancer-specific Signalling: a Case Study of Deregulated PDGF Signalling in GIST

Preface: Different types of mathematical models in systems biology have been applied to study and analyse biological systems where each of them has distinct strong and weak points. For the analysis of signal transduction networks with quantitative steady-state proteomic measurements, applying a qualitative modelling approach, such as Boolean network, usually fails to capture quantitative changes of molecules due to its inherent qualitative nature of state and time. In contrast, applying a quantitative method such as ODE-based models might need a richer set of data to identify kinetic parameters. Also, there might exist a large number of potential model variants which fit equally well to a single set of steady-state data when the majority of model structure is not yet known.

Given the drawbacks of each modelling approach, we see probabilistic Boolean networks (PBNs), which can capture the quantitative changes of molecular states at steady-state and can deal with uncertainty within the network structure, as an alternative compromise. We previously demonstrated in (Trairatphisan et al., 2014) that the PBN approach could be applied to study and analyse an apoptotic signalling network in the physiological context of hepatocytes. In this article, we applied the same approach to model and analyse the deregulated PDGF signalling in the pathological context of GIST. An extensive set of steady-state protein profiles of PDGFRα mutants in combination with signalling inhibitors was generated and integrated with a literature-derived PDGF signalling network applying the tool *optPBN*. The derived context-specific PDGF signalling in GIST points out to the importance of some crosstalk interactions including the ones from PI3K to MAPK pathway and another one from PKC to MEK1,2.

To evaluate our modelling results, we performed an additional set of experiments by inhibiting PI3K and PKC with signalling inhibitors and observing their crosstalk targets. The results revealed that PKC is independent of PDGFRα activation which is in contrast to the assumption applied to the initial model structure. We subsequently revised the model structure and integrated additional data into model training. The refined final model demonstrated that only the crosstalk interactions from PI3K towards MAPKs are essential. The refined model was also evaluated with a validation dataset for predictive power where it delivers very good predictions.

A Probabilistic Boolean Network Approach for the Analysis of Cancer-specific Signalling: a Case Study of Deregulated PDGF Signalling in GIST

Short title: A PBN Study on Deregulated PDGF Signalling in GIST

Panuwat Trairatphisan[1], Monique Wiesinger[1], Christelle Bahlawane[2], Serge Haan[2], and Thomas Sauter[1]*

1 Systems Biology group, Life Sciences Research Unit, University of Luxembourg, Luxembourg, Luxembourg

2 Molecular disease mechanisms group, Life Sciences Research Unit, University of Luxembourg, Luxembourg, Luxembourg

* Corresponding author (thomas.sauter@uni.lu)

Abstract

Signal transduction networks are increasingly studied with mathematical modelling approaches. Modelling the network behaviour with qualitative approaches, like Boolean networks, does not usually capture quantitative effects while applying detailed mechanistic modelling requires an extensive amount of data to infer the respective kinetic parameters. Probabilistic Boolean network (PBN) modelling might therefore be a promising compromise, as it allows capturing quantitative changes of molecular states at steady-state with minimal parameterisation. PBN modelling so far has not been explored for studying signalling networks.

We successfully applied the PBN approach to model and analyse the deregulated Platelet-Derived Growth Factor (PDGF) signalling pathway in Gastrointestinal Stromal Tumour (GIST). We experimentally determined a rich and accurate dataset of steady-state profiles of selected downstream kinases of PDGF-receptor-alpha mutants in combination with inhibitor treatments. Applying the tool optPBN, we fitted a literature-derived candidate network model to the training dataset consisting of single perturbation conditions. Model analysis suggested several important crosstalk interactions. The validity of these predictions was further investigated experimentally pointing to relevant ongoing crosstalk from PI3K to MAPK signalling in tumour cells. The refined model was evaluated with a validation dataset comprising multiple perturbation conditions. The model thereby showed excellent performance allowing to quantitatively predict the combinatorial responses from the individual treatment results in this cancer setting.

We therefore propose the PBN approach as a promising method for analysing signal transduction networks, as it allows for modelling these networks in a simplified manner while still delivering quantitative predictions. The established optPBN pipeline is widely applicable to gain a better understanding of signalling network properties at steady-state in a context-specific fashion.

Keywords: probabilistic Boolean networks, Systems Biology, signal transduction networks, PDGF signalling, Gastrointestinal Stromal Tumour

Author summary

Probabilistic Boolean network (PBN) modelling is an alternative modelling approach which allows capturing quantitative changes of molecular states at steady-state with minimal parameterisation. PBN has been mainly applied to study gene regulatory networks in the context of uncertainty. In this research article, we applied PBN to model and analyse a cancer-specific signalling network with uncertain influences from crosstalk interactions. We demonstrated that PBN models can fit well to steady-state protein data. Subsequent modelling analysis revealed the importance of certain crosstalk interactions where the validity of these predictions was further investigated experimentally. The refined model was evaluated with a validation dataset where it showed very good predictive power. With its simplicity yet effectiveness of the PBN framework, we propose PBN modelling as an alternative approach for analysing signal transduction networks with steady-state measurements. The same analytical pipeline as presented in this article could also be applied for the study of other signalling pathways in order to gain a better understanding on their contextualised properties at steady-state.

Introduction

Signal transduction networks are one of the core functional layers in cells. They conveys intra- and extra-cellular signals towards cellular regulators which modulate the expression of cellular phenotypes corresponding to the types and concentrations of the stimuli [1]. In general, a signal transduction network is a large and highly complex network comprising multiple intracellular signalling pathways such as mitogen-activated protein kinases (MAPK), phosphatidyl-inositide-3-kinases/AKT/mammalian-target-of-Rapamycin (PI3K/AKT/mTOR), and phospholipase C gamma/protein kinase C (PLCγ/PKC) pathways [2]. Also, there exist a number of crosstalk interactions between these signalling pathways which help to fine-tune signals and to preserve the whole networks' integrity upon perturbations [3–5]. In a physiological condition, signals from various stimuli are optimally integrated and transduced to regulate cellular functions and to maintain homeostasis of cellular components [6]. Once the transfer processes of these signals are deregulated, the resulting aberrant signals often lead to abnormalities of cellular functions which were identified as the etiology of many diseases including neurodegenerative diseases, metabolic diseases, as well as cancers [7–9].

Gastrointestinal stromal tumours (GISTs) are the most common primary mesenchymal neoplasia of the gastrointestinal tract. These tumours commonly arise from gain of function mutations of type III receptor tyrosine kinases, i.e. KIT at 78-90% of the cases and platelet-derived growth factor alpha (PDGFRα) at 5-7% of the cases [10]. As constitutively active surface receptors, these mutated proteins are the most upstream components of the cellular signal transduction network. The first-line therapy for GISTs is surgical resection combined with administration of Imatinib mesylate, a tyrosine kinase inhibitor with activities against ABL, BCR-ABL, KIT, and PDGFRα/β. The clinical outcome of the first-line therapy is mostly satisfactory with 35%-49% 9 years survival [11]. However, certain point mutations in GIST, e.g. an Aspartate (D) to Valine (V) mutation at amino acid 842 on the PDGFRA gene, were also shown to be associated with drug resistance [12]. Several hypotheses were proposed to explain the underlying molecular mechanisms of oncogenic PDGFRα-driven GIST formation such as the retention of oncogenic PDGFRα in intracellular compartments [13] or the deregulation of downstream oncogenic signalling pathways, see also [14]. Such hypotheses still require further investigation at the molecular level to understand how signals are transduced and processed mechanistically in this cancer setting.

In recent years, various modelling approaches in Systems Biology were applied to model and analyse the properties of signal transduction networks in both physiological and pathological conditions. This includes Bayesian networks [15], Boolean networks [16], fuzzy logic models [17], ordinary differential equation (ODE)-based models [18], partial differential equation (PDE)-based models [19] and also stochastic models [20], to list only a few examples. Some modelling work also demonstrated the connections between the deregulation of signal transduction networks to the pathophysiology of diseases, e.g. in metabolic disease [21] and in cancers [22]. Even if there exist many modelling approaches for the study of signal transduction networks, each of them has certain advantages and disadvantages. Describing the connections between signalling molecules with conditional probabilities in Bayesian networks framework express their relationships quantitatively but such values could not explain the underlying regulatory mechanisms between molecules. Boolean networks could depict simple regulatory mechanisms of biochemical interactions with logic operators but the results are nonetheless strictly qualitative. More quantitative approaches such as fuzzy logic or ODE-based models provide additional mechanistic details of the networks on a continuous scale. However, they require an extensive amount of prior knowledge to define suitable mathematical formula and they also need an extensive set of experimental data in order to infer kinetic parameters.

Given the pros and cons of different modelling approaches, we envisage probabilistic Boolean network (PBN) modelling which allows capturing quantitative changes of molecular states at steady-state with minimal parameterisation as an alternative framework for modelling and analysis of signal transduction networks. Initially, PBN was introduced by Shmulevich et al. to model gene regulatory networks in the context of uncertainty [23]. With the assignment of probabilities on each interaction, PBN is capable to depict individual regulatory effects in a stochastic manner. The concept of PBN can also be applied to study other biological networks including metabolic networks and physiological networks, for a recent review see [24]. In terms of applications on signal transduction networks, Bayesian networks coupled with a probabilistic Boolean threshold function were used to infer the topology of the signal transduction network from experimental data [25]. Nevertheless, the concept of PBN as originally introduced by Shmulevich et al. has not been applied to model and analyse signal transduction networks so far. We therefore investigated the application of the original PBN concept as introduced in [23] to model and analyse a contextualised signal transduction network in a pathological condition with our tool optPBN [26]. As a case study, we investigated the deregulation of PDGF signalling in GIST. We demonstrate its excellent performance in modelling and

parameter fitting, as well as its predictive capacity, by correctly predicting combinatorial treatment results from individual stimulation measurements.

Materials and Methods

Biological materials

The PDGFRα-mutant proteins were generated based on the pLNCX2-PDGFRα wild-type expression plasmid generously provided by Prof. Andrius Kazlauskas (Boston). This sequence was cloned to pcDNA5/FRT/TO-vector (Invitrogen™) and the constitutively active oncogenic PDGFRα mutant occurring in GIST was generated by introducing the D842V point mutation (PDGFRα-D842V-wild type, "DV-WT"). Based on the PDGFRα-D842V mutant, two PDGFRα-D842V-"knock-out" mutants were constructed by introducing tyrosine (Y) to phenylalanine (F) point mutations, i.e. Y720F (PDGFRα-D842V-Y720F, "DV-dMAPK") and YY731/742FF (PDGFRα-D842V-YY731/742FF, "DV-dPI3K") which have been shown to abrogate the recruitment of signalling molecules such as SHP2 and PI3K in the PDGFRα-wild type receptor, respectively. All point mutations were introduced using the QuikChange kit (Stratagene) following the manufacturer's recommendations.

For analysing the signalling behaviour of the mutant PDGFRα protein, an isogenic Flp-In™ cell line was constructed based on the Hek293 cell line which is naturally devoid of endogenous expression of PDGFRα/β proteins. As such, Hek293 cells were co-transfected with the Flp-In™ target site vector (pFRT/lacZeo, Invitrogen™) and the regulatory vector (encoding Tetracycline repressor protein, pcDNA™6/TR/invitrogen) using the TransIT®-LT1 transfection reagent (Mirus) according to the manufacturer's recommendations. Cell clones were selected by cultivation in presence of 10µg/ml Blasticidin and 100µg/ml Zeocin™ (both InvivoGen). Based on this isogenic parental cell line ("293FR"), stable cell lines were generated by site directed recombination by co-transfecting the transgene expression plasmid (pcDNA5/FRT/TO-based) in combination with the Flp recombinase expression plasmid (pOG44, Invitogen). Stably transfected cells were selected and cultivated in presence of 100µg/ml Hygromycin and 10µg/ml Blasticidin [13].

Cell treatment and Western blot analysis

Experiments were conducted by seeding 250,000 cells/well (12 well plate) in DMEM containing 10% FBS, 2% L-Glutamine, and 25mM HEPES on 12-well plates for 24-30 hours. Then, growth media was exchanged and protein expression was induced by

adding 5ng/ml doxycycline (Sigma) under serum reduced (1% FBS) conditions for 14 hours and for additional 3 hours under serum free (0% FBS) conditions. Pharmacological inhibition was performed by adding either 1μM of the PI3K inhibitor Wortmannin, 10μM of the MEK1,2 inhibitor U0126, 1μM of the pan-PKC inhibitor GF109203X, or 500nM for the PKC-α and PKC-β1 specific inhibitor Gö6976. Wortmannin and Gö6976 were purchased from Sigma-aldrich while GF109203X and U0126 were purchased from Calbiochem.

Cells were lysed on the dish with 300μl 1x Lämmli buffer. Cellular proteins were subjected to SDS-PAGE, transferred to a nylon membrane (Amersham Hybond™-N/GE Healthcare by ThermoFisher Scientific), blocked with 10% BSA and probed with the respective antibodies. Alpha-tubulin was also probed as loading control. Phospho-specific antibodies against ERK1/2 (pThr202/pTyr204), PDGFRα (pTyr849)/β(pTyr857), AKT (pSer473) and PKC substrates (pSer) were purchased from Cell Signaling. Phospho-specific antibodies for STAT5 (pTyr694) and PLCγ1 (pTyr783) were purchased from BD Biosciences. Antibodies against PDGFRα (C-20) were purchased from Santa Cruz Biotechnology®. Each gel was probed simultaneously against alpha-tubulin to allow compensation for loading variations (using either antibody DM1A/Santa Cruz Biotechnology® or PA1-38814/Pierce™). The secondary antibodies were coupled with IRdye, allowing imaging with the LI-COR Odyssey system. The relative intensities of Western blot images were quantified with Image Studio Lite version 4.0 using left-right and top-bottom background subtraction. The experiments were done in 3 biological replicates with 3 technical replicates on each Western blot. The calibrator sample was prepared from an early passage of DV-WT cell line induced by 5ng/ml doxycycline.

Normalisation pipeline and datasets for modelling

Starting from the obtained raw experimental data, a quality control step was applied by discarding 1) the data points overlapping with unspecific stains and 2) the data points with low signals due to blotting issues. Furthermore, data points with a corresponding tubulin signal of less than 20 percent compared to the maximal signal within the same blot were removed from the analysis due to low signal to noise ratio.

The remaining data points were first normalised to tubulin (loading control) and then to a calibrator sample in order to correct for the differences across multiple blots. The normalised mean of technical triplicates from each biological replicate were pooled and re-normalised to the maximal value to generate the final mean and standard deviation for the modelling task.

We divided the initial set of experimental data into two parts. The training dataset comprises 6 experimental conditions including negative control (all signals assumed to be zero), positive control (DV-WT) and 4 experimental conditions with single perturbations by YF point mutations (DV-dMAPK and DV-dPI3K) or signalling inhibitors (DV-WT-Wortmannin and DV-WT-U0126). The validation dataset comprises the 4 remaining experimental conditions with combined perturbations.

Literature-derived PDGF signalling network

We built a mutant PDGF signalling network including the major downstream signalling pathways including MAPK, PI3K/AKT/mTOR, PLCγ/PKC pathways, as well as STAT5 as another prominent signalling target downstream of the oncogenic mutant (but not downstream of the wild-type receptor) [13]. Interactions and regulatory mechanisms were modelled according to well-established knowledge [27,28]. The recruitment sites of upstream signalling molecules to PDGFRα were previously described [29,30]. Among the three major downstream signalling pathways in PDGF signalling, there also exist a number of crosstalk interactions suggested by literature as listed in Table 1. In addition, apart from depicting the effects of inhibitors on their main targets, we also included the information on an off-target effect of Wortmannin on conventional PKC [31] which might be essential to explain the obtained signalling profiles in the context of GIST.

Table 1. References for the investigated crosstalk interactions in the PDGF signalling network.

Crosstalk number	Descriptions	References
1	Ras -> PI3K	(Rodriguez-Viciana et al., 1994)
2	MEK1,2 -| Grb2SOS	(Kolch, Calder, & Gilbert, 2005)
3	PI3K -> Ras	(C.-C. Wang et al., 2009)
4	PI3K -> MEK1,2	(Grammer & Blenis, 1997)
5	PIP3 -> GabSOS	(Markevich et al., 2004)
6	Akt -| GabSOS	(Lynch & Daly, 2002)
7	Akt -| Raf	(Zimmermann & Moelling, 1999)
8	PLCγ -> PI3K	(Z. Wang & Moran, 2002)
9	PKC -> MEK1,2	(Grammer & Blenis, 1997)

The crosstalk interactions among MAPK, PI3K/AKT/mTOR and PLCγ/PKC pathways within PDGF signalling as suggested by literature are listed. The annotation '->' refers to activation and the annotation '-|' refers to inhibition.

PBN description of PDGF signalling network

Based on the topology of the literature-derived PDGF signalling network, we built a corresponding PBN model which comprises 27 nodes (molecules) and 40 edges (interactions). Multiple interactions directing onto one node were usually modelled as separated Boolean rules with the corresponding selection probabilities. We applied the Boolean logic gate "OR" to combine non-exclusive inputs in the same class while inhibitions were represented by the "AND NOT" gate. Additionally, selection probabilities of the main pathways were assigned to be high with the flag "H", while being fixed to low with the flag "L" for crosstalk interactions. This assignment will ensure that the boundary of selection probabilities for the interaction(s) with the flag "H" will always be higher compared to the interaction(s) with the flag "L". This implementation is integrated in the latest version of the optPBN toolbox (version 2.2.3) available on http://sourceforge.net/projects/optpbn. The full set of model descriptions can be found in the Supporting Information: S1 File and the computational scripts of the PBN models, as well as the example of results are included in Supporting Information: S2 File.

Optimisation

We applied the grid-based version of the optPBN toolbox (version 2.2.3) to perform the optimisation [26]. 23 selection probabilities within the PBN models were optimised on the Grid'5000 infrastructure using 160 parallel cores (Intel CPU @2.50Ghz, 16GB ram). 10,000 samples of parameter sets were evaluated by two optimisation algorithms, i.e. differential evolution (DE) and evolutionary algorithm (EA). Three rounds of optimisation were performed where the best fitting cost from DE algorithm was chosen for further analyses.

Results

Incomplete inhibition and potential influence of crosstalk interactions were observed from experimental data

We investigated the signalling profiles of three PDGFRα mutants, i.e. DV-WT, DV-dMAPK and DV-dPI3K in combinatorial treatment with two signalling inhibitors, i.e. Wortmannin and U0126. According to the results from the Western blot investigation (Fig. 1), the cell system and the signalling inhibitors functioned properly. We observed that PDGFRα transgenes were expressed exclusively upon doxycycline induction and that basal phosphorylation of the signalling molecules was very low. Also, phosphorylated PDGFRα (pPDGFRα) signals were found to be comparable between the different mutants (Fig. 1). This observation also applies to phosphorylated STAT5 (pSTAT5) signals. The phosphorylated PLCγ (pPLCγ) signals dropped in the DV-dMAPK mutant, indicating that the SHP-2 recruiting motif Y720 also affects PLCγ activation (see Fig. 1). In parallel, decreased activities of downstream signalling molecules, i.e. phosphorylated ERK1,2 (pERK1,2) and phosphorylated AKT (pAKT) were observed in accordance to the inhibitory effects of YF mutants and the signalling inhibitors. These results indicate that the abrogation of recruitment sites by point mutations and the signalling inhibitors worked effectively on their primary targets. In addition, we found that neither the constructions of our cell system nor the inhibitor treatments used in the study affect the expression level of the investigated signalling components (see [13] and Supporting Information: S1 Figure).

Among the inhibitory effects that we investigated, we observed that both Wortmannin and U0126 reduced the phosphorylation of their downstream signalling targets almost completely. In contrast, the DV mutants with additional YF point mutations delivered only partial inhibition, e.g. the signals of pERK1,2 and pPLCγ in DV-dMAPK mutants are only reduced by half. In parallel, some evidences of crosstalk interaction were observed in the dataset. For instance, inhibiting PI3K with Wortmannin also decreased the pERK1,2 signal in the MAPK pathway while inhibiting MEK1,2 with U0126 additionally increased the pAKT signal in the PI3K/AKT/mTOR pathway (Fig. 1B). It should be noted that the cross-regulation through Wortmannin inhibition is stronger than the one via U0126, with up to 39% decreased pERK1,2 signal versus up to 25% increased pAKT signal compared to the untreated conditions. Also, the cross-regulation by Wortmannin is

significant in 2 out of 3 mutants while the crossed inhibitory effect mediated by U0126 was significant only in the DV-WT condition (Fig. 1).

Initial data integration in the PBN framework suggested important crosstalk interactions from candidate network

We integrated this initial dataset into the literature-derived PBN model of PDGF signalling by applying the tool *optPBN*. The model structure and the splitting of datasets are shown in Fig. 2. Initially, we started with a model topology without any crosstalk interaction to explore if the model that contains only the major oncogenic pathways would already be sufficient to fit the training dataset. The results showed that the initial model fitted well to pPDGFR, pSTAT5 and pAKT data, but still could not capture the decrease of pERK1/2 signals after Wortmannin treatment. Also, we found that a model variant with all-or-none inhibition could not fit well to either the pERK1/2 or the pPLCγ data (see Supporting Information: S2 Figure).

Next, we investigated 9 model variants where the crosstalk interactions proposed in the literature (Table 1) were added one-at-a-time and optimisation was re-performed (Table 2).

Table 2. Fitting costs of the initial model variants with and without crosstalk interactions.

Crosstalk number	Descriptions	Fitting cost
-	No crosstalk	0.179
1	Ras -> PI3K	0.174
2	MEK1,2 -| Grb2SOS	0.190
3	**PI3K -> Ras**	**0.127**
4	**PI3K -> MEK1,2**	**0.131**
5	PIP3 -> GabSOS	0.188
6	Akt -| GabSOS	0.193
7	Akt -| Raf	0.182
8	PLCγ -> PI3K	0.188
9	**PKC -> MEK1,2**	**0.112**

The fitting cost from the initial PBN model without any crosstalk interaction was compared to the ones after adding crosstalk interactions one-at-a-time. The fitting cost is the sum of squared error

between the simulated values and the mean values of measurement data in 6 experimental conditions within the training dataset. The annotation '->' refers to activation and the annotation '-|' refers to inhibition.

Only the positive crosstalk interactions from PI3K to Ras (Number 3), from PI3K to MEK1,2 (Number 4) and from PKC to MEK1,2 (Number 9) could further improve model fitting and allowed to distinguish the signal intensities of the DV-WT and DV-WT-Wortmannin conditions (results not shown).

Iterative experimental investigation and model refinement revealed that PKC activity is independent of PDGFRα activation and the crosstalk interactions from PI3K towards MAPK pathway being the most relevant

To investigate the validity of these modelling results, we set-up an additional experiment by treating the DV-WT mutant with PKC inhibitors (Pan-PKC inhibitor 'GF109231X' or conventional-PKC specific inhibitor 'Gö6976') thereby also probing for phosphorylated PKC (pPKC) substrates to investigate if the PKC inhibitors worked effectively (Fig. 3).

While the PKC inhibitors exhibited a strong effect on their downstream primary target, the results showed that PKC activity is independent from PDGFRα activation. Considering the positive crosstalk interaction from PKC towards MEK1,2, we initially expected to observe decreased activities of pERK1,2 upon the reduction of pPKC substrates signals. However, we witnessed in contrary that the pERK1,2 signals increased after PKC-inhibitors treatments. Regarding the fact that our initial model assumed pPKC signals to be dependent on pPDGFR activation, we concluded that the crosstalk interaction from PKC towards MEK1,2 is not important in our cellular context.

On the counterpart, we still observed that pERK1,2 signals were decreased upon Wortmannin treatment as previously shown in the initial set of experiments (Fig. 1 and Fig. 3). This finding supported the validity of the positive crosstalk interactions from PI3K towards MAPK pathways. In addition, we saw that pAKT and pERK1,2 signals are significantly up-regulated following PKC-inhibitor treatment. On one hand, it could be an off-target effect of the two PKC inhibitors on the negative regulators of both oncogenic pathways. On other hand, the off-target effect might only be applied to PI3K/AKT/mTOR pathway while the changes in pERK1,2 signals were modulated by the positive crosstalk as suggested by the model.

Further model refinement by introducing basal PKC activity and by adding the 3 data points from negative control, positive control and Wortmannin treatment conditions to the training dataset confirmed in the one-at-a-time crosstalk adding experiment (Table 3, see also Table 2 for comparison) that the crosstalk interactions from PI3K to MAPK pathway are the only relevant interactions which improve model fitting. We chose the refined model with the crosstalk from PI3K to MEK1,2 which has the best fitting cost as the final model for further analysis.

Table 3. Fitting costs from the model variants with and without crosstalk interactions integrated after model refinement.

Crosstalk number	Descriptions	Fitting cost
-	No crosstalk	0.181
1	Ras -> PI3K	0.182
2	MEK1,2 -| Grb2SOS	0.180
3	**PI3K -> Ras**	**0.152**
4	**PI3K -> MEK1,2**	**0.139**
5	PIP3 -> GabSOS	0.188
6	Akt -| GabSOS	0.189
7	Akt -| Raf	0.192
8	PLCγ -> PI3K	0.191
9	PKC -> MEK1,2	0.184

The fitting cost from the refined PBN model variants without crosstalk interaction was compared to the ones after adding crosstalk interactions one-at-a-time. The fitting cost is the sum of square error between the simulated values and the mean values of measurement data in 6 experimental conditions within the training dataset. The annotation '->' refers to activation and the annotation '-|' refers to inhibition.

We found that the quality of plotted fitting results generated from the final model is very good when compared against the training dataset (Fig. 4) where similar results can be obtained from the model variant with the crosstalk from PI3K to Ras (results not shown). We checked that a model variant with both crosstalk interactions from PI3K to the MAPK pathway integrated could not further improve the fitting cost (results not shown), possibly due to redundancy. Either one of these two crosstalk interactions is sufficient to explain the signalling profiles of the cancer-specific deregulated PDGF signalling in GIST.

The final PBN model of deregulated PDGF signalling in GIST is highly predictive for combinatorial stimulations and it provides quantitative insights on signal flows.

In order to evaluate the predictive power of the final PDGF model, we simulated the model with combined perturbations conditions integrating YF point mutations (dMAPK and dPI3K) with signalling inhibitors (Wortmannin and U0126) and compared the respective experimental data (validation dataset). The results are shown in Fig. 5.

It was shown that our final PBN model predicts the molecular profiles of the combined perturbations conditions in the validation dataset accurately. 19 out of 20 data points in combined perturbation conditions were precisely predicted within the standard deviation of the experimental data. Also, the prediction on pSTAT5 in the DV-dPI3K-U0126 condition is qualitatively captured correctly. We suspect that this data point could be an experimental outlier due to its distinctly low signal compared to the others on the same panel (see Fig. 1). Overall, this shows the excellent performance of the final PBN model on predicting combinatorial treatments.

The model variant without any crosstalk interaction also returned a relatively good prediction, except in the experimental conditions where the necessary crosstalk interaction takes place (see Supporting Information: S3 Figure). For instance, the discrepancy can be observed in the DV-dPI3K-Wortmannin condition where pERK1,2 signals were modulated by the crosstalk interaction from PI3K/AKT/mTOR pathway.

Besides its predictive power, the final PBN model also summarises the quantitative signal flow within the deregulated PDGF signalling network in GIST by considering the weights of interactions represented by the optimised selection probabilities as shown in Fig. 6.

It demonstrates that, among multiple crosstalk interactions which were proposed in literature to be involved in the PDGF signalling pathway, only one crosstalk is necessary to explain the investigated experimental data in the context of GIST mutants in our cell system. Also, it is shown that the weight of the crosstalk interaction from PI3K to MEK1,2 is not very strong (0.259) compared to the main pathway (0.741). In parallel, it is shown that the weight of basal PKC activity which is independent from the PDGFRα activation completely overrides the activation signals from inositol 1,4,5-trisphosphate (IP3), diacylglycerol (DAG), and calcium ion (Ca) (1.000 versus 0.000). The known inhibitory effects of inhibitor are also reported quantitatively, e.g. the effects of Wortmannin on its primary target PI3K (0.870) is stronger than the one on its off-target molecule PKC (0.113)

while U0126 has a very strong inhibitory effect on MEK1,2 (0.917) [5,37,38]. The complete list of the weights of the interactions from the final PBN model can be found in Supporting Information: S3 File.

Discussion

Advantages of applying the PBN approach for analysing signal transduction networks

In this research article, we demonstrated several advantages of applying the PBN approach for analysing signal transduction networks. First, we experimentally showed that the inhibitory effects of biological interventions do not work in an all-or-none fashion. Such data call for modelling approaches which can capture quantitative changes of signalling molecules where PBN approach serves the purpose well for capturing quantitative changes at steady-state. Second, we showed that a PBN model with minimal parameterisation is sufficient to explain the signalling profiles observed at steady-state. Hence, PBN might be a better candidate for modelling signal transduction networks with steady-state data compared to more quantitative approaches where additional information on prior knowledge and parameters might be needed. Third, we demonstrated that we can study the importance of crosstalk interactions within the PBN framework by analysing the fitting quality after adding an individual interaction. Also, the weights of interaction can be derived quantitatively. Such results lead to the generation of new hypotheses on the importance of crosstalk interactions which can subsequently be tested experimentally. Lastly, we showed that we can predict the signalling profiles of combined perturbation experiments based on a PBN model which was only trained on single perturbation experiments. This might be a promising idea for future studies to assess the effects of combinatorial perturbations or drugs treatments by performing *in silico* prediction on PBN models prior to wet-lab experiments.

Unresolved issues and open questions from the study

There are a couple of unresolved issues and open questions from the study. First, we observed in our initial experimental that pAKT was up-regulated after U0126 treatment (significant in DV-WT mutant). Some evidences were shown that MEK1,2 inhibition can lead to AKT up-regulation but it was only demonstrated in breast and lung cancer models [39,40]. In addition, this observation was not significant in the other two mutants in our initial set of experiment. Second, Wortmannin was shown to potentially inhibit MEK1 (MKK1) [38] and this interaction could replace the need of crosstalk interactions to fit our datasets (results not shown). However, the study is based on in-vitro kinase assays so the threshold for the inhibitory effect might differ substantially comparing to our cell-based

experiments. Furthermore, other studies failed to detect this potential inhibitory effect of Wortmannin on MEK1 [37] and [41]. Lastly, we found that the model variant without any crosstalk interaction also has a relatively good predictive power. One might argue that we do not need to investigate the importance of crosstalk interactions nor to include them in the model. However, we still believe that crosstalk interaction is an essential element that could not be left out from the models of signal transduction networks. To support our argument, we showed in Fig. 5 that at least one essential crosstalk interaction is required in the model to explain the observed signalling profiles. In addition, recent literature supports the role of crosstalk interactions in providing a compensatory mechanism within signal transduction networks which was proven to be an important factor for the discovery and design of therapeutic agents as well [42].

Technical issues, limitations, and outlook on PBN modelling

There are some technical issues and limitations on the current implementation of PBN modelling that should be noted. For instance, the assignment of Boolean rules in our PBN model was still performed manually. Hence, the rules might be inconsistent between different modellers. Also, the assignment of importance on main pathways over crosstalk interactions might be biased, especially once dealing with less well-studied signalling pathways. On the technical side, optPBN generated results based on an approximation of the steady-state distributions. The approximation error might not be negligible for certain examples and the respective parameters (precision and number of optimisation rounds) would have to be tuned accordingly. In addition, even if the grid-based pipeline of optPBN can generate optimisation results in an acceptable timeframe, large computational resources are needed to perform the calculation. This might consider to be a rate-limiting step in the workflow once there is a limited amount of computational resources. This issue also posts a big challenge for analysing larger signalling networks in the future.

To account for these issues and limitations, we foresee several potential improvements to be implemented in the optPBN toolbox. First of all, we plan to build a standardised pipeline for converting network interactions into PBN model structures. Such pipeline should eliminate the variability of PBN modelling among different modellers. In terms of computation, a recent PBN simulation tool on Java called “ASSA-PBN” was developed which allows for a more precise steady-state approximation and for faster calculations [43]. These features of ASSA-PBN could be integrated into the optPBN framework in order to improve the efficiency and accuracy of the results. As for an outlook into future applications, we envisage that the same analytical pipeline as presented here

could also be applied to study the properties of different signalling networks at steady-state. Furthermore, since PBN is also widely used to model gene regulatory networks, it is possible to build multi-scale models which connect the two functional layers within the PBN framework in order to get more comprehensive insights.

Conclusion

We demonstrated a novel pipeline for modelling and analysing signal transduction networks with the PBN approach. By integrating normalised measurement data at steady-state into an *a priori* model topology derived from literature using the tool *optPBN*, we could generate a PBN model which fits well to signalling data at steady-state quantitatively. An additional experiment was performed to investigate the validity of predicted crosstalk interactions. After a new round of experimental and modelling work, we generated a final PBN model of the PDGF signalling network contextualized for GIST with a high predictive power for predicting the combinatorial treatments from the individual stimulation measurements. Such model could be further used to guide and reduce the number of experimental conditions to be investigated in the lab. In addition, the analytical pipeline we presented could also be used to study and analyse other signal transduction networks.

Acknowledgements

- The authors would like to thank Dr. Andrzej Mizera and Dr. Jun Pang for fruitful discussions on qualitative and quantitative modelling.
- Experiments presented in this paper were carried out using the Grid'5000 testbed, supported by a scientific interest group hosted by Inria and including CNRS, RENATER and several Universities as well as other organizations (see https://www.grid5000.fr).

References

1. Bradshaw R., Dennis E. Handbook of Cell signalling. 2nd ed. Academic Press; 2009.

2. Lemmon M a., Schlessinger J. Cell signaling by receptor tyrosine kinases. Cell. 2010;141: 1117–1134. doi:10.1016/j.cell.2010.06.011

3. Wang C-C, Cirit M, Haugh JM. PI3K-dependent cross-talk interactions converge with Ras as quantifiable inputs integrated by Erk. Mol Syst Biol. 2009;5: 246. doi:10.1038/msb.2009.4

4. Zimmermann S, Moelling K. Phosphorylation and regulation of Raf by Akt (protein kinase B). Science. 1999;286: 1741–1744. doi:10.1126/science.286.5445.1741

5. Grammer TC, Blenis J. Evidence for MEK-independent pathways regulating the prolonged activation of the ERK-MAP kinases. Oncogene. 1997;14: 1635–1642. doi:10.1038/sj.onc.1201000

6. Helikar T, Konvalina J, Heidel J, Rogers J a. Emergent decision-making in biological signal transduction networks. Proc Natl Acad Sci U S A. 2008;105: 1913–1918. doi:10.1073/pnas.0705088105

7. Kim EK, Choi E-J. Pathological roles of MAPK signaling pathways in human diseases. Biochim Biophys Acta. Elsevier B.V.; 2010;1802: 396–405. doi:10.1016/j.bbadis.2009.12.009

8. Lawrence MC, Jivan A, Shao C, Duan L, Goad D, Zaganjor E, et al. The roles of MAPKs in disease. Cell Res. 2008;18: 436–442. doi:10.1038/cr.2008.37

9. Vivanco I, Sawyers CL. The phosphatidylinositol 3-Kinase AKT pathway in human cancer. Nat Rev Cancer. 2002;2: 489–501. doi:10.1038/nrc839

10. Tan CB, Zhi W, Shahzad G, Mustacchia P. Gastrointestinal Stromal Tumors: A Review of Case Reports, Diagnosis, Treatment, and Future Directions. ISRN Gastroenterol. 2012;2012: 1–16. doi:10.5402/2012/595968

11. Rammohan A, Sathyanesan J, Rajendran K, Pitchaimuthu A, Perumal S-K, Srinivasan U, et al. A gist of gastrointestinal stromal tumors: A review. World J Gastrointest Oncol. 2013;5: 102–12. doi:10.4251/wjgo.v5.i6.102

12. Markku M, Jerzy L. Gastrointestinal Stromal Tumors Review on Morphology, Molecular Pathology, Prognosis, and Differential Diagnosis. Virchows Arch. 2006;130: 1466–1478. doi:10.1016/B978-1-4160-5544-0.00081-7

13. Bahlawane C, Eulenfeld R, Wiesinger MY, Wang J, Muller A, Girod A, et al. Constitutive activation of oncogenic PDGFRα-mutant proteins occurring in GIST patients induces receptor mislocalisation and alters PDGFRα signalling characteristics. Cell Commun Signal. 2015;13. doi:10.1186/s12964-015-0096-8

14. Heinrich MC, Corless CL, Duensing A, McGreevey L, Chen C-J, Joseph N, et al. PDGFRA activating mutations in gastrointestinal stromal tumors. Science. 2003;299: 708–710. doi:10.1126/science.1079666

15. Hill SM, Lu Y, Molina J, Heiser LM, Spellman PT, Speed TP, et al. Bayesian Inference of Signaling Network Topology in a Cancer Cell Line. Bioinformatics. 2012;28: 2804–2810. doi:10.1093/bioinformatics/bts514

16. Saez-Rodriguez J, Simeoni L, Lindquist J a., Hemenway R, Bommhardt U, Arndt B, et al. A logical model provides insights into T cell receptor signaling. PLoS Comput Biol. 2007;3: 1580–1590. doi:10.1371/journal.pcbi.0030163

17. Aldridge BB, Saez-Rodriguez J, Muhlich JL, Sorger PK, Lauffenburger D a. Fuzzy Logic Analysis of Kinase Pathway Crosstalk in TNF/EGF/Insulin-Induced Signaling. PLoS Comput Biol. 2009;5. doi:10.1371/journal.pcbi.1000340

18. Konrath F, Witt J, Sauter T, Kulms D. Identification of New IkBa Complexes by an Iterative Experimental and Mathematical Modeling Approach. PLoS Comput Biol. 2014;10. doi:10.1371/journal.pcbi.1003528

19. Rangamani P, Iyengar R. Modelling spatio-temporal interactions within the cell. J Biosci. 2007;32: 157–167. doi:10.1007/s12038-007-0014-3

20. Yuan Q, Trairatphisan P, Pang J. Probabilistic Model Checking of the PDGF Signaling Pathway. … Systems Biology XIV. 2012. pp. 151–180. Available: http://www.springerlink.com/index/N01M1GR5H55615V7.pdf

21. Park HS, Cho SB. Evolutionary attribute ordering in Bayesian networks for predicting the metabolic syndrome. Expert Syst Appl. Elsevier Ltd; 2012;39: 4240–4249. doi:10.1016/j.eswa.2011.09.110

22. Bachmann J, Raue a., Schilling M, Becker V, Timmer J, Klingmüller U. Predictive mathematical models of cancer signalling pathways. J Intern Med. 2012;271: 155–165. doi:10.1111/j.1365-2796.2011.02492.x

23. Shmulevich I, Dougherty ER, Kim S, Zhang W. Probabilistic Boolean Networks: a rule-based uncertainty model for gene regulatory networks. Bioinformatics. 2002;18: 261–274. doi:10.1093/bioinformatics/18.2.261

24. Trairatphisan P, Mizera A, Pang J, Tantar AA, Schneider J, Sauter T. Recent development and biomedical applications of probabilistic Boolean networks. Cell Commun Signal. 2013;11: 46. doi:10.1186/1478-811X-11-46

25. Kaderali L, Dazert E, Zeuge U, Frese M, Bartenschlager R. Reconstructing signaling pathways from RNAi data using probabilistic boolean threshold networks. Bioinformatics. 2009;25: 2229–2235. doi:10.1093/bioinformatics/btp375

26. Trairatphisan P, Mizera A, Pang J, Tantar AA, Sauter T. optPBN: An optimisation toolbox for probabilistic Boolean networks. PLoS One. 2014;9. doi:10.1371/journal.pone.0098001

27. Andrae J, Gallini R, Betsholtz C. Role of platelet-derived growth factors in physiology and medicine. 2008; 1276–1312. doi:10.1101/gad.1653708.revealing

28. Heldin CH. Targeting the PDGF signaling pathway in the treatment of non-malignant diseases. J Neuroimmune Pharmacol. 2014;9: 69–79. doi:10.1007/s11481-013-9484-2

29. Bazenet CE, Gelderloos J a, Kazlauskas a. Phosphorylation of tyrosine 720 in the platelet-derived growth factor alpha receptor is required for binding of Grb2 and SHP-2 but not for activation of Ras or cell proliferation. Mol Cell Biol. 1996;16: 6926–6936. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=231696&tool=pmcentrez&rendertype=abstract

30. Yu JC, Gutkind JS, Mahadevan D, Li W, Meyers K a, Pierce JH, et al. Biological function of PDGF-induced PI-3 kinase activity: its role in alpha PDGF receptor-mediated mitogenic signaling. J Cell Biol. 1994;127: 479–87. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2120211&tool=pmcentrez&rendertype=abstract

31. Ishizuka T, Nagashima T, Yamamoto M, Kajita K, Yamada K, Wada H, et al. Effects of wortmannin on glucose uptake and protein kinase C activity in rat adipocytes. Diabetes Res Clin Pract. 1995;29: 143–152.

32. Rodriguez-Viciana P, Warne PH, Dhand R, Vanhaesebroeck B, Gout I, Fry MJ, et al. Phosphatidylinositol-3-OH kinase as a direct target of Ras. Nature. Nature Publishing Group; 1994;370: 527–532. doi:10.1038/370527a0

33. Kolch W, Calder M, Gilbert D. When kinases meet mathematics: The systems biology of MAPK signalling. FEBS Lett. 2005;579: 1891–1895. doi:10.1016/j.febslet.2005.02.002

34. Markevich N, Moehren G, Demin O, Kiyatkin A, Hoek J, Kholodenko B. Signal processing at the Ras circuit: what shapes Ras activation patterns? Syst Biol (Stevenage). 2004;1: 104–113. doi:10.1049/sb

35. Lynch DK, Daly RJ. PKB-mediated negative feedback tightly regulates mitogenic signalling via Gab2. EMBO J. 2002;21: 72–82. doi:10.1093/emboj/21.1.72

36. Wang Z, Moran MF. Phospholipase C-gamma1: a phospholipase and guanine nucleotide exchange factor. Mol Interv. 2002;2: 352–355,338. doi:10.1124/mi.2.6.352

37. Davies SP, Reddy H, Caivano M, Cohen P. Specificity and mechanism of action of some commonly used protein kinase inhibitors. Biochem J. Portland Press Ltd; 2000;351: 95–105. doi:10.1517/13543776.11.3.405

38. Bain J, Plater L, Elliott M, Shpiro N, Hastie CJ, McLauchlan H, et al. The selectivity of protein kinase inhibitors: a further update. Biochem J. 2007;408: 297–315. doi:10.1042/BJ20070797

39. Turke AB, Song Y, Costa C, Cook R, Arteaga CL, Asara JM, et al. MEK inhibition leads to PI3K/AKT activation by relieving a negative feedback on ERBB receptors. Cancer Res. 2012;72: 3228–3237. doi:10.1158/0008-5472.CAN-11-3747

40. Hoeflich KP, O'Brien C, Boyd Z, Cavet G, Guerrero S, Jung K, et al. In vivo antitumor activity of MEK and phosphatidylinositol 3-kinase inhibitors in basal-like breast cancer models. Clin Cancer Res. 2009;15: 4649–4664. doi:10.1158/1078-0432.CCR-09-0317

41. Anastassiadis T, Deacon SW, Devarajan K, Ma H, Peterson JR. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. Nat Biotechnol. Nature Publishing Group; 2011;29: 1039–1045. doi:10.1038/nbt.2017

42. Mendoza MC, Er EE, Blenis J. The Ras-ERK and PI3K-mTOR pathways: Cross-talk and compensation. Trends Biochem Sci. Elsevier Ltd; 2011;36: 320–328. doi:10.1016/j.tibs.2011.03.006

43. Mizera A, Pang J, Yuan Q. Reviving the Two-state Markov Chain Approach (Technical Report). 2015. Available: arXiv:1501.01779. Accessed 17 April 2015.

Supporting information

S1 Figure | Expression level of signalling molecules in PDGFRα mutants.
A panel of PDGFRα-D842V mutants with YF point mutations in a separated set of experiment is shown. The experiments were performed following the same protocol as described in Materials and Methods. We observed that different constructions of mutants, including the Y720F and YY731/742FF mutants in our study, did not affect the expression level of the signalling molecules that we investigated.

S2 Figure | Compared fitting quality of the initial PBN variants with all-or-none inhibition versus with partial inhibition.
Model simulations from two initial PBN model variants, one with all-or-none inhibition and another with partial inhibition, were compared against the training dataset in 6 experimental conditions. The model variant with all-or-none inhibition could not be fitted to the data points of pPLCγ and pERK1,2 in DV-dMAPK condition while the model variant with partial inhibition fitted well to these data points. Nevertheless, the latter model still could not distinguish the differences in pERK1,2 signals comparing between DV-WT and DV-WT-Wortmannin conditions.

S3 Figure | Predictive power of the final PBN model versus the model variant without crosstalk.
Model simulations from the final model and the refined model variant without any crosstalk interaction were compared against the validation dataset in 4 experimental conditions. Both model variants have very good predictive power on most of the experimental

conditions. Nevertheless, the refined model variant without any crosstalk interaction could not accurately predict the change of pERK1,2 signal in DV-dPI3K-Wort condition which is modulated via crosstalk signalling.

S1 File | Model description of the PDGF signalling network in PBN format.
This document provides a detailed description of the PBN models of PDGF signalling presented in our study. The model descriptions are divided into 3 sections: 1) The core model structure with major intracellular signalling pathways, 2) The integration of crosstalk interactions proposed in literature, and 3) The final model structure after integrating data from the additional experiment. Small examples within the computational scripts are also included in each section to illustrate how each type of network interaction can be coded in the PBN format.

S2 File | Computational scripts of the PBN models and the result files from the modelling studies.
This compressed zip file comprises the following elements from modelling study: 1) PBN model descriptions in the form of computational scripts, 2) Examples of result file from grid-based computation, and 3) Saved model structures for grid computation. Instructions on how to further analyse the result files and how to re-perform the optimisation are included.

S3 File | Weights of interaction of the final PBN model.
This spreadsheet provides a list of Boolean interactions in the final PBN model together with the mean and standard deviation of the optimised selection probabilities, i.e. weights of interaction, generated from the best 500 parameter sets of the optimisation. Very low standard deviations of the optimised weights (all < 0.025) indicate the high consistency of the optimisation result.

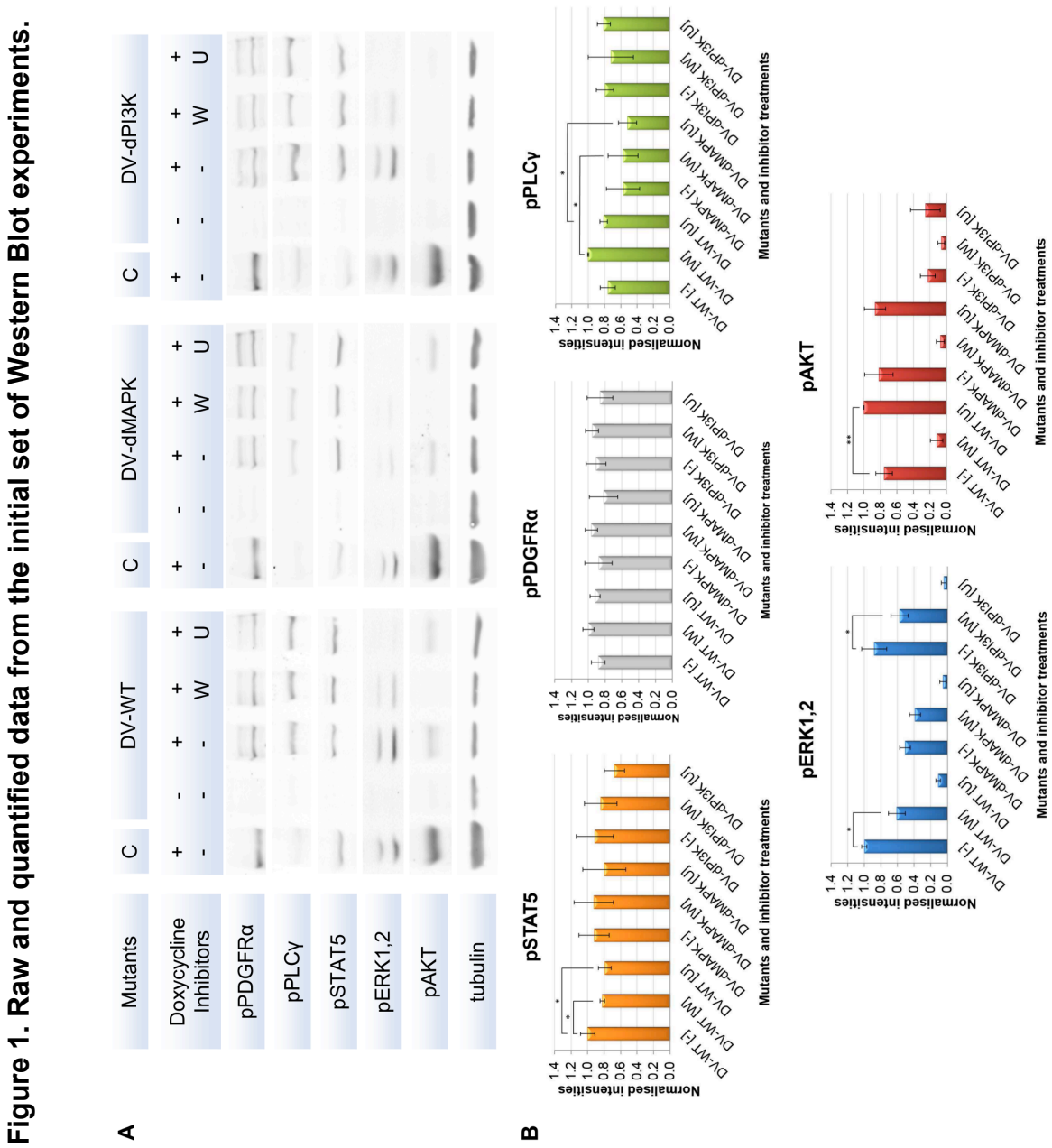


Figure 1. Raw and quantified data from the initial set of Western Blot experiments.

[A] An example set of raw data from Western blot experiments is shown in Fig. 1A. Three PDGFRα mutants containing the D842V point mutation were investigated. One mutant contains no additional YF point mutation, i.e. wild type (DV-WT) while the other two contain either Y720F (DV-dMAPK) or YY731/742FF (DV-dPI3K) point mutations which abrogate the recruitment sites of upstream signalling molecule for MAPK and PI3K/AKT/mTOR pathways, respectively. These experimental conditions were combined with the treatment with two signalling inhibitors, i.e. Wortmannin (W) [1 μM], which mainly inhibits PI3K, and U0126 (U) [10 μM], which mainly inhibits MEK1,2. Doxycycline is used to induce the transcription of the constitutively phosphorylated PDGFRα mutants. The experiment was performed in 3 biological replicates with 3 technical replicates. The signals from the calibrator sample (C) derived from the induced DV-WT cell line in an early passage were applied to calibrate relative intensities between Western blots. **[B]** The summarised quantified Western blot data are shown in Fig. 1B. The signals were normalised against tubulin and subsequently calibrated by the calibrator signals. The normalised mean of technical triplicates from each biological replicate were pooled and re-normalised to the maximal value to generate the final mean and standard deviation values for modelling task. The significances between two data points were assessed with the Student's T-test; p-value < 0.05 (*) and p-value < 0.01 (**).

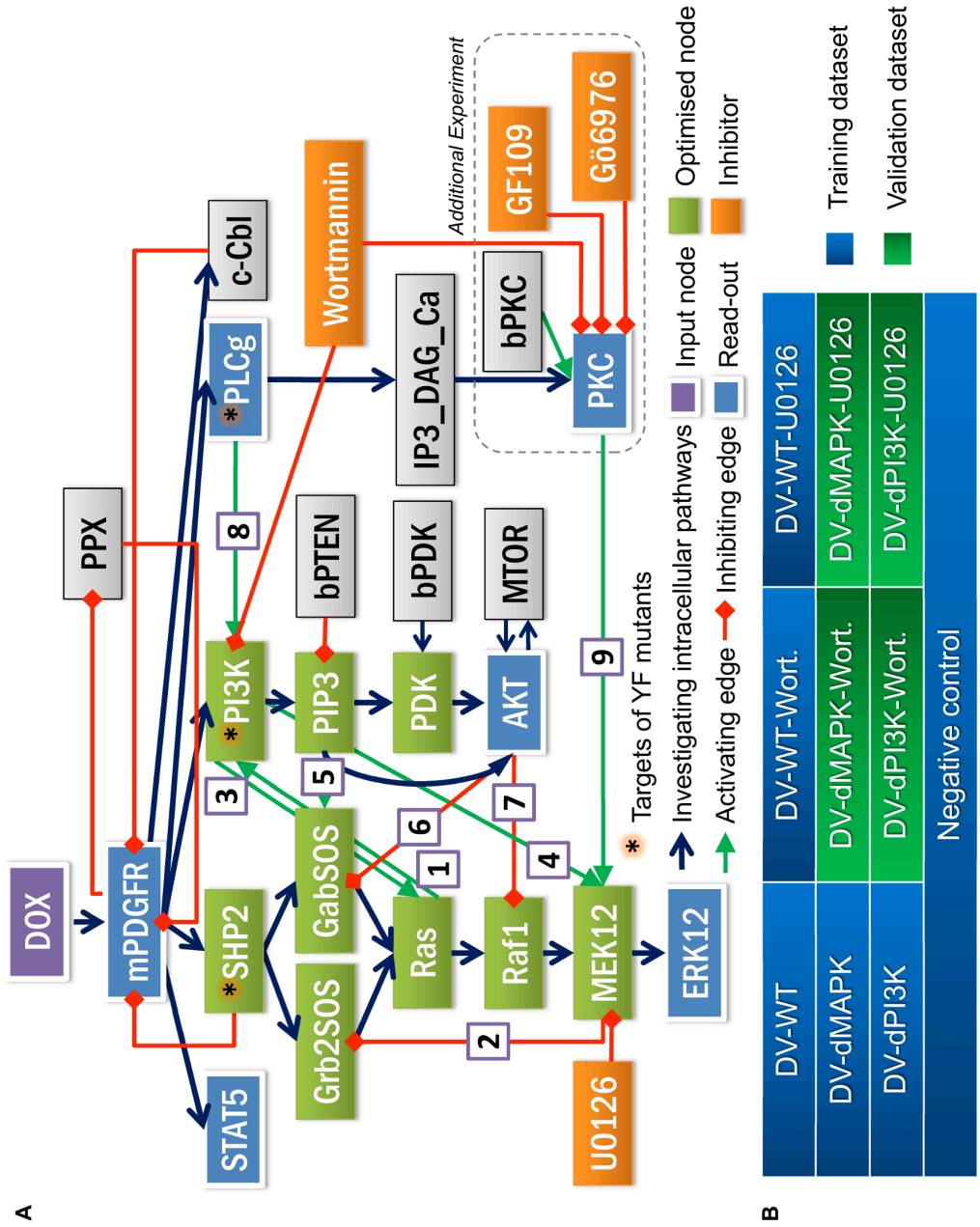


Figure 2: Model structure and datasets for modelling.

[A] The model structure of the literature-derived PDGF signalling is shown in Fig. 2A. The constitutive activity of mutated PDGFRα (mPDGFR) is inducible by doxycycline (DOX). Downstream of mPDGFR includes 3 canonical intracellular signalling pathways: MAPK, PI3K/AKT/mTOR and PLCγ/PKC pathways as well as regulatory mechanisms on PDGFRα. In addition, we included, STAT5, which was shown to be rather activated by oncogenic PDGFRα mutants (Bahlawane et al., 2015). SHP2, PI3K and PLCγ, marked with asterisks, are the targets of YF mutants. The basal activities of PTEN and PDK were represented by bPTEN and bPDK, respectively. Four signalling inhibitors, i.e. Wortmannin, U0126, GF109231X (GF109), and Gö6976, and their targets are also depicted. The phosphorylation state of five downstream signalling molecules, i.e. PDGFRα, STAT5, PLCγ, ERK1,2 and AKT were used as read-outs. There were 9 crosstalk interactions proposed in literature (numbered 1-9) that might be involved in PDGF signalling. Note that the phosphorylation of PKC substrates (PKC) was also included as a read-out in the additional experiment (Fig. 3) and the read-out nodes with multiple inputs, i.e. PDGFRα, PLCγ, PKC and AKT are also optimised. **[B]** The quantified Western blot data from 9 experimentally-investigated conditions together with a negative control condition were applied for modelling task. Experimental data were split into a training set (single perturbation plus positive and negative controls) and a validation dataset (combinatorial perturbations).

This experiment was performed by Dr. Christelle Bahlawane (Figure 3A), the quantification of Western blot images was performed by P. Trairatphisan (Figure 3B).

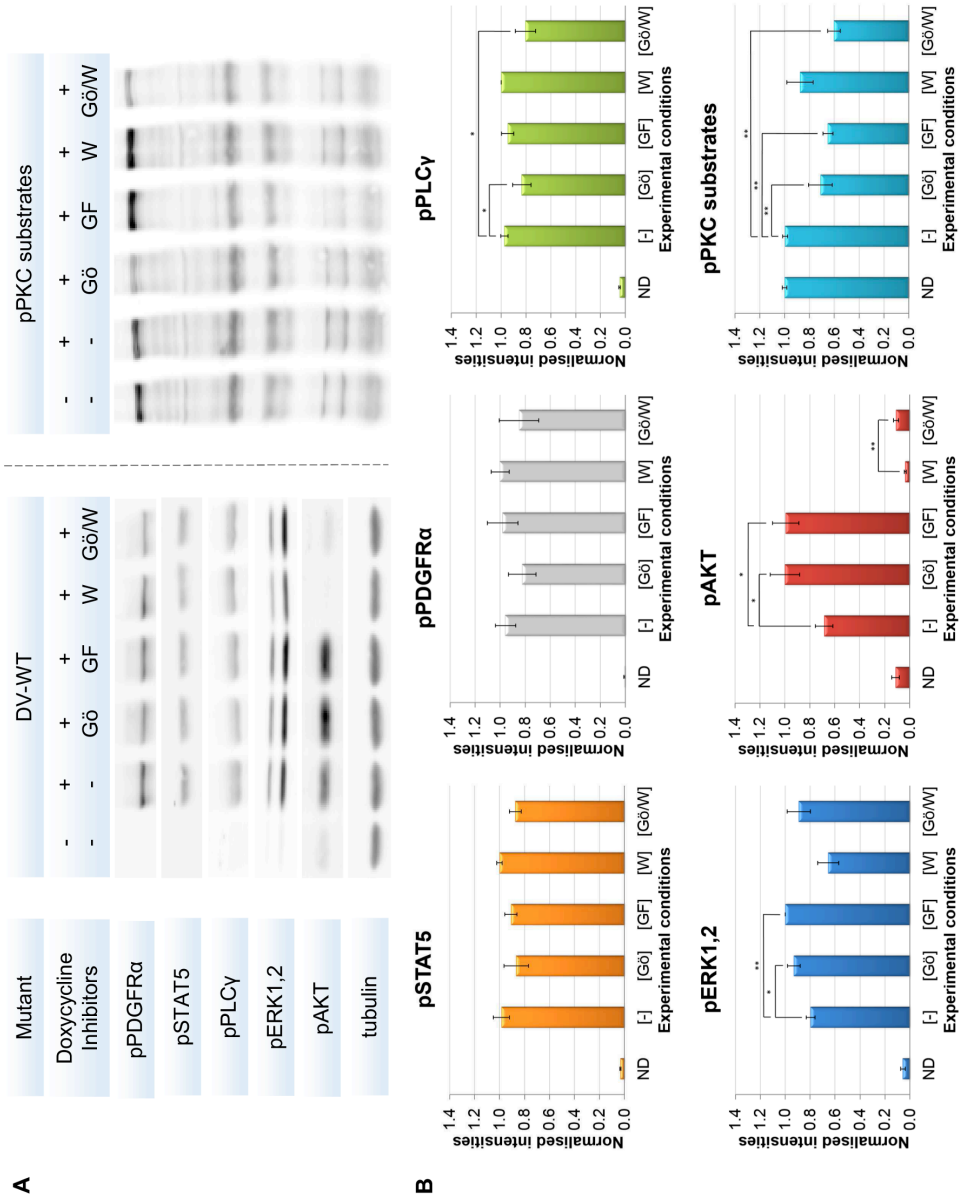


Figure 3: Raw and quantified data of the additional set of Western Blot experiments

[A] An example set of raw data from the additional Western blot experiment to investigate the validity of modelling results are shown in Fig. 3A. The PDGFRα mutant with D842V point mutation (DV-WT) was investigated. Two additional signalling inhibitors, i.e. PKC-α and PKC-β1 specific inhibitor Gö6976 and pan-PKC inhibitor GF109203X together with the detection of phosphorylated PKC (pPKC) substrates were included. Six experimental conditions were assigned, i.e. negative control without doxycycline induction (ND), positive control with doxycycline induction but without inhibitor treatment (-), doxycycline induction with the treatment of Gö6976 (Gö) [500nM], of GF109203X (GF) [1μM], of Wortmannin (W) [1μM] and of combined Gö6976 and Wortmannin (Gö/W) [500nM/1μM]. The experiment was performed in 3 biological replicates with 3 technical replicates. **[B]** The summarised quantified data from the additional Western blot experiment are shown in Fig. 3B. Data were normalised against tubulin, then the mean values of technical triplicates from each biological replicates were pooled and re-normalised to maximal values. The significances between two data points were assessed with the Student's T-test; p-value < 0.05 (*) and p-value < 0.01 (**).

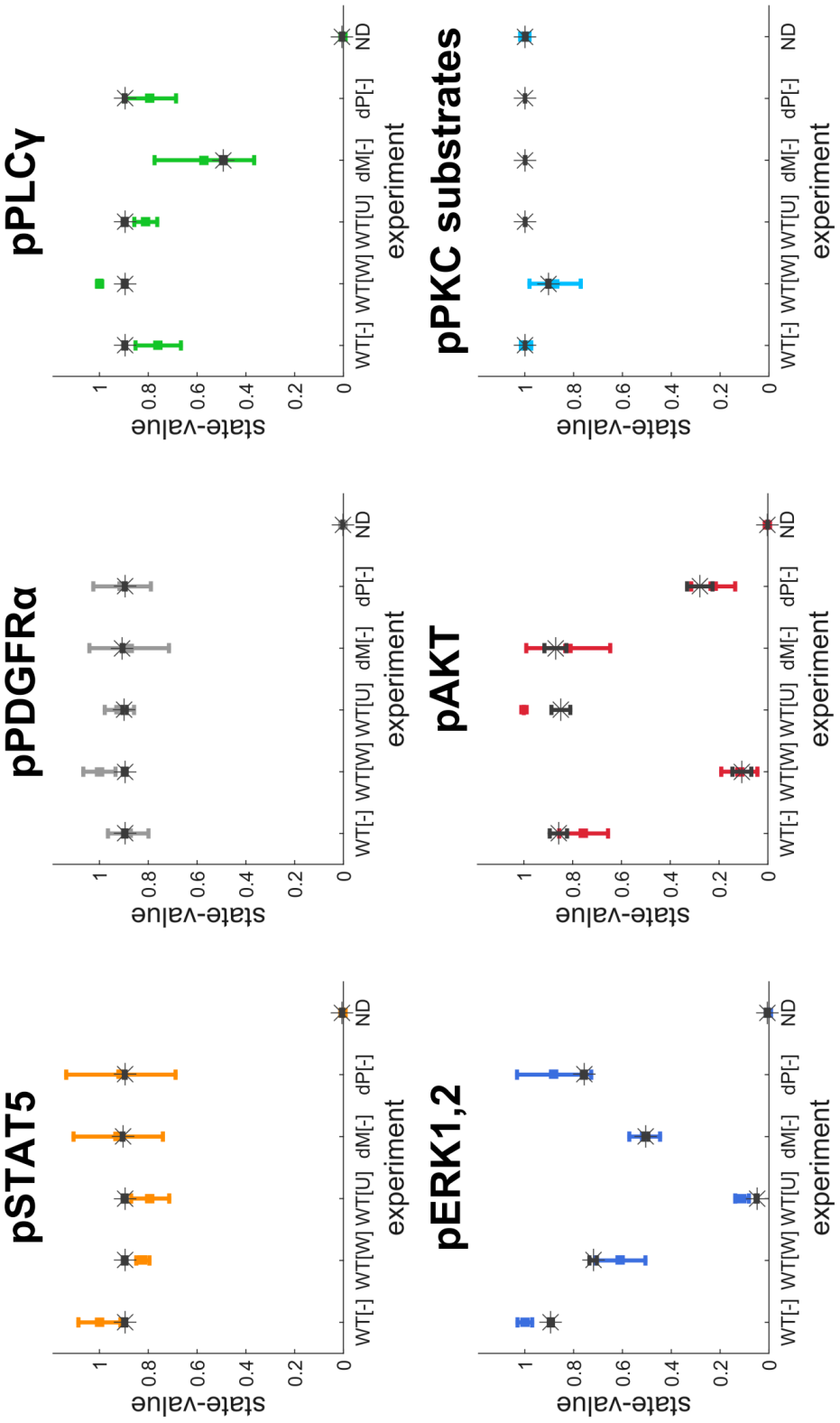


Figure 4: Plotted fitting results of the final PBN model.

The mean and standard deviation (SD) of ten simulated values generated from the final PBN model (black stars [mean] and error bars [SD] on top) were compared against the experimental data from the training dataset (multi-coloured squares [mean] and error bars [SD] on bottom). Six experimental conditions as labelled on the x-axis are in the following order: DV-WT (WT[-]), DV-WT-Wortmannin (WT[W]), DV-WT-U0126 (WT[U]), DV-dMAPK (dM[-]), DV-dPI3K (dP[-]), and negative control (no doxycycline induction, ND).

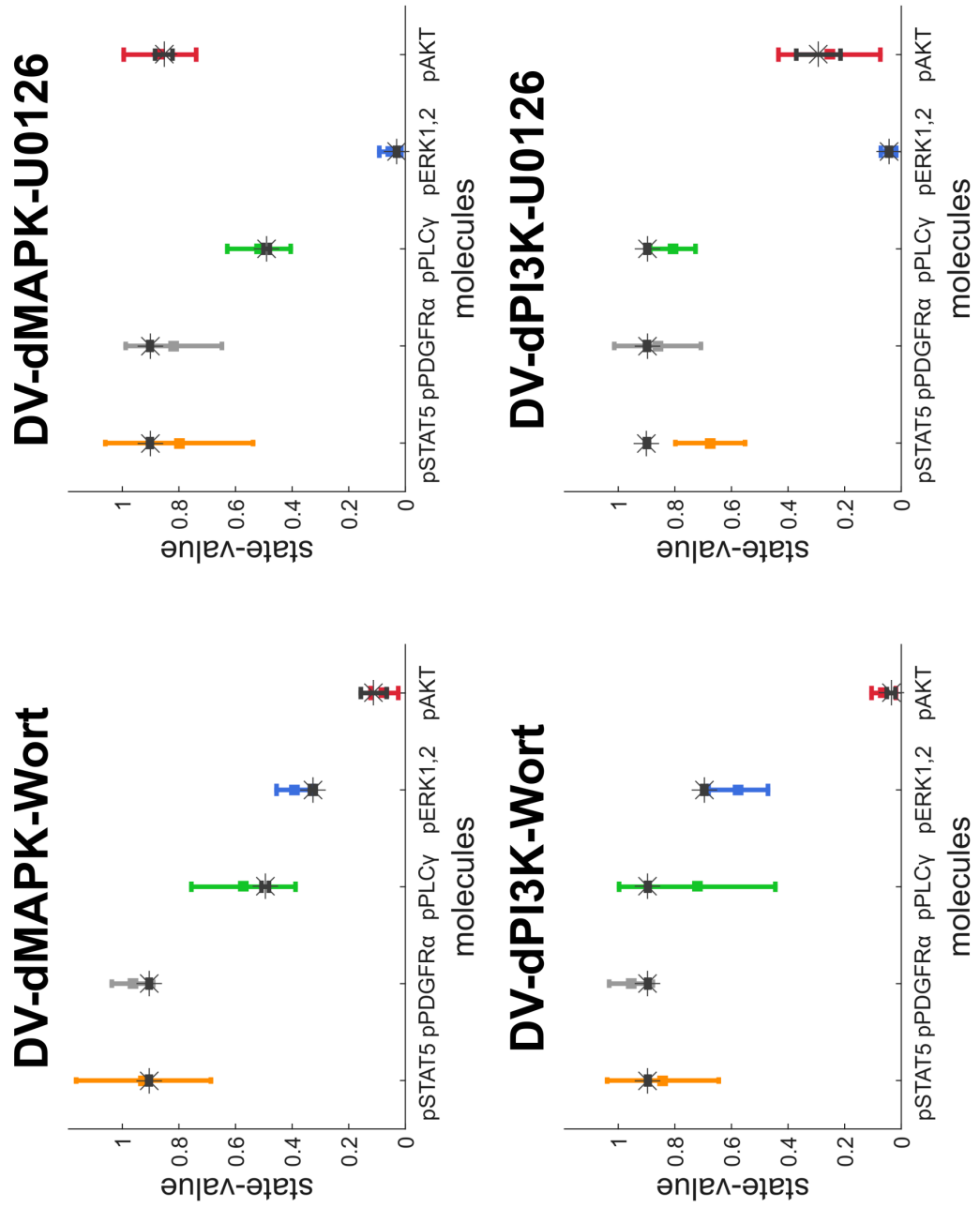


Figure 5: Predictions of the final PBN model compared to the validation dataset.

The mean and standard deviation of ten simulated values from the final PBN model (black stars [means] and error bars [SD] on top) were compared against the experimental data from the validation dataset (multi-coloured squares [mean] and error bars [SD] on bottom). Five molecules as labelled on the x-axis in four experimental conditions are in the following order: pSTAT5, pPDGFRα, pPLCγ, pERK1,2 and pAKT.

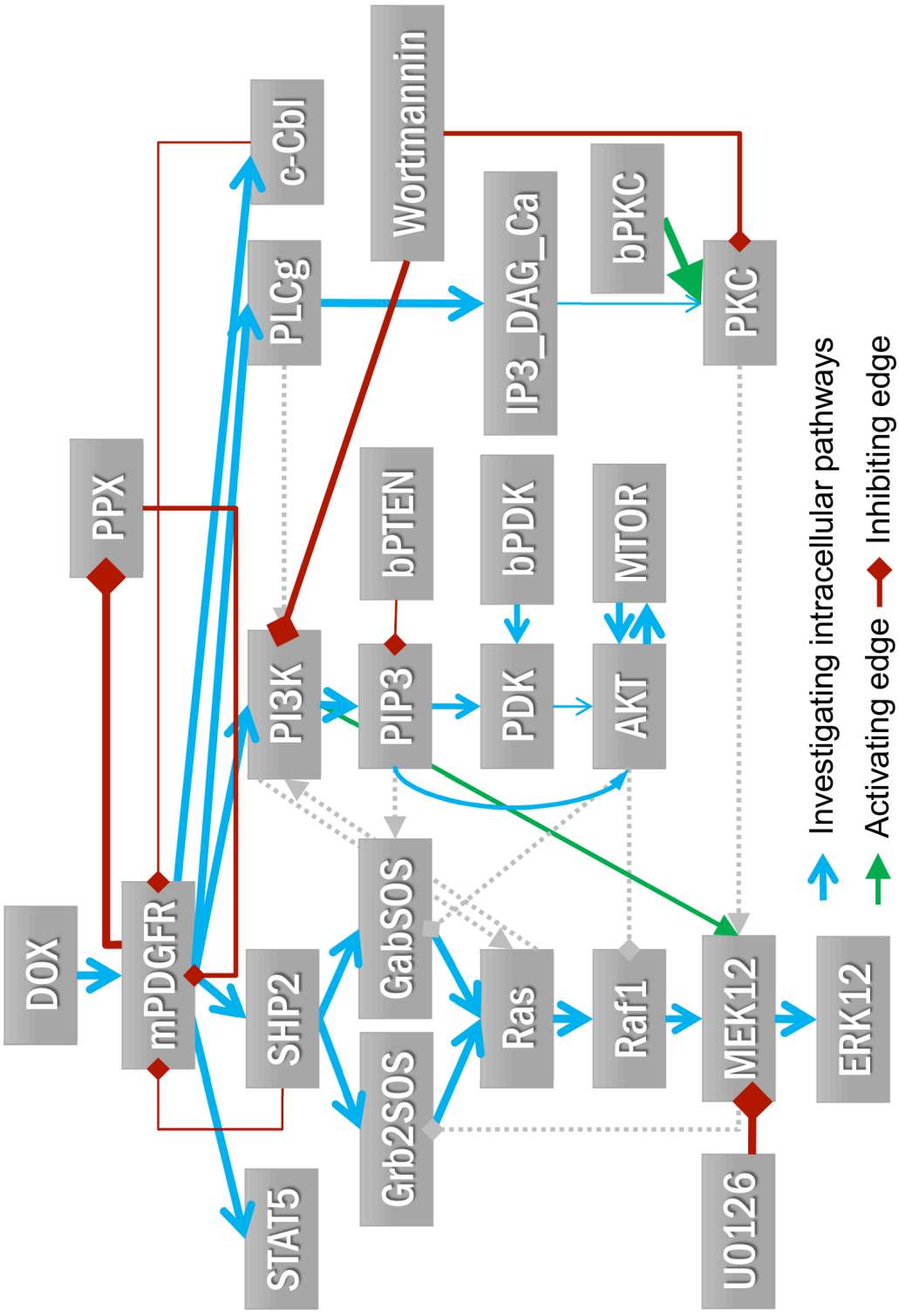


Figure 6: Signal flow in the final PBN model.

A quantitative overview of the signal flow in the final PBN is shown. The weights of interaction as represented by the optimised selection probabilities in PBN model were classified into 5 categories: very low (< 0.1), low (0.1 to 0.3), medium (0.3 to 0.7), high (0.7 to 0.9), and very high (> 0.9) where the sizes of arrow were adjusted accordingly. The crosstalk interactions which are not necessary in the context of GIST were replaced with dotted grey arrows.

4.3. A comparative study of L-plastin signalling in breast cancer cell lines with a probabilistic Boolean approach

Manuscript title: L-plastin Ser5 phosphorylation in breast cancer cells and in vitro is mediated by RSKs, the downstream effectors of the MAPK-ERK pathway

Preface: Cell invasion and metastasis is one of the ultimate hallmarks that cancer cells need to acquire to be able to spread their colonies into adjacent and distant tissues (Hanahan & Weinberg, 2000, 2011). This cellular process is principally governed by the changing expressions of actin-binding proteins such as cofilin, filamin, fastin and L-plastin and these proteins were found to be over-expressed in several types of cancer.

To uncover a novel therapeutic target which could impede the metastatic process, an investigation on the controlling mechanisms of actin-binding proteins at the molecular level is required. In this study, we comparatively investigated the L-plastin signalling in four breast cancer cell lines with different invasive capacities using a combined experimental and modelling approach. A large panel of experimental conditions combining the activators and inhibitors of the L-plastin signalling were initially investigated where several hypotheses on network interaction were formulated. Then, these hypotheses were incorporated into a literature-derived L-plastin signalling model via the MAPK-ERK pathway as additional edges. The integrated model was subsequently optimised by an improved version of the *optPBN* toolbox to investigate the relevance of the hypothesised interactions observed from the data and to compare the differences of signalling paths between non-invasive and invasive cell lines.

According to the optimisation results, only two hypothesised interactions, i.e. PKC activating PKA and ribosomal protein S6 kinase (RSK) activating L-plastin were shown to be relevant in all four cancer cell lines. The validity of these interactions was also confirmed by 100 rounds of bootstrapping. This finding points out to RSK as the key regulator for controlling the phosphorylation of L-plastin, and possibly, for controlling invasion and metastasis. In addition, comparative analyses of the identified selection probabilities revealed that the activities of PKA in invasive cancer cell lines are depended on 8-bromo-cAMP stronger than in the non-invasive ones while the crosstalk between PKA towards PKC might only exist in invasive cell lines. These findings allow researchers to identify more targets to further design novel therapeutic agents.

Remark: Only Figure 5 and result on modelling part were contributed by P. Trairatphisan.

L-plastin Ser5 phosphorylation in breast cancer cells and *in vitro* is mediated by RSKs, the downstream effectors of the MAPK-ERK pathway

Maiti Lommel[1], Panuwat Trairatphisan[2], Karoline Gäbler[1], Christina Laurini[1], Arnaud Muller[3], Tony Kaoma[3], Laurent Vallar[3], Thomas Sauter[2], Elisabeth Schaffner-Reckinger[1*].

[1]Laboratory of Cytoskeleton and Cell Plasticity, Life Sciences Research Unit, University of Luxembourg, Luxembourg [1]Systems Biology Group, Life Sciences Research Unit, University of Luxembourg, Luxembourg; [3]Genomics Research Unit, Luxembourg Institute of Health, Luxembourg

*Corresponding author: elisabeth.schaffner@uni.lu, Tel: (+352) 46 66 44 6581, Fax: (+352) 46 66 44 6442

Abstract

Deregulated cell migration and invasion, which are hallmarks of metastatic cancer cells, are correlated to structural and functional alterations of the actin cytoskeleton which is regulated by numerous actin-binding proteins. Among these, the actin-bundling protein L-plastin, initially detected in leukocytes, is ectopically expressed in several solid tumours and is often considered as a metastatic marker. Phosphorylation of L-plastin on residue serine 5 (Ser5) has been shown to activate L-plastin and to be crucial for invasion and metastasis formation. Here, we investigate the signalling pathway leading to L-plastin phosphorylation using four breast cancer cell lines. Whole genome microarray analysis comparing cell lines with different invasive capacities and corresponding variations in L-plastin Ser5 phosphorylation level revealed that genes of the ERK/MAPK pathway are differentially expressed. In line, in vitro kinase assays showed that ERK/MAPK pathway downstream kinases RSK1 and RSK2 are able to directly phosphorylate L-plastin on Ser5. In parallel to a knockdown approach, activation and inhibition studies of signalling molecules of the ERK/MAPK pathway, followed by computational modelling analysis, confirmed that RSK is an important activator of L-plastin in all four studied cell lines. Finally and rounding up our study, RSK knockdown significantly impaired migratory and invasive capacities of a selected invasive cell line, thus consolidating RSK as a promising therapeutic target in certain invasive carcinomas. Altogether, our data provide substantial evidence that the ERK/MAPK pathway is involved in L-plastin Ser5 phosphorylation in breast cancer cells with its downstream kinases RSK1 and RSK2 being able to directly phosphorylate L-plastin on Ser5.

Keywords

L-plastin

Ser5 phosphorylation

Mitogen-activated protein kinase

Ribosomal S6 kinase

Epidermal growth factor receptor

Invasion

Breast cancer

Abbreviations

MAPK	Mitogen-activated protein kinase
RSK1	Ribosomal protein S6 kinase alpha-1
RSK2	Ribosomal protein S6 kinase alpha-3
MSK1	Ribosomal protein S6 kinase alpha-5
MEK	Dual specificity mitogen-activated protein kinase kinase (MAPKK)
ERK	Extracellular signal regulated kinase (MAPK3)
RAF	RAF proto-oncogene serine/threonine-protein kinase
EGFR	Epidermal growth factor receptor
EGF	Epidermal growth factor
PMA	Phorbol 12-myristate 13-acetate
PKC	Protein kinase C
PKA	Protein kinase A
DEG	Differentially expressed gene
SDE	Significantly differentially expressed
IPA	Ingenuity Pathway Analysis
FC	Fold change
FDR	False discovery rate
TBS	Tris-buffered saline
MT	Mutant
WT	Wild type
SEM	Standard error of the mean

Introduction

Tumor cell migration and invasion are largely dependent on actin cytoskeleton changes [41], which are under the control of a large panel of actin-binding proteins such as cofilin, α-actinin, filamin, fascin or the plastins [41, 44]. In this context, our previous studies have shown that the actin-bundling protein L-plastin regulates the turnover of the actin filaments in vitro and in vivo in addition to its crosslinking activities [1]. Although L-plastin is typically expressed in haematopoietic cells where it plays a role in the immune response (reviewed in [35]), it is also frequently ectopically expressed in carcinoma cells (reviewed in [45]). Here L-plastin localises to actin-rich structures such as focal adhesions, filopodia and membrane ruffles which are involved in adhesion, signalling or locomotion [4, 21].

L-plastin activity has been shown to be increased following phosphorylation on residue serine 5 (Ser5) in vitro and in vivo. In leukocytes, phosphorylation of L-plastin is part of the immune response [23, 33]. Moreover, F-actin-binding and -bundling activities of L-plastin are increased upon Ser5 phosphorylation and this phosphorylation is required for its efficient targeting to focal adhesions [21]. Most interestingly, recent findings have demonstrated that L-plastin Ser5 phosphorylation is crucial for in vitro invasion [21, 28] and in in vivo metastasis formation [42]. Distinct protein kinases have been reported to play a role in L-plastin Ser5 phosphorylation depending on the cell type and environment. Candidate kinases include protein kinase A (PKA) [21, 33, 55], protein kinase C (PKC) [1, 16, 22, 24, 32, 37, 38] and phosphoinositide 3-kinase (PI3K) [24, 37]. Nevertheless, the detailed signalling pathway(s) leading to L-plastin Ser5 phosphorylation remain(s) to be resolved.

Besides the deregulation of the actin cytoskeleton, also the deregulation of signalling pathways is an important feature of many cancers. Breast cancers as well as other cancers frequently display an upregulation of members of the epidermal growth factor family [9]. The epidermal growth factor receptor (EGFR) (also called HER1) and its relatives HER2, HER3 and HER4 are known as oncogenic drivers in various cancers, including lung cancer [34], breast cancer [5] and glioblastoma [30, 31, 54]. Particularly EGFR and HER2 are mutated to constitutively active forms in a large number of epithelial tumors. Signalling pathways downstream of these receptors which are frequently deregulated in cancer are the Ras/Raf/MEK/ERK (MAPK) and the PTEN/PI3K/AKT/mTOR pathways (for review see [20] and [14]).

In this study, we have investigated the signalling pathways leading to L-plastin Ser5 phosphorylation in cancer cells and we have provided evidence that the ERK/MAPK pathway is crucial for L-plastin Ser5 phosphorylation. In vitro kinase assays have revealed that phosphorylation of L-plastin residue Ser5 can be directly mediated by the MAPK-activated protein kinases RSK1 and RSK2 which may provide a useful therapeutic target in the future.

Materials and Methods

Cell Culture

SK-BR-3 and BT-20 cells were grown in Mc Coy's 5A and EMEM (Eagle's minimal essential medium) media respectively and MCF7 and MDA-MB-435S cells were grown in RPMI (Roswell Park Memorial Institute) medium. All media were supplemented with 10% fetal bovine serum, 2 mM of L-glutamine and 100 units/ml of penicillin and streptomycin. Cells were grown at 37°C under 5% CO2 atmosphere. All cells were bought from or authentified by ATCC (American Type Culture Collection, Manassas, VA, US).

Antibodies and Reagents

Mouse monoclonal IgG1 against L-plastin (LPL4A.1, MA5-11921) was purchased from Thermo Scientific (Erembodegem, Belgium). Rabbit polyclonal anti-Ser5-P antibody recognising specifically L-plastin phosphorylated at Ser5 was raised against a peptide encoding L-plastin residues 2-17 in which Ser5 was phosphorylated (ARGS(P)VSDEEMMELREA) (characterised in [21]). Rabbit polyclonal antibodies against RSK1 (sc-231) or RSK2 (#9340) were purchased from Santa Cruz Biotechnology (Santa Cruz, Heidelberg, Germany) and Cell Signalling Technology (Leiden, Netherlands) respectively. Phorbol 12-myristate 13-acetate (PMA), epidermal growth factor (EGF), 8-Bromo-cAMP, H89 and PP2 were obtained from Sigma-Aldrich (Diegem, Belgium), GF109203X, SL0101 and BI-D1870 from Calbiochem Merck Millipore (Nothingham, UK) and PD98059 from Cell Signalling Technology (Leiden, Netherlands).

Treatment of Cells with Pharmacological Agents

Cells were treated with PMA at a concentration of 0.1 µM for 1 hour, with EGF at 1 ng/ml for 15 minutes, with GF109203X at 1 µM for 3 hours, with H89 at 50 µM for 1 hour, with PP2 at 10 µM for 1 hour, with PD98059 at 10 µM for 1 hour, with BI-D1870 at 5 µM for 30 minutes, with SL0101 at 80 µM for 4 hours. In case of combined treatment with activators and inhibitors, the incubation with the inhibitor was performed before the incubation with the activator. For serum starvation prior to EGF treatment, cells were cultured in the absence of serum for 24 hours.

Microarrays

RNA extraction was performed using Trizol reagent. RNA quality and concentration were evaluated spectroscopically using a NanoDrop 2000c instrument (Thermo Scientific, Erembodegem, Belgium). RNA integrity was subsequently analysed on an Agilent 2100

Bioanalyser (Agilent Technologies, Palo Alto, CA, USA). Only good quality RNA with integrity numbers > 9 was used in this study. Transcriptome profiling assays were performed using the Affymetrix Human GeneChip 1.0 ST arrays (Affymetrix, Santa Clara, CA, USA). Briefly, 250 ng of total RNA was reverse transcribed into cDNA, then transcribed into cRNA and labeled into biotinylated cRNA using the GeneChip WT PLUS Reagent kit (Affymetrix) according to the manufacturer's standard protocols (P/N 4425209 Rev.B 05/2009 and P/N 702808 Rev.6). Labeled cRNA products were randomly fragmented and hybridised onto Affymetrix GeneChips. Upon hybridisation, arrays were washed and stained using the Affymetrix GeneChip WT Terminal Labeling and Hybridisation kit, before being scanned using a GeneChip Scanner 3000. CEL files containing hybridisation raw signal intensities were imported into the Partek GS software (Partek, St. Louis, MO, USA) using Partek default options. Resulting expression data (transcript cluster level) were imported into R statistical environment for further analysis. Transcript clusters with no chromosome location were removed. Quality of the data was assessed through boxplot, relative log expression and Pearson's correlation. For differential expression analysis, LIMMA was used to compare gene expression between different conditions, according to author's recommendations [LIMMA User's Guide section 9.5]. Resulting p-values were adjusted for false discovery rate with Benjamini and Hochberg's FDR [7] and transcript clusters with FDR < 0.05 and absolute FC ≥ 1.5 were considered as significantly differentially expressed (SDE). The Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems, Redwood City, CA, USA, www.ingenuity.com) was used for transcript cluster mapping which led to the identification of differentially expressed genes (DEGs) and for data mining, including functional analyses and gene network reconstruction. Right-tailed Fisher's exact test was used to calculate a p-value for functional enrichment analysis (threshold: -log(p-value) > 1.301). Microarray expression data are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under the accession number E-MTAB-3487.

Kinase screening

A phosphorylation prediction algorithm, Kinase substrate predictor version 2.0 (KSPv2) from KINEXUS Bioinformatics Corporation (Vancouver, Canada), identified the 50 best scored candidate kinases for L-plastin phosphorylation on residue Ser5, out of which 43 were subsequently screened by KINEXUS. The following peptides corresponding to the L-plastin N-terminus and comprising residue Ser5 were chosen to be used in the kinase assays: ARGSVSDEERR (WT) and ARGSVADEERR (MT), both starting with an N-acetylalanine

(taking into account co-translational modifications described by the UniProtKB/Swiss-Prot database), as well as native MARGSVSDEERR (M-WT) and MARGSVADEERR (M-MT), both still comprising the initial methionine residue. The substitution of residue Ser7 by an alanine was done in order to be able to distinguish between Ser5 and Ser7 phosphorylation and to exclude false positives. The addition of two arginine residues (R) at the C-terminal end of the peptides was meant to ensure the adhesion of the peptides to the capture phosphocellulose filter paper following the kinase assay, and placed far enough away from the Ser5 site so as not to affect the kinase recognition of this site. Briefly, L-plastin peptides were mixed with individual protein kinases in the presence of [γ-33P] ATP for 20-40 minutes, depending on the protein kinase tested. The assay was terminated by spotting 10 µl of the reaction mixture onto a multiscreen phosphocellulose P81 plate. After removing unreacted [γ-33P] ATP from the reaction, radioactivity was quantified in a scintillation counter.

In vitro kinase assays

10 µg of recombinant L-plastin were incubated with 100 ng of recombinant kinase (RSK1, RSK2 or MSK1) obtained from SignalChem (Richmond, British Columbia, Canada) in a reaction volume of 25 µl according to the manufacturer's protocol. Following an incubation of 15 minutes at 30°C, Laemmli buffer was added, the samples were boiled at 100°C for 5 minutes and then analysed by immunoblotting.

Immunoblotting

Cells were lysed in situ in ice-cold lysis buffer (50 mM Tris-HCL pH 7.5, 150 mM NaCl, 0.1% SDS, , 5 mM EDTA, 1% Nonidet P-40, 1% Triton X-100, 1% Na-deoxycholate, 1 mM Na3VO4, 10 mM NaF, 100 µM leupeptin, 100 µM E64D) containing a cocktail of protease inhibitors (Roche Diagnostics GmbH, Howald, Luxembourg). Lysates were cleared by centrifugation at 13200 rpm for 15 minutes at 4°C. The total protein concentration was determined using the Bradford assay (Bio-Rad, Temse, Belgium). Protein separation was performed by SDS-PAGE gel electrophoresis under reducing conditions and proteins were then transferred onto nitrocellulose membranes by semi-dry transfer. The membranes were saturated with 1% BSA in Tris-buffered saline (TBS) 1x buffer supplemented with Tween 0.1% for 1 hour at RT, then incubated with primary antibodies overnight at 4°C and with secondary antibodies coupled to a fluorescent dye for 1h at RT. Antibody incubations were followed by membrane washings with TBS 1X buffer supplemented with Tween 0.1%. Signal intensities were detected by the Odyssey infrared image system (Licor, Westburg, Leusden, Netherlands). For quantification, the ratio between the intensities obtained for phosphorylated L-plastin (or phosphorylated PKA

substrates) versus total L-plastin was determined and then normalised to the mean of all the values obtained in one experiment to account for experimental variability. After scaling to the highest signal, graphical representations show means and standard deviations. Statistical significance was determined by an unpaired T-test with Welch's correction. $P<0.05$ was considered significant.

siRNA knockdowns

Small interfering RNA against RSK-1 (Hs_RPS6KA1_10) and RSK-2 (Hs_RPS6KA3_5) were purchased from Qiagen GmbH (Venio, Netherlands). 60 nM of siRNA was transfected using Lipofectamine® 2000 (Life Technologies, Gent, Belgium) according to the manufacturer's protocol.

Invasion and migration assays

Cancer cell lines were seeded in collagen I-coated (200 µg/ml) 96-well plates (Essen Imagelock, Essen Bioscience, Hertfordshire, UK). At appoximately 90% confluence, a wound was scratched across each well with the Cellplayer 96-well woundmaker (Essen Bioscience). (For siRNA knockdown, cells were transfected with siRNA 24 h before the wound was scratched.) To study invasion, cells were covered with collagen I (1.5 mg/ml) diluted in cell culture medium. To study migration, cell culture medium was added to the cells. Wound confluence was monitored with the Incucyte LiveCell Imaging System (Essen Bioscience) by measuring cell confluence every 3 hours for a total of 72 hours. The graphs depict means from three biological replicates each with at least five technical replicates +/- standard error of the mean (SEM).

**** The following paragraph is written by P.Trairatphisan**

Modelling

The literature-derived model topology of the L-plastin signalling including the interactions between src, PKC, PKA and the ERK/MAPK pathway were described as Boolean rules with corresponding selection probabilities in the probabilistic Boolean network (PBN) framework (see review in [52]). The experimental data obtained for ratios of phosphorylated L-plastin versus total L-plastin (P-LPL/LPL) and phosphorylated PKA substrates versus total L-plastin (P-PKAsubstrates/LPL) from the four cell lines were used for model contextualization. Normalisation of western blot data was performed as described above in the immunoblotting section. Data generated from different experimental sets were normalised to the calibrator PMA, subsequently pooled and scaled to the maximal value. We applied an improved version of the optPBN toolbox [51] to optimise the selection probabilities of the L-plastin signalling model in the PBN format.

Optimisation was performed on a stand-alone machine (Intel CPU Xeon @3.50GHz, 16GB Ram) for the different model variants based on random initial conditions. Bootstrapping was performed by randomly sampling 100 artificial data sets based on mean and standard deviation as acquired from our experimental datasets. Optimisation was subsequently performed 100 times to identify the distribution of the identified selection probabilities.

Results

Invasive breast cancer cell lines display higher baseline L-plastin Ser5 phosphorylation than non-invasive breast cancer cell lines

The four breast cancer cell lines MCF7, SK-BR-3, BT-20 and MDA-MB-435S chosen as a working model for this study were carefully selected in order to cover different clinicopathological features [25, 27, 47] with the main prerequisite being the expression of our protein of interest L-plastin. It is noteworthy however, that considerable doubt had been raised about the origin of the cell clone MDA-MB-435S. Indeed evidence was provided that MDA-MB-435S cells and the M14 melanoma cell line are genetically identical, which was confirmed by microsatellite analysis [18, 40]. Even though the correct origin of the two cell lines is still under debate, the most recent findings strongly suggest that both cell lines may be of breast origin, which substantiates our selection [10, 19]. For this study the invasive phenotype of the chosen breast cancer cell lines is of particular importance, with MCF7 and SK-BR-3 being considered as non- or merely weakly invasive cell lines in contrast to BT-20 and MDA-MB-435S which have been described as invasive cell lines [29, 39, 49, 60]. By performing *in vitro* scratch wound assays (Essen BioScience CellPlayer™) we reassessed the invasive capacities of our four cell lines. In line with the literature, our results confirmed that MCF7 and SK-BR-3 cells are only weakly invasive whereas invasiveness was considerably increased in BT-20 and highly increased in MDA-MB-435S cells (Fig. 1a). Since a correlation between the invasive capacity of melanoma cells and the phosphorylation state of L-plastin on residue Ser5 has been suggested before [28], we continued by investigating L-plastin Ser5 phosphorylation in the four breast cancer cell lines. To this end, we used an antibody recognising specifically L-plastin phosphorylated on residue Ser5 (anti-Ser5-P antibody) raised and characterised by our group [1, 21, 22]. Although the L-plastin expression level is higher in the invasive as compared to the non-invasive cell lines, our results clearly show high baseline L-plastin Ser5 phosphorylation in the invasive cell lines BT-20 and MDA-MB-435S as compared to absent or extremely weak phosphorylation in the non-invasive cell lines MCF7 and SK-BR-3 (Fig. 1b).

The ERK/MAPK pathway is enriched in DEGs when comparing breast cancer cells with differential Ser5 L-plastin phosphorylation

Whole genome microarray analysis was performed to detect differences in signal transduction pathways leading to differential L-plastin Ser5 phosphorylation in non-invasive breast cancer cells with low baseline Ser5 phosphorylated L-plastin, versus invasive breast cancer cells, with constitutively Ser5 phosphorylated L-plastin. In addition, we treated the two non-invasive cell lines with phorbol-12-myristate-13-acetate (PMA), as this treatment had been shown previously to increase L-plastin Ser5 phosphorylation in MCF7 cells [1]. As expected, PMA treatment led to a considerable increase in L-plastin Ser5 phosphorylation in both cell lines MCF7 and SK-BR-3 as compared to their respective untreated control. (Fig. 2 left). In order to perform microarray analysis, total RNA was isolated from triplicate cultures and microarray data were obtained using Affymetrix technologies. Lists of differentially expressed transcript clusters were established as described in the materials and methods section and transcript cluster mapping was performed with IPA. For the comparison of PMA-treated versus untreated controls, we thus obtained a first list of DEGs for MCF7 and a second for SK-BR-3 cells. For the comparison of our two model invasive and two model non-invasive cancer cell lines, we obtained another four lists of DEGs (Fig. 2 right), the intersection of which was retained as the third list of DEGs to be analysed. Taking the intersection of the 4 lists enabled us to focus on the genes that are differentially expressed between all comparisons of invasive versus non-invasive cells, thus setting a stringent filter. At the end, a detailed analysis using Ingenuity Pathway Analysis® (IPA) was performed with the three lists of DEGs described above. The three lists of differentially expressed genes can be found in electronic supplementary material (ESM_List1, ESM_List2 and ESM_List3).

A heatmap based on the four sample sets obtained with the untreated four cell lines and integrating the differentially expressed transcript clusters of the intersection of the 4 corresponding lists shows that the vast majority of these genes are regulated in opposite directions depending on the invasive capacities of the cell line (ESM_Fig1). Thus the DEGs of the intersection clearly partition the cell lines into two distinct groups: invasive and non-invasive.

Most importantly, in order to assess a difference in signalling pathways in the compared conditions, we focused on canonical signalling pathways in IPA. For each of the 3 comparisons of interest, IPA revealed a list of canonical signalling pathways (ESM_CP_List1, ESM_CP_List2, ESM_CP_List3) from which we selected those that were

significant (*P*-value ≤ 0.05) and thus enriched by genes that are significantly differentially expressed. Interestingly, we identified 3 canonical signalling pathways that were common to the 3 lists: ERK/MAPK Signalling, Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis and UVA-Induced MAPK Signalling (Fig. 2). These results suggest an involvement of one or more of these pathways in L-plastin Ser5 phosphorylation.

RSK1 as well as RSK2 specifically phosphorylate residue Ser5 of recombinant full-length L-plastin *in vitro*

L-plastin peptides were synthesised and screened for Ser5 phosphorylation by 43 candidate kinases identified using the phosphorylation prediction algorithm, *Kinase substrate predictor version 2.0 (KSPv2)* from KINEXUS Bioinformatics Corporation (Vancouver, Canada). In addition to PKA, which was previously shown to be able to phosphorylate L-plastin Ser5 *in vitro* [21, 55], our screening identified RSKs as well as MSK1 as candidate kinases for phosphorylating this residue. As shown in Table 1, these kinases led to high radioactivity counts for the wild type L-plastin peptide (WT) as well as for the Ser7-to-alanine mutated peptide (MT), both peptides being devoid of the initiator methionine and acetylated on the N-terminal alanine, thus taking into account information about co-translational modifications of L-plastin described by the UniProtKB/Swiss-Prot database. The WT peptide displayed higher counts for the named kinases than the MT peptide, indicating that phosphorylation occurs not only at residue Ser5, but also at Ser7. However, notably Ser7 phosphorylation is not required for Ser5 phosphorylation by the investigated kinases and, importantly, the same kinases appear in the top positions for both peptides. Validation experiments as well as experiments with the WT and MT peptides still containing the initiator methionine were performed and showed similar rankings (data not shown).

As the *in vitro* kinase assays performed on L-plastin peptides identified RSK1, RSK2 and MSK1 as candidate kinases (Table 1), we also tested these kinases for their ability to phosphorylate recombinant full-length L-plastin on residue Ser5 *in vitro*. As shown in Figure 3, both RSK1 and RSK2 were able to strongly phosphorylate L-plastin on residue Ser5 while MSK1 was merely able to induce a weak L-plastin Ser5 phosphorylation.

Table 1 For *in vitro* kinase assays on L-plastin peptides, the peptides were mixed with individual protein kinases in the presence of [γ-^{33}P] ATP for 20-40 minutes, depending on the protein kinase tested. The assay was terminated by spotting 10 µl of the reaction mixture onto a multiscreen phosphocellulose P81 plate. After removing unreacted [γ-^{33}P] ATP from the reaction, radioactivity was quantified in a scintillation counter.

Ranking	Kinases	Counts (cpm)	Kinases	Counts (cpm)
		*ARGSVSDEERR (WT)		*ARGSVADEERR (MT)
1	RSK2	116698	RSK1	55774
2	PKAca	109766	PKAca	53986
3	RSK1	101952	MSK1	53254
4	MSK1	86947	RSK2	48800
5	RSK3	71917	PKAcb	34148
6	PKAcb	54673	PKAcg	33572
7	RSK4	45174	RSK3	32917
8	PKAcg	40167	PRKG2	26041
9	SGK3	32646	RSK4	22477
10	SGK2	32129	SGK2	22249
11	PRKG2	29921	SGK3	20290
12	PKCh	29273	STK33	10761
13	p70S6Kb	19894	PKCh	10007
14	AURORA B	12688	AURORA B	9696
15	STK33	9990	PRKG1	9110
16	PKCq	9671	PRKX	8113
17	VRK1	9054	DCAMKL1	5899

18	PRKX	8310	VRK1	5745
19	PRKG1	8189	VRK2	5470
20	CAMK1b	8126	MNK1	5430
21	VRK2	7876	PKCe	5053
22	CAMK4	6955	CAMK4	4668
23	PKCe	6779	DCAMKL2	4447
24	DCAMKL2	5726	p70S6Kb	4242
25	MNK1	5718	CAMK1b	3996
26	DCAMKL1	5437	NDR	3797
27	NDR	5072	PKCq	3589
28	PKCd	4597	IKKe	3422
29	PIM2	4405	NDR2	3284
30	SGK1	4243	ASK1	2771
31	NDR2	3821	SGK1	2540
32	MSK2	3741	PKCd	2329
33	IKKe	3437	AKT1	1920
34	AKT1	3257	p70S6K	1911
35	ASK1	3241	PIM2	1813
36	PIM3	2732	PIM3	1537
37	p70S6K	2454	AKT3	1460
38	AKT3	2196	AKT2	1242
39	AKT2	2142	MSK2	1005
40	PIM1	892	PIM1	922
41	CK2a2	892	AURORA C	663

42	AURORA C	588	CHK1	256
43	CHK1	176	CK2a2	44

*N-terminal alanine is acetylated

The ERK/MAPK pathway and its downstream RSK kinases are involved in L-plastin Ser5 phosphorylation in breast cancer cells

The two newly identified prime candidate kinases capable of L-plastin Ser5 *in vitro* phosphorylation, RSK1 and RSK2, are downstream effectors of the ERK/MAPK pathway. In addition, the results of our microarray experiments have also pointed to a potential involvement of the ERK/MAPK pathway in L-plastin Ser5 phosphorylation. Altogether these findings prompted us to further investigate the role of the ERK/MAPK pathway in L-plastin Ser5 phosphorylation. In a first step, we inhibited selected signaling molecules of this pathway and we showed that PKC inhibition with the widely used pan-PKC inhibitor GF109203X, MEK1/MEK2 inhibition with PD98059 as well as RSK inhibition with BI-D1870 decreased baseline L-plastin Ser5 phosphorylation in invasive cells (Fig. 4a). In a second step, we took advantage of the fact that the ERK/MAPK pathway is one of the major signalling pathways stimulated upon binding of various growth factors to the corresponding receptor tyrosine kinases. Knowing that the EGFR/ErbB family, a prominent receptor tyrosine kinase family, plays a key role in normal and malignant breast development [15] and is capable to trigger cell migration through ERK/MAPK pathway signalling [50], we checked for EGFR expression in our four model cell lines. As previously decribed by others [43, 47], both SK-BR-3 and BT-20 cells expressed high levels of EGFR, whereas for MCF7 and MDA-MB-435S cells the expression was very low to absent (ESM_Fig2a). As expected, stimulation with EGF as a key EGFR binding ligand did not increase L-plastin phosphorylation in the two EGFR-negative cell lines MCF7 and MDA-MB-435S, even when stimulated with high EGF concentrations (ESM_Fig2b). However, EGF treatment highly increased L-plastin Ser5 phosphorylation in SK-BR-3 and BT-20 cells at all tested concentrations (ESM_Fig2b). We have shown previously that treatment with the phorbol ester PMA, a well-known activator of PKCs which has also been described to activate the ERK/MAPK pathway [3, 8, 26], was able to induce L-plastin Ser5 phosphorylation in MCF7 [1] and SK-BR-3 cells (Fig. 2). Accordingly, PMA treatment of the two invasive cell lines BT-20 and MDA-MB-435S, displaying already high baseline L-plastin Ser5 phosphorylation, was able to further increase this phosphorylation level (Fig. 4b). Importantly, preincubation with inhibitors of the ERK/MAPK pathway impaired L-

plastin phosphorylation upon PMA or EGF stimulation in all tested cell lines (Fig. 4b). To exclude off-target effects, we confirmed our results with a second RSK inhibitor, SL0101, which impaired EGF-triggered L-plastin phosphorylation in SK-BR-3 and BT-20 cells and decreased baseline L-plastin phosphorylation in invasive cell lines (data not shown). Interestingly, the strongest decrease of L-plastin Ser5 phosphorylation was obtained for the combined inhibition of RSK and PKC (Fig. 4b). And finally, Trametinib, a clinical MEK inhibitor approved by the US Food and Drug Administration (FDA) for melanoma treatment (trade name Mekinist), clearly reduced L-plastin Ser5 phosphorylation in invasive cell lines and prevented an EGF-triggered increase in L-plastin Ser5 phosphorylation (Fig. 4c). Altogether these results point to a major involvement of the ERK/MAPK signalling pathway in L-plastin Ser5 phosphorylation.

**** The following section, Table 2, and Table 3 are contributed by P.Trairatphisan**

Modelling of the L-plastin signalling pathway with *probabilistic Boolean networks*

The signalling pathways leading to L-plastin Ser5 phosphorylation are not clearly understood and, so far, it was assumed that the identity of the protein kinase responsible for this phosphorylation event depends on the cell type and environment. In addition to reports on the involvement of PKA [21, 33, 55] and PKC [1, 16, 22, 24, 32, 37, 38], our results presented here highlight a role of RSK kinases in this process. In order to get a better quantitative understanding of the signalling pathways upstream of L-plastin Ser5 phosphorylation, we applied the *optPBN* toolbox to further study and analyse these pathways in the probabilistic Boolean network (PBN) framework [51]. We built a PBN model which represents the network topology of signalling pathways upstream of L-plastin Ser5 phosphorylation based on literature information and own findings (Fig. 5a). Our model is mainly focused on the ERK/MAPK pathway (downstream of the EGFR) and includes src, PKC and PKA kinases which are known to interact with this pathway. Then, we fitted the PBN model to an extensive dataset comprising activation and inhibition of various network nodes which modulate the signals towards the two measured output nodes, i.e. L-plastin Ser5 phosphorylation and PKA substrate phosphorylation in four breast cancer cell lines. To this end, we took into account the described off-target effects of the inhibitors GF109203X and H89 on RSK [2, 11]. As a result, we obtained a model which explained well our experimental data for all four cell lines as shown in Figure 5b.

As we found RSK kinases to be essential for L-plastin Ser5 phosphorylation, the fitted PBN model was used to investigate whether RSK kinases have more influence on L-plastin in the four cell lines than PKA and PKC. In addition, we checked whether the

crosstalk interactions between PKC and PKA suggested in the literature [48, 59] are important to modulate the signal transduction within the L-plastin signalling pathway. We therefore applied an *in silico* knock-out approach where we removed an interaction from the model one-at-a-time and checked if the removal affected the fitting quality. Five interactions situated in close proximity to our output nodes were analysed, i.e. RSK→L-plastin, PKC→L-plastin, PKA→L-plastin, PKC→PKA and PKA→PKC (Table 2). We found that removing the interactions RSK→L-plastin and PKC→PKA led to a dramatic increase of the model fitting costs in all four cell lines, meaning that the networks missing one of these two interactions fitted our experimental data less well. The individual knock-out of the other three interactions only led to minor changes in fitting costs, suggesting that none of these three interactions is necessary to explain our experimental data.

We then proceeded by examining the optimised selection probability (weight) of each interaction and their distributions obtained via bootstrapping (see Methods section). A subset of the obtained weight distributions of the interactions is shown in Table 3. All weights can be found in electronic supplementary material (ESM_Fig3). Strikingly, in all four cell lines, the activation of L-plastin by RSK is largely predominant as compared to its activation by PKC or PKA. In addition, the weights attributed to the investigated interactions indicate that in all 4 cell lines the crosstalk appears to be directed from PKC to PKA rather than from PKA to PKC. The relatively low standard deviations on the weights ensured that these findings are robust against experimental variation in the dataset.

Table 2 The fitting costs of model variants after removing individual interactions were compared to those of the initial model structure prior to removal. The fitting cost is the sum of squared error between simulated state values and the mean values of the experimental data.

	MCF7	SK-BR-3	BT-20	MDA-MB-435S
Initial model	0.2298	0.2942	0.2961	0.3126
RSK -> LPL removed	0.7872	1.0344	0.6921	1.2369
PKC -> LPL removed	0.2298	0.2942	0.2962	0.3126

PKA -> LPL removed	0.2318	0.2942	0.2966	0.3126
PKC -> PKA removed	0.7163	0.4520	0.4245	0.4622
PKA -> PKC removed	0.2298	0.2942	0.3045	0.3452

Table 3 The distributions of identified weights obtained via bootstrapping are shown for the five interactions under investigation in the four cell lines shown as mean and (standard deviation).

	MCF7	SK-BR-3	BT-20	MDA-MB-435S
RSK -> LPL	0.888 (0.081)	1.000 (0.000)	0.948 (0.070)	1.000 (0.000)
PKC -> LPL	0.004 (0.023)	0.000 (0.000)	0.025 (0.058)	0.000 (0.000)
PKA -> LPL	0.108 (0.073)	0.000 (0.000)	0.027 (0.023)	0.000 (0.000)
PKC -> PKA	0.407 (0.060)	0.236 (0.049)	0.262 (0.059)	0.252 (0.093)
PKA -> PKC	0.072 (0.112)	0.005 (0.017)	0.256 (0.189)	0.204 (0.068)

Combined RSK1 and RSK2 knockdown decreases L-plastin Ser5 phosphorylation and impairs migration and invasion in breast cancer cells

Altogether, we provide evidence that RSK kinases are involved in L-plastin Ser5 phosphorylation by *in vitro* kinase assays, by inhibition with two different RSK inhibitors as well as by computational modelling. To consolidate our findings, we simultaneously knocked down RSK1 and RSK2 by an siRNA approach, which resulted in an important decrease of L-plastin Ser5 phosphorylation in our four breast cancer cell lines (Fig. 6a). Even though phosphorylation was not completely abolished, our data clearly reveal an important role for RSKs in L-plastin Ser5 phosphorylation in all four model cell lines. Some RSK1 and RSK2 protein was still expressed after knockdown and may account for the remaining L-plastin phosphorylation.

Finally, we investigated the effect of RSK knockdown on migration and invasion in MDA-MB-435S. This cell line was selected as it has the highest invasive capacity as shown in Figure 1a. *In vitro* scratch wound assays (Essen BioScience CellPlayer™) revealed that

the combined knockdown of RSK1 and RSK2 considerably slowed down both migration and invasion of MDA-MB-435S cells (Fig. 6b), whereas cell proliferation remained largely unaffected (data not shown).

Discussion

In our former studies we have shown that phosphorylation of L-plastin on residue Ser5 plays a critical role for L-plastin activation [1, 21]. In this study, we have unraveled a major signalling pathway responsible for L-plastin Ser5 phosphorylation in cancer cells. We have revealed the involvement of the ERK/MAPK pathway with its downstream target kinases RSK1 and RSK2 being able to directly phosphorylate L-plastin on Ser5. In line with previous studies showing L-plastin Ser5 phosphorylation to be crucial for cell invasion and metastasis formation [21, 28, 42], we have consolidated in our model an involvement of RSK1 and RSK2 in cancer cell invasion.

In this study, we chose to take a microarray-based gene expression profiling approach in order to correlate conditions and cell lines presenting differential L-plastin Ser5 phosphorylation levels with patterns of changes in gene expression. In the context of the cofilin pathway in breast cancer invasion and metastasis [56], the authors have pointed out that not only individual genes, but whole pathways with differential regulation and activity states of the corresponding molecules should be taken into account for phenotype interpretation. Similarly, in our study the whole genome microarray analysis approach allowed us to identify three canonical pathways enriched in DEGs, namely ERK/MAPK Signalling, Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis and UVA-Induced MAPK. In addition, the in vitro kinase assays carried out in parallel using L-plastin peptides identified the ERK/MAPK pathway downstream kinases RSK1 and RSK2 as the most prominent candidate kinases for L-plastin Ser5 phosphorylation. These findings together with a previous report describing L-plastin as a MAP kinase pathway regulated protein (Lewis, TS, 2000), prompted us to proceed to an in depth investigation of this pathway, whose deregulation has been described in a large subset of human breast cancers [57].

Several approaches were taken to unravel the involvement of the ERK/MAPK pathway with its downstream kinases RSK1 and RSK2 in the L-plastin Ser5 phosphorylation event. To trigger this pathway we stimulated the cells with EGF or with PMA, both described as activators of the ERK/MAPK pathway. Evidence for a direct involvement of the ERK/MAPK pathway downstream kinases RSK1 and RSK2 in L-plastin Ser5 phosphorylation was provided, based both on a siRNA knockdown approach and on the use of two different RSK inhibitors BI-D1870 and SL0101, thus reducing the risk that the inhibitor-dependent phosphorylation decrease is due to off-target effects. Indeed, BI-

D1870 and SL0101 appear to have only one common off-target, Aurora B [6], which showed considerably weaker potency in phosphorylating L-plastin peptides than RSK1 or RSK2 as shown by the in vitro kinase assays from KINEXUS (Table 1). In line and most importantly, the two kinases RSK1 as well as RSK2 were able to directly phosphorylate the recombinant L-plastin protein on residue Ser5 in an in vitro kinase assay. Further evidence for the importance of the ERK/MAPK pathway comes from the data obtained with the MEK inhibitor PD98059 as well as with the clinically used inhibitor Trametinib (trade name Mekinist). Moreover, our finding that L-plastin phosphorylation can be mediated by RSK upon activation of the ERK/MAPK pathway was confirmed by our computational model which suggests that RSK is the most important activator of L-plastin in all 4 cell lines. Our PBN modelling approach proved to be a useful tool to integrate the information from literature together with our vast experimental dataset, in particular because signalling tends to occur in complex networks.

Although our results provide strong evidence for a role of the ERK/MAPK pathway with the downstream kinases RSK1 and RSK2 being able to directly phosphorylate L-plastin on residue Ser5, they do not rule out that L-plastin Ser5 phosphorylation can also be mediated by other pathways. So far, only PKA has been reported to directly phosphorylate L-plastin in vitro [21, 55], whereas other kinases such as PKCδ, catalytic domain of PKC, casein kinase II, Pak1, protein kinase B, M2K, M3K and p38-regulated/activated protein kinase failed to directly phosphorylate L-plastin in vitro [24, 55]. In cells, by contrast, distinct protein kinases have been reported to play a role in triggering L-plastin phosphorylation depending on the cell type and environment. Most frequent are reports of the involvement of PKA [21, 33, 55] and PKC [1, 16, 22, 24, 32, 37, 38]. PI3K has also been reported to play a role in L-plastin phosphorylation in human neutrophils [24, 37], but not in T lymphocytes [16]. Even though an siRNA knockdown approach was successful in reducing L-plastin Ser5 phosphorylation for PKCδ [1, 22] and PKCβII [38], it has to be considered that most studies were based on activation and/or inhibition studies. Strikingly, all inhibitors used to demonstrate the involvement of PKA and PKC (H89, GF109203X, Gö6976 and Ro-31-8220) in L-plastin phosphorylation also strongly inhibit RSK2 [2, 11]. Moreover, PMA, used as a PKC activator, does not only activate PKC but has also been shown to activate the ERK1/2 module either through PKC and c-Src or through RasGRP [3, 8, 26]. Finally, the effect of the PKA activator cAMP on the ERK1/2 module appears to depend on the cellular context since cAMP has been demonstrated to stimulate ERK in a B-Raf dependent manner or to suppress ERK signalling in many cells through its ability to target C-Raf (reviewed in [13]). In the present study we have demonstrated a clear and direct activity of RSK1 and RSK2 kinases in L-plastin phosphorylation, supported by a

modelling approach. A potential involvement of the previously described kinases, PKA, PKC and PI3K, cannot be ruled out, although, in our model, PKA and PKC appear to play a more indirect role in L-plastin activation than RSK. This result is in line with previous findings of Hagi et al., who reported that cAMP stimulation was not able to trigger L-plastin Ser5 phosphorylation in macrophage [17]. Regarding a potential involvement of PI3K, future studies, which were out of the scope of the present work, have to be dedicated to the investigation of the PTEN/PI3K/AKT/mTOR pathway in mediating L-plastin Ser5 phosphorylation.

RSKs have been described as versatile regulators controlling migration and invasion downstream of ERK/MAPK activation (reviewed in Sulzmaier and Ramos, Cancer Research, 2013), by altering the transcription of many genes involved in EMT, by regulating cell adhesion through phosphorylation events with subsequent integrin activity modulation and/or by remodeling the actin cytoskeleton. In this context, RSK2 expression has been correlated to the expression of the actin-bundling protein Fascin-1 in tumor samples from HNSCC patients and to filopodia formation in cancer cells (Li D, JBC, 2013), suggesting that the RSK2-CREB pathway increases proinvasive and prometastatic capacities. Interestingly, RSK1 has been described to phosphorylate a further actin-bundling protein, filamin A [36], on residue Ser2152 [58], a site required for membrane ruffling [53]. In our study, both RSK1 and RSK2 were able to phosphorylate the actin-binding protein L-plastin and combined knockdown of the two isoforms led to a significant decrease in cell migration and invasion. The identification of L-plastin as a target of RSK kinases consolidates a role for RSK in the regulation of the actin cytoskeleton. Our results extend the findings of Doehn and colleagues, who investigated the effects of RSK1 and RSK2 on migration and invasion in epithelial breast cells and who showed that combined knockdown of RSK1 and RSK2 or RSK inhibition suppressed the ERK pathway dependent induction of promotile and proinvasive genes [12]. We suggest that the effect of RSK knockdown on the promotile/invasive capacities of cancer cells may also be due, in part, to the decrease of L-plastin Ser5 phosphorylation upon RSK knockdown. The link between L-plastin Ser5 phosphorylation and the motile capacities of cells may consist in the improved capacity of the phosphorylated form to regulate the actin cytoskeleton.

Overall, our results show the first evidence, to our knowledge, of the involvement of the ERK/MAPK pathway in L-plastin Ser5 phosphorylation. These findings corroborate Ser5 phosphorylated L-plastin as a molecular marker for invasive carcinomas with deregulated ERK/MAPK pathway signaling. Furthermore, our identification of the signalling molecules leading to L-plastin Ser5 phosphorylation consolidates them as potential targets for cancer

therapies as L-plastin phosphorylation has been shown to promote metastasis and invasion [21, 28, 42]. As opposed to MEK and Raf, RSKs are not “global regulators” and thus their inhibition by therapeutic drugs may have less severe side effects. Knowing that RSKs are overexpressed in approximately 50% of human breast cancer tissues [46], this kinase family can be considered as a highly promising therapeutic drug target in certain invasive carcinomas.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

This work was supported by a grant from the University of Luxembourg (IRP: PhosphoPlast) and through funds from the Luxembourg Ministry for Higher Education and Research. Maiti Lommel and Panuwat Trairatphisan are recipients of fellowships allocated by the Fonds National de la Recherche (AFR grant numbers 2748401 and 1233900, respectively). The authors are grateful to François Bernardin for technical assistance with microarray experiments and to Nicolas Beaume and Aurélien Ginolhac for help with statistical data analysis. The authors wish to thank Maria Koffa and Serge Haan for proof-reading of the manuscript and for constructive discussions as well as Evelyne Friederich as an initiator of this research project.

Electronic supplementary material

ESM_Fig1 Heatmap of differentially expressed transcript clusters between invasive and non-invasive cell lines. BT-20 and MDA-MB-435S were successively compared to MCF7 and to SK-BR-3. FDR and absolute FC thresholds of 0.05 and 1.5 were used to select differentially expressed features and expression profiles of 1858 transcript clusters detected in the four comparisons are depicted in the figure as heatmap. Transcript clusters are shown in rows and samples (triplicates for each cell line) in columns. Row and column orders were determined by hierarchical clustering and colors represent transcript cluster expression level as z-score.

ESM_Fig2 Analysis of EGF-dependent L-plastin Ser5 phosphorylation. a EGF receptor (EGFR) is expressed in SK-BR-3 and BT-20 breast cancer cell lines. Cell extracts were subjected to immunoblot analysis using antibodies specific for the EGFR and β-actin, used as a loading control. b Cells were stimulated with different EGF concentrations as indicated and cell extracts were subjected to immunoblot analysis using antibodies specific for Ser5 phosphorylated L-plastin (P-LPL) and total L-plastin (LPL).

ESM_Fig3 Optimised interactions' weights for the L-plastin signalling network. Bootstrapping was performed by randomly sampling 100 artificial data sets based on mean and standard deviation as acquired from the wet lab experiments. Optimisation was subsequently performed 100 times to identify the distribution of the identified selection probabilities. Means and standard deviations of the interactions' weights were compared among the four cell lines.

ESM_ListDEG1 List of differentially expressed genes between MCF7+PMA and MCF7 identified by whole genome microarray analysis using Affymetrix Human GeneChip 1.0 ST arrays.

ESM_ListDEG2 List of differentially expressed genes between SK-BR-3+PMA and SK-BR-3 identified by whole genome microarray analysis using Affymetrix Human GeneChip 1.0 ST arrays.

ESM_ListDEG3 List of differentially expressed genes, between invasive and non-invasive breast cancer cells (intersection of comparisons BT-20 vs MCF7, MDA-MB-435S vs MCF7, BT-20 vs SK-BR-3 and MDA-MB-435S vs SK-BR-3), identified by whole genome microarray analysis using Affymetrix Human GeneChip 1.0 ST arrays.

ESM_CP_List1 List of canonical pathways identified by IPA for the comparison of MCF7+PMA and MCF7.

ESM_CP_List2 List of canonical pathways identified by IPA for the comparison of SK-BR-3+PMA and SK-BR-3.

ESM_CP_List3 List of canonical pathways identified by IPA for the comparison of BT-20 vs MCF7, MDA-MB-435S vs MCF7, BT-20 vs SK-BR-3 and MDA-MB-435S vs SK-BR-3.

References

1. Al Tanoury Z, Schaffner-Reckinger E, Halavatyi A et al. (2010) Quantitative kinetic study of the actin-bundling protein L-plastin and of its impact on actin turn-over. PLoS One 5:e9210
2. Alessi DR (1997) The protein kinase C inhibitors Ro 318220 and GF 109203X are equally potent inhibitors of MAPKAP kinase-1beta (Rsk-2) and p70 S6 kinase. FEBS letters 402:121-123
3. Amos S, Martin PM, Polar GA et al. (2005) Phorbol 12-myristate 13-acetate induces epidermal growth factor receptor transactivation via protein kinase Cdelta/c-Src pathways in glioblastoma cells. J Biol Chem 280:7729-7738
4. Arpin M, Friederich E, Algrain M et al. (1994) Functional differences between L- and T-plastin isoforms. J Cell Biol 127:1995-2008
5. Arteaga CL, Sliwkowski MX, Osborne CK et al. (2012) Treatment of HER2-positive breast cancer: current status and future perspectives. Nature reviews. Clinical oncology 9:16-32
6. Bain J, Plater L, Elliott M et al. (2007) The selectivity of protein kinase inhibitors: a further update. Biochem J 408:297-315
7. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) 57:289-300
8. Brose N, Rosenmund C (2002) Move over protein kinase C, you've got company: alternative cellular effectors of diacylglycerol and phorbol esters. Journal of cell science 115:4399-4411
9. Campbell PM, Der CJ (2004) Oncogenic Ras and its role in tumor cell invasion and metastasis. Seminars in cancer biology 14:105-114
10. Chambers AF (2009) MDA-MB-435 and M14 cell lines: identical but not M14 melanoma? Cancer Res 69:5292-5293
11. Davies SP, Reddy H, Caivano M et al. (2000) Specificity and mechanism of action of some commonly used protein kinase inhibitors. Biochem J 351:95-105
12. Doehn U, Hauge C, Frank SR et al. (2009) RSK is a principal effector of the RAS-ERK pathway for eliciting a coordinate promotile/invasive gene program and phenotype in epithelial cells. Molecular cell 35:511-522
13. Dumaz N, Marais R (2005) Integrating signals between cAMP and the RAS/RAF/MEK/ERK signalling pathways. Based on the anniversary prize of the

Gesellschaft fur Biochemie und Molekularbiologie Lecture delivered on 5 July 2003 at the Special FEBS Meeting in Brussels. The FEBS journal 272:3491-3504

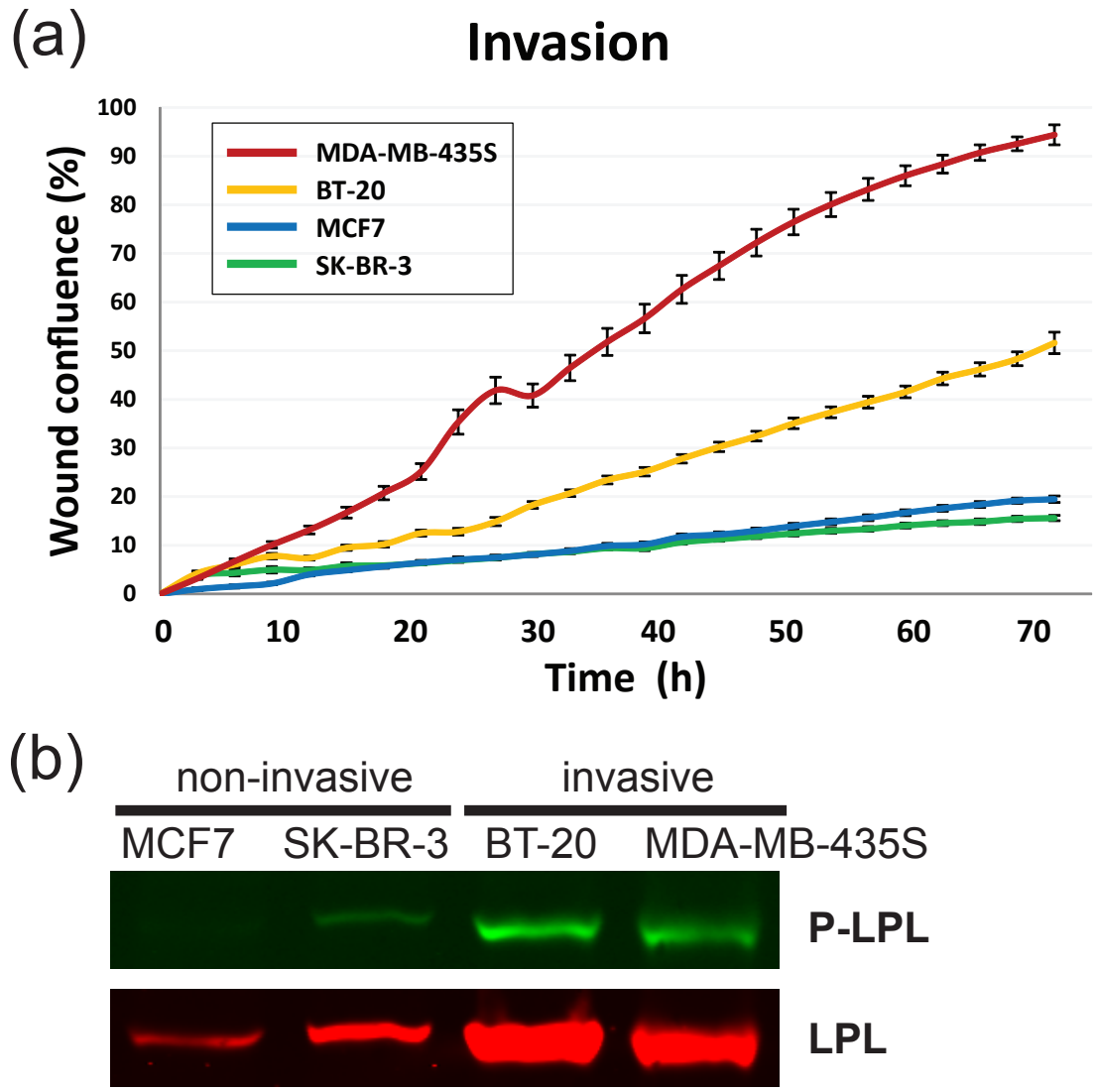
14. Dunn KL, Espino PS, Drobic B et al. (2005) The Ras-MAPK signal transduction pathway, cancer and chromatin remodeling. Biochemistry and cell biology = Biochimie et biologie cellulaire 83:1-14
15. Eccles SA (2011) The epidermal growth factor receptor/Erb-B/HER family in normal and malignant breast biology. The International journal of developmental biology 55:685-696
16. Freeley M, O'dowd F, Paul T et al. (2012) L-plastin regulates polarization and migration in chemokine-stimulated human T lymphocytes. J Immunol 188:6357-6370
17. Hagi A, Hirata H, Shinomiya H (2006) Analysis of a bacterial lipopolysaccharide-activated serine kinase that phosphorylates p65/L-plastin in macrophages. Microbiology and immunology 50:331-335
18. Hollestelle A, Nagel JH, Smid M et al. (2010) Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. Breast cancer research and treatment 121:53-64
19. Hollestelle A, Schutte M (2009) Comment Re: MDA-MB-435 and M14 cell lines: identical but not M14 Melanoma? Cancer Res 69:7893
20. Hynes NE, Macdonald G (2009) ErbB receptors and signaling pathways in cancer. Current opinion in cell biology 21:177-184
21. Janji B, Giganti A, De Corte V et al. (2006) Phosphorylation on Ser5 increases the F-actin-binding activity of L-plastin and promotes its targeting to sites of actin assembly in cells. Journal of cell science 119:1947-1960
22. Janji B, Vallar L, Al-Tanoury Z et al. (2010) The actin filament cross-linker L-plastin confers resistance to TNF-alpha in MCF-7 breast cancer cells in a phosphorylation-dependent manner. J Cell Mol Med 14:1264-1275
23. Jones SL, Brown EJ (1996) FcgammaRII-mediated adhesion and phagocytosis induce L-plastin phosphorylation in human neutrophils. J Biol Chem 271:14623-14630
24. Jones SL, Wang J, Turck CW et al. (1998) A role for the actin-bundling protein L-plastin in the regulation of leukocyte integrin function. Proc Natl Acad Sci U S A 95:9331-9336
25. Kao J, Salari K, Bocanegra M et al. (2009) Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. PLoS One 4:e6146

26. Kazanietz MG (2000) Eyes wide shut: protein kinase C isozymes are not the only receptors for the phorbol ester tumor promoters. Mol Carcinog 28:5-11
27. Kenny PA, Lee GY, Myers CA et al. (2007) The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression. Molecular oncology 1:84-96
28. Klemke M, Rafael MT, Wabnitz GH et al. (2007) Phosphorylation of ectopically expressed L-plastin enhances invasiveness of human melanoma cells. Int J Cancer 120:2590-2599
29. Lacroix M, Leclercq G (2004) Relevance of breast cancer cell lines as models for breast tumours: an update. Breast cancer research and treatment 83:249-289
30. Lee JC, Vivanco I, Beroukhim R et al. (2006) Epidermal growth factor receptor activation in glioblastoma through novel missense mutations in the extracellular domain. PLoS medicine 3:e485
31. Libermann TA, Nusbaum HR, Razon N et al. (1985) Amplification, enhanced expression and possible rearrangement of EGF receptor gene in primary human brain tumours of glial origin. Nature 313:144-147
32. Lin CS, Lau A, Lue TF (1998) Analysis and mapping of plastin phosphorylation. DNA Cell Biol 17:1041-1046
33. Matsushima K, Kobayashi Y, Copeland TD et al. (1987) Phosphorylation of a cytosolic 65-kDa protein induced by interleukin 1 in glucocorticoid pretreated normal human peripheral blood mononuclear leukocytes. J Immunol 139:3367-3374
34. Mok TS (2011) Personalized medicine in lung cancer: what we need to know. Nature reviews. Clinical oncology 8:661-668
35. Morley SC (2012) The actin-bundling protein L-plastin: a critical regulator of immune cell function. International journal of cell biology 2012:935173
36. Ohta Y, Hartwig JH (1996) Phosphorylation of actin-binding protein 280 by growth factors is mediated by p90 ribosomal protein S6 kinase. J Biol Chem 271:11858-11864
37. Paclet MH, Davis C, Kotsonis P et al. (2004) N-Formyl peptide receptor subtypes in human neutrophils activate L-plastin phosphorylation through different signal transduction intermediates. Biochem J 377:469-477
38. Pazdrak K, Young TW, Straub C et al. (2011) Priming of eosinophils by GM-CSF is mediated by protein kinase CbetaII-phosphorylated L-plastin. J Immunol 186:6485-6496
39. Pishvaian MJ, Feltes CM, Thompson P et al. (1999) Cadherin-11 is expressed in invasive breast cancer cell lines. Cancer Res 59:947-952

40. Rae JM, Creighton CJ, Meck JM et al. (2007) MDA-MB-435 cells are derived from M14 melanoma cells--a loss for breast cancer, but a boon for melanoma research. Breast cancer research and treatment 104:13-19
41. Ridley AJ, Schwartz MA, Burridge K et al. (2003) Cell migration: integrating signals from front to back. Science 302:1704-1709
42. Riplinger SM, Wabnitz GH, Kirchgessner H et al. (2014) Metastasis of prostate cancer and melanoma cells in a preclinical in vivo mouse model is enhanced by L-plastin expression and phosphorylation. Molecular cancer 13:10
43. Roos W, Fabbro D, Kung W et al. (1986) Correlation between hormone dependency and the regulation of epidermal growth factor receptor by tumor promoters in human mammary carcinoma cells. Proc Natl Acad Sci U S A 83:991-995
44. Samstag Y, Eibert SM, Klemke M et al. (2003) Actin cytoskeletal dynamics in T lymphocyte activation and migration. Journal of leukocyte biology 73:30-48
45. Shinomiya H (2012) Plastin family of actin-bundling proteins: its functions in leukocytes, neurons, intestines, and cancer. International journal of cell biology 2012:213492
46. Smith JA, Poteet-Smith CE, Xu Y et al. (2005) Identification of the first specific inhibitor of p90 ribosomal S6 kinase (RSK) reveals an unexpected role for RSK in cancer cell proliferation. Cancer Res 65:1027-1034
47. Subik K, Lee JF, Baxter L et al. (2010) The Expression Patterns of ER, PR, HER2, CK5/6, EGFR, Ki-67 and AR by Immunohistochemical Analysis in Breast Cancer Cell Lines. Breast cancer : basic and clinical research 4:35-41
48. Sugita S, Baxter DA, Byrne JH (1997) Modulation of a cAMP/protein kinase A cascade by protein kinase C in sensory neurons of Aplysia. The Journal of neuroscience : the official journal of the Society for Neuroscience 17:7237-7244
49. Sugiyama N, Gucciardo E, Tatti O et al. (2013) EphA2 cleavage by MT1-MMP triggers single cancer cell invasion via homotypic cell repulsion. J Cell Biol 201:467-484
50. Tarcic G, Avraham R, Pines G et al. (2012) EGR1 and the ERK-ERF axis drive mammary cell migration in response to EGF. FASEB J 26:1582-1592
51. Trairatphisan P, Mizera A, Pang J et al. (2014) optPBN: an optimisation toolbox for probabilistic Boolean networks. PLoS One 9:e98001
52. Trairatphisan P, Mizera A, Pang J et al. (2013) Recent development and biomedical applications of probabilistic Boolean networks. Cell communication and signaling : CCS 11:46

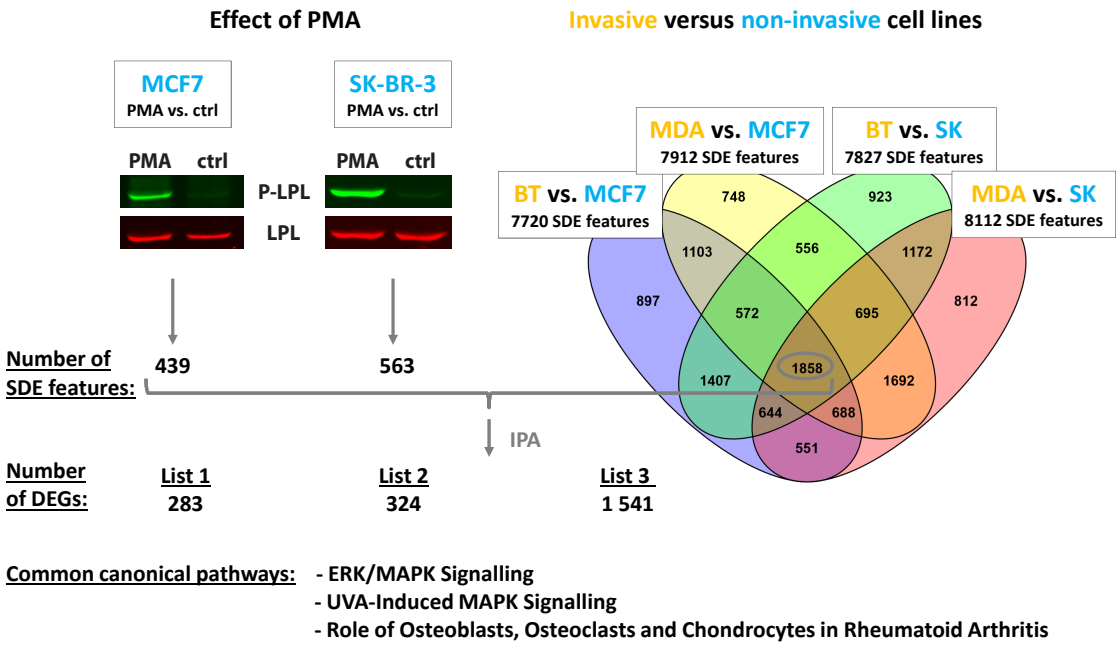
53. Vadlamudi RK, Li F, Adam L et al. (2002) Filamin is essential in actin cytoskeletal assembly mediated by p21-activated kinase 1. Nat Cell Biol 4:681-690
54. Vivanco I, Robins HI, Rohle D et al. (2012) Differential sensitivity of glioma- versus lung cancer-specific EGFR mutations to EGFR kinase inhibitors. Cancer discovery 2:458-471
55. Wang J, Brown EJ (1999) Immune complex-induced integrin activation and L-plastin phosphorylation require protein kinase A. J Biol Chem 274:24349-24356
56. Wang W, Eddy R, Condeelis J (2007) The cofilin pathway in breast cancer invasion and metastasis. Nat Rev Cancer 7:429-440
57. Whyte J, Bergin O, Bianchi A et al. (2009) Key signalling nodes in mammary gland development and cancer. Mitogen-activated protein kinase signalling in experimental models of breast cancer progression and in mammary gland development. Breast cancer research : BCR 11:209
58. Woo MS, Ohta Y, Rabinovitz I et al. (2004) Ribosomal S6 kinase (RSK) regulates phosphorylation of filamin A on an important regulatory site. Molecular and cellular biology 24:3025-3035
59. Yao L, Fan P, Jiang Z et al. (2008) Dopamine and ethanol cause translocation of epsilonPKC associated with epsilonRACK: cross-talk between cAMP-dependent protein kinase A and protein kinase C signaling pathways. Molecular pharmacology 73:1105-1112
60. Zajchowski DA, Bartholdi MF, Gong Y et al. (2001) Identification of gene expression profiles that predict the aggressive behavior of breast cancer cells. Cancer Res 61:5168-5178

Fig. 1 High baseline L-plastin Ser5 phosphorylation in highly invasive versus low baseline L-plastin Ser5 phosphorylation in non- or weakly invasive breast cancer cell lines.



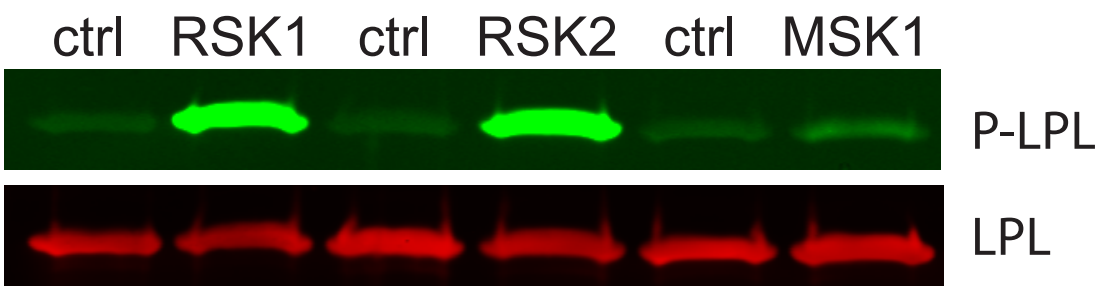
High baseline L-plastin Ser5 phosphorylation in highly invasive versus low baseline L-plastin Ser5 phosphorylation in non- or weakly invasive breast cancer cell lines. a Four breast cancer cell lines were seeded in collagen I-coated (200 µg/ml) 96-well plates (Essen Imagelock, Essen Bioscience). At appoximately 90% confluence, a wound was scratched across each well with the Cellplayer 96-well woundmaker (Essen Bioscience). Cells were covered with collagen I (1.5 mg/ml) diluted in growth medium. Wound confluence was monitored with the Incucyte LiveCell Imaging System (Essen Bioscience) by measuring cell confluence every 3 hours over a total period of 72 hours. The graphs depict means +/- SEM from three biological replicates with at least 5 technical replicates. b Cell extracts were subjected to immunoblot analysis using antibodies specific for Ser5 phosphorylated L-plastin (anti-Ser5-P antibody, P-LPL) and total L-plastin (LPL) and signals were analysed using the Odyssey infrared image system (Licor,Westburg).

Fig. 2 MAPK pathways among the top canonical pathways for the 3 comparisons of interest.



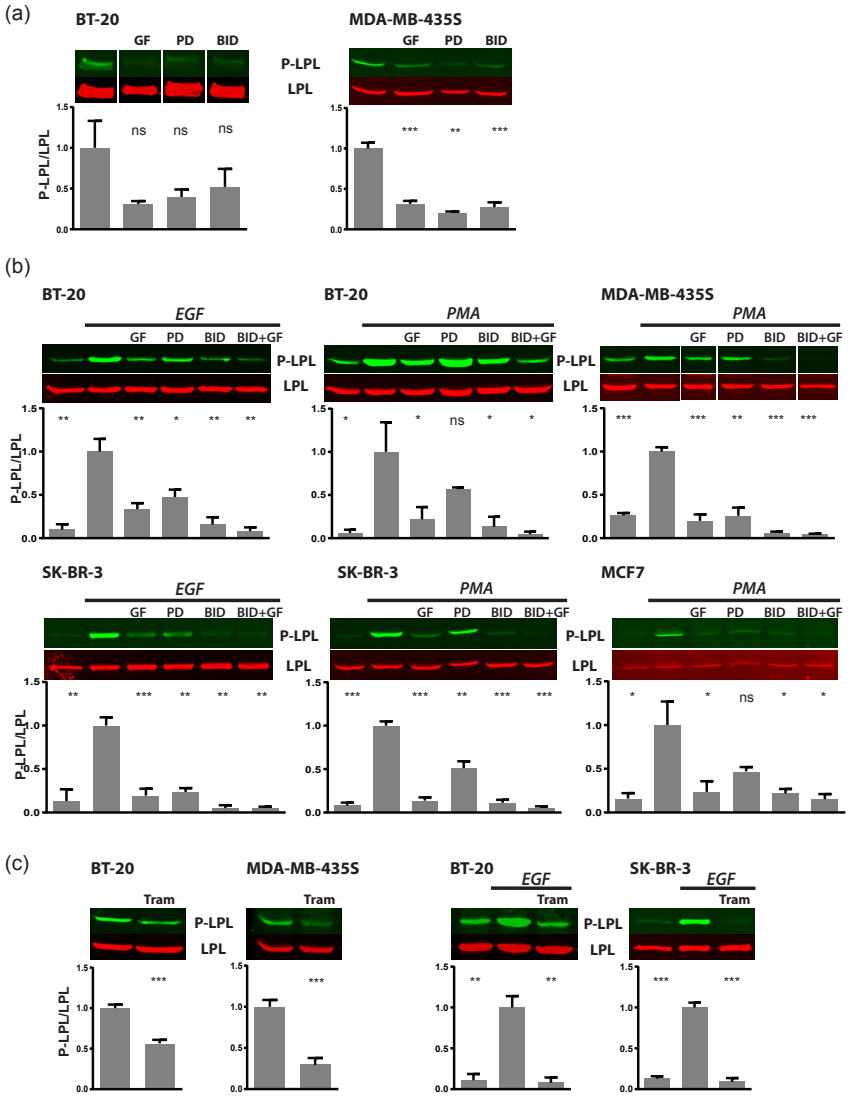
MAPK pathways among the top canonical pathways for the 3 comparisons of interest. Whole genome microarray analysis revealed significantly differentially expressed transcript clusters between the indicated conditions. The lists 1-3 were mapped and analysed by Ingenuity Pathway Analysis which revealed 3 canonical pathways that were ranked over the treshold of -log(p-val) ≥ 1.3010 common to the 3 comparisons.

Fig. 3 RSK1 and RSK2 are able to directly phosphorylate L-plastin Ser5 *in vitro*.



RSK1 and RSK2 are able to directly phosphorylate L-plastin Ser5 in vitro. 10 µg of recombinant full-length L-plastin were incubated with 100 ng of recombinant kinase and with 50 µM of ATP in a reaction volume of 25 µl according to the manufacturer's protocol. Samples were incubated for 15 minutes at 30°C. Following incubation, Laemmli buffer was added, the samples were boiled at 100°C for 5 minutes and analysed by immunoblot visualising Ser5 phosphorylated L-plastin (P-LPL) and total L-plastin (LPL).

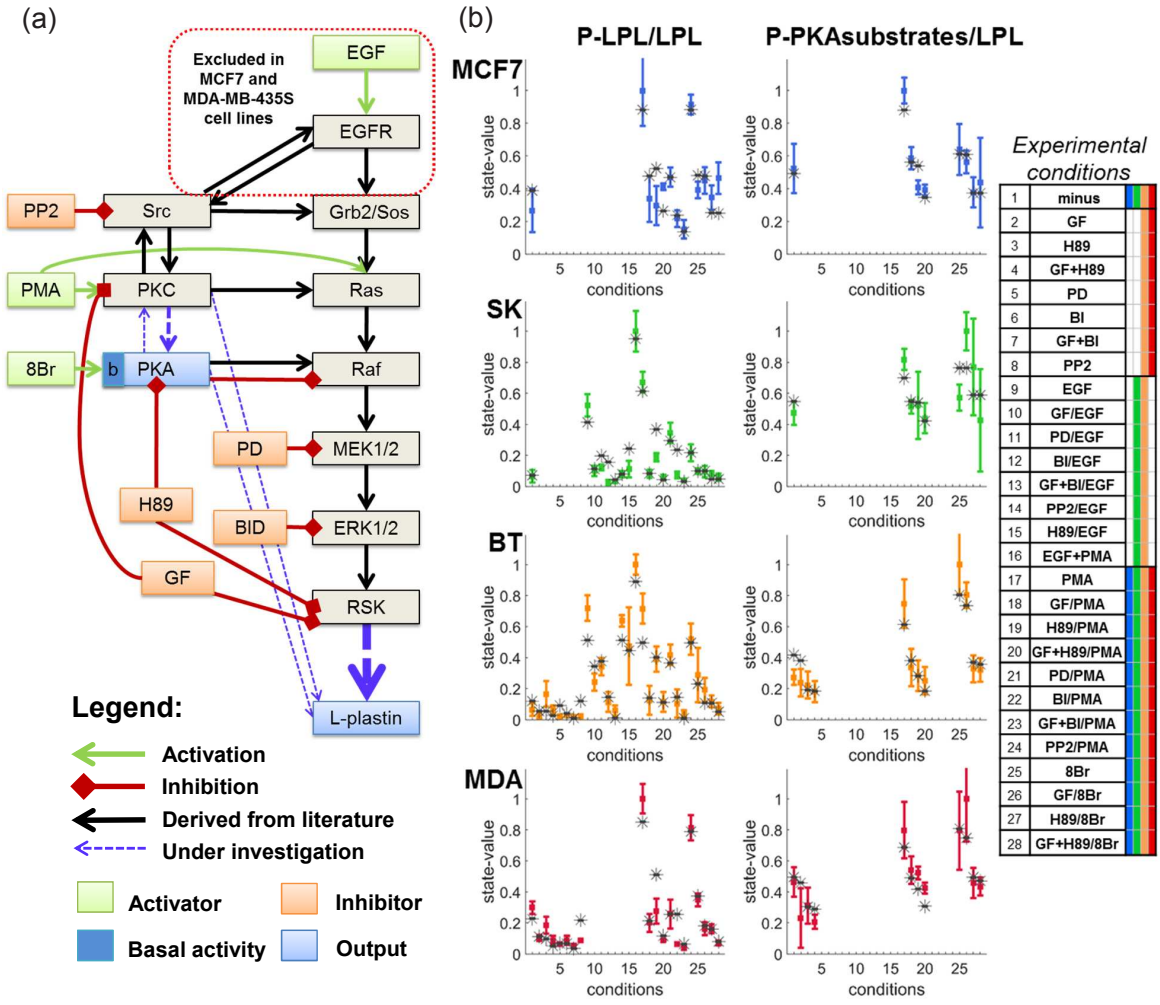
Fig. 4 Involvement of the ERK/MAPK pathway in L-plastin Ser5 phosphorylation in breast cancer cells.



Involvement of the ERK/MAPK pathway in L-plastin Ser5 phosphorylation in breast cancer cells. (a) Invasive cell lines exhibiting high baseline L-plastin Ser5 phosphorylation were treated with inhibitiors of selected signalling molecules of the ERK/MAPK pathway GF109203X (GF), PD98059 (PD) and BI-D1870 (BID). Subsequent to inhibitor treatment, residual L-plastin Ser5 phosphorylation was determined by immunoblot analysis. For quantification, the ratio between the intensities obtained for phosphorylated L-plastin (P-LPL) versus total L-plastin (LPL) was determined and then normalised to the mean of all the values obtained in one experiment to account for experimental variability. After scaling to the highest signal, graphical representations show means and standard deviations. Statistical significance was determined by an unpaired T-test with Welch's correction. *P<0.05, **P<0.01 and ***P<0.001 (b) Cells were preincubated with inhibitors of the ERK/MAPK pathway and then stimulated with EGF (for the two EGFR expressing cell lines) or PMA (for the four cell lines). L-plastin Ser5 phosphorylation was quantified as described under (a). (c) Cells were treated with Trametinib with or without subsequent EGF stimulation and L-plastin Ser5 phosphorylation was quantified as described under (a). For each immunoblot shown, samples were run on the same gel, but for BT-20 shown under (a) and MDA-MB-435S shown under (b), bands were cut and put in another order for representative purposes.

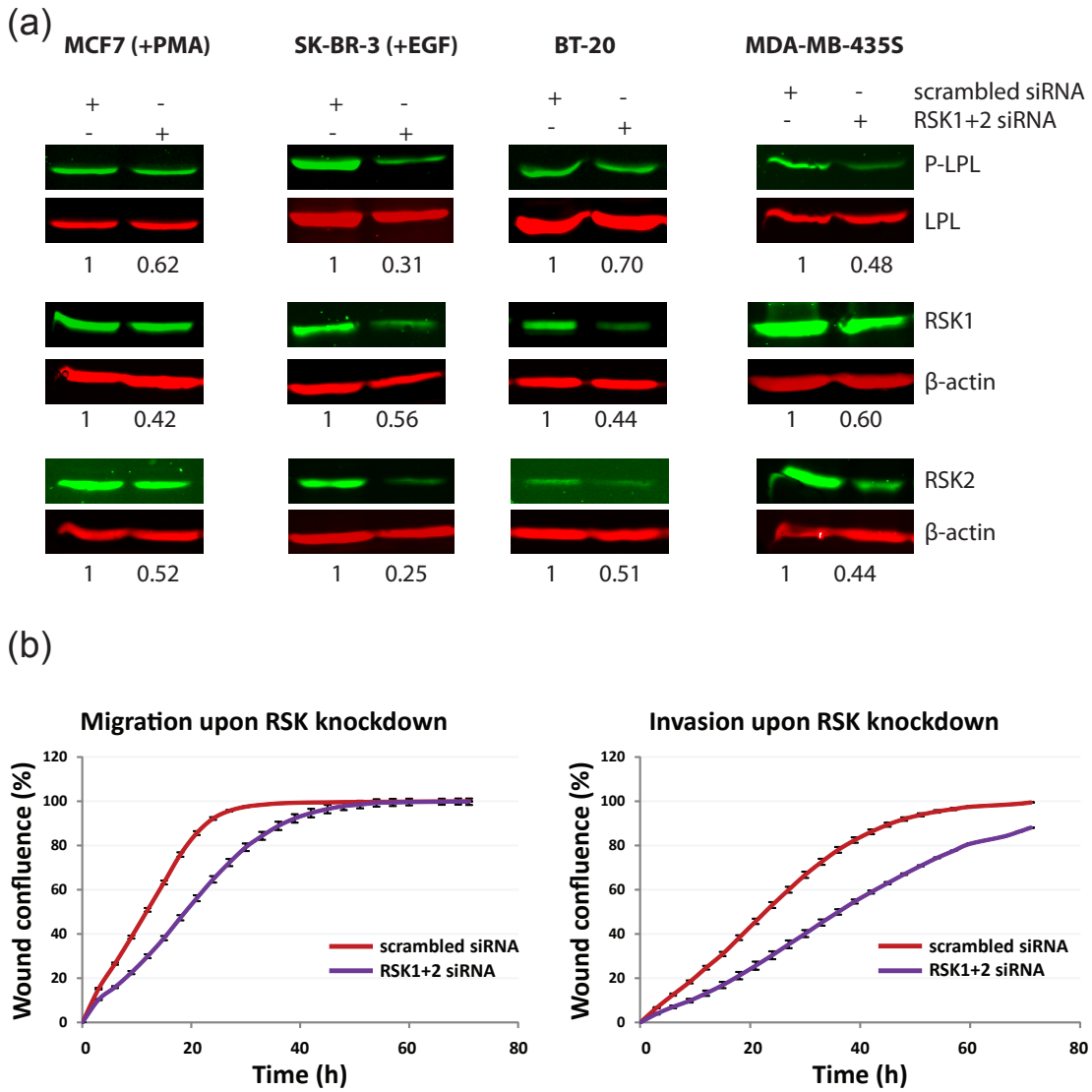
**** Figure 5 is prepared by P.Trairatphisan**

Fig. 5 Literature-derived and experiment-based L-plastin signalling network and model fitting results.



Literature-derived and experiment-based L-plastin signalling network and model fitting results. (a) A candidate network for the signalling pathways upstream of L-plastin phosphorylation was built based on literature information and own findings. The network interactions were analysed by applying a probabilistic Boolean network approach taking into account cell line specific immunoblot-based quantifications of phosphorylated L-plastin and phosphorylated PKA substrates. Various conditions were tested in the four cell lines MCF7, SK-BR-3 (SK), BT-20 (BT), MDA-MB-435S (MDA) including activation by EGF, PMA or 8-Bromo-cAMP (8Br) and/or inhibition by GF109203X(GF), H89, PP2, PD98059 (PD) or BI-D1870 (BID). (b) PBN model fitting results in comparison to experimental data for the various tested conditions. Coloured bars indicate the conditions tested for the individual cell lines. For quantification, the ratio between the intensities obtained for phosphorylated L-plastin (P-LPL) respectively phosphorylated PKA substrates (P-PKAsubstrates) versus total L-plastin (LPL) was determined and normalised to the mean of all the values obtained in one experiment to account for experimental variability. Data generated from different experimental sets were normalised to the calibrator PMA, subsequently pooled and scaled to the highest signal (of each respective graph). Means of ten simulated values from the PBN model (black stars) were compared against the experimental data (multi-coloured squares [mean] and error bars [SD]).

Fig. 6 Combined RSK1 and RSK2 knockdown decreases L-plastin Ser5 phosphorylation and impairs migration and invasion in breast cancer cells.



Combined RSK1 and RSK2 knockdown decreases L-plastin Ser5 phosphorylation and impairs migration and invasion in breast cancer cells. A siRNA against RSK1 and RSK2 or scrambled siRNA were transfected in the four cancer cell lines as indicated and L-plastin Ser5 phosphorylation, RSK1 and RSK2 expression were assessed by immunoblot analysis 72 hours post transfection. Non-invasive cell lines were treated with EGF or PMA as indicated. b MDA-MB-435S cells were seeded in collagen I-coated (200 µg/ml) 96-well plates (Essen Imagelock, Essen Bioscience) and transfected with siRNA against RSK1 and RSK2 or with scrambled siRNA. After 24 hours, a wound was scratched across each well with the Cellplayer 96-well woundmaker (Essen Bioscience). To study invasion, cells were covered with collagen I (1.5 mg/ml) diluted in cell culture medium. To study migration, cell culture medium was added to the cells. Migration and invasion were monitored by measuring wound confluence every 3 hours for a total of 72 hours with the Incucyte LiveCell Imaging System (Essen Bioscience). The graph depicts means +/- SEM from three biological replicates each including at least 11 technical replicates. Efficient knockdown of RSK1 and RSK2 as well as efficient decrease of L-plastin Ser5 phosphorylation were confirmed by immunoblot analysis (data no shown).

4.4. Potential applications of PBNs in pharmaceutical industry

4.4.1 Detailed mechanistic models and their applications in pharmaceutical industry

In the drug discovery process, various strategies are applied to identify novel therapeutic targets including target-based screening, phenotypic screening, modification of natural substances and biologic-based approaches (Swinney & Anthony, 2011). In recent years, pharmaceutical industry started to apply computational approaches in systems biology to aid the drug discovery process via performing model-based data integration approaches to identify and explore the molecular mechanism of action of therapeutic agents (Iskar, Zeller, Zhao, van Noort, & Bork, 2012). Merrimack Pharmaceuticals (Boston, MA, USA, http://merrimackpharma.com) is one of the leading pharmaceutical companies which applies modelling approaches in systems biology known as *Network Biology* for the discovery and assessment of therapeutic agents which are tailored for the treatments on different types of cancers.

In this sub-chapter, I present a modelling study which I performed during my internship at Merrimack to demonstrate how quantitative mechanistic modelling can be applied to analyse the properties of a therapeutic agent in a pharmaceutical setting. Here, we focused on the study of c-MET signalling which is one of the important signalling pathways that drives tumour growth, proliferation and metastasis in many cancer types. To get a better understanding on the dynamics and the other properties such as the sensitivity of molecules in this signalling pathway, we built a mechanistic ODE-based model of c-MET signalling based on a literature-derived model topology. Then, we performed an optimisation against a training dataset which comprises the read-out from c-MET and downstream signalling molecules in response to hepatocyte growth factor (HGF) stimulation from 0.037nM to 9nM over the course of 2 hours on the ACHN (renal cancer) cell line. Once we obtained the model with optimised parameters that fits well to the data, the model was subsequently used to assess the effectiveness of a therapeutic agent.

To start building a dynamical model of c-MET signalling, we derived a set of ODEs based on the network topology as proposed in literature (Schoeberl et al., 2009, 2002). The network topology of the receptor level and the downstream signalling pathways is shown in Figure 3. The c-MET signalling model consists of 109 species, 141 parameters and 104 reactions which include detailed interactions and regulations of c-MET receptor

(MET) and downstream intracellular signalling pathways, i.e. MAPK and PI3K/AKT/mTOR pathways with their crosstalks. Apart from fitting the model to experimental data, prior knowledge from literature such as the HGF-cMET affinity (Kd), the fraction of internalised c-MET to total c-MET, the fraction of degraded c-MET to total c-MET and the minimal ratio of phosphorylated c-MET were integrated as additional soft constraints. The optimisation was performed by applying a modified version of particle swarm algorithms *(confidential)* to identify the corresponding parameter values. The plotted results from the optimisation in time-course and dose-response representations are shown in Figure 4 and Figure 5.

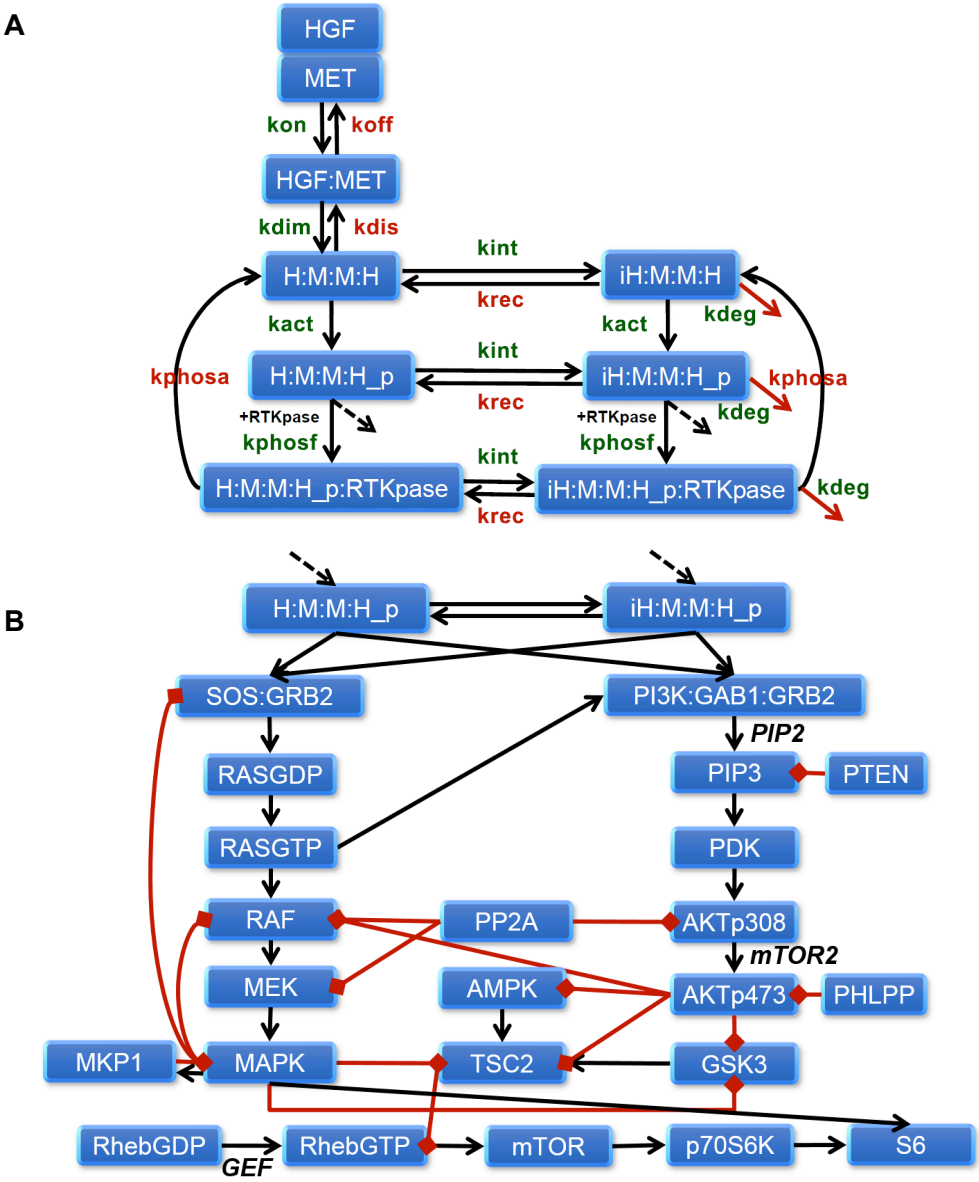


Figure 3. Network topology of the literature-derived c-MET signalling network. A) upstream signalling pathway at the c-MET receptor level with parameter names. B) downstream signalling pathways including MAPK and PI3K/AKT/mTOR pathways with their crosstalk interactions. The dashed arrows in A and B link the upstream signalling pathway to the downstream signalling pathway.

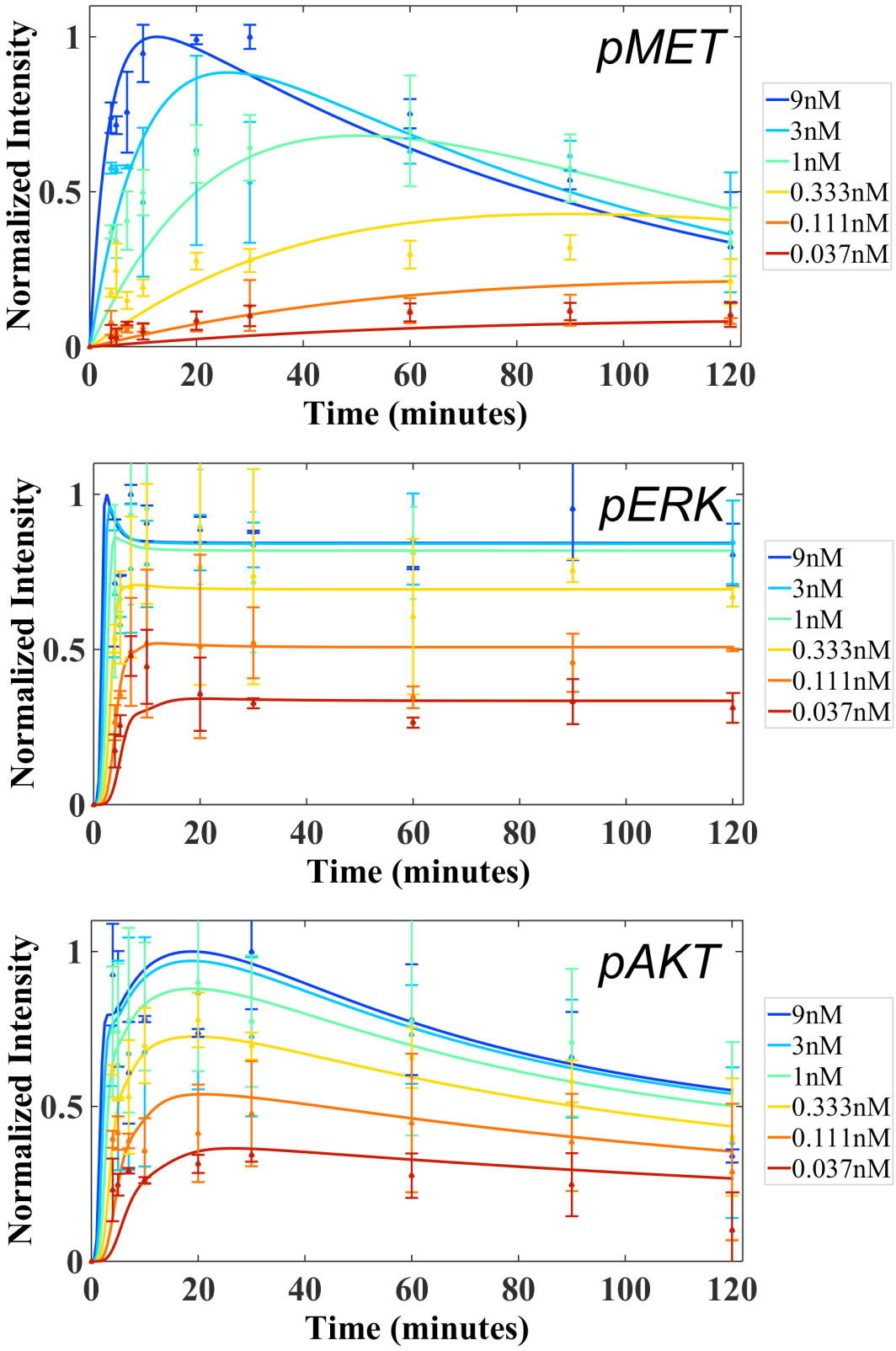


Figure 4. Plotted optimisation results of c-MET signalling model after HGF stimulation at concentrations from 0.037nM to 9nM over the course of two hours in time-course representation. Error-bars depict experimental data and solid-lines depict model simulations.

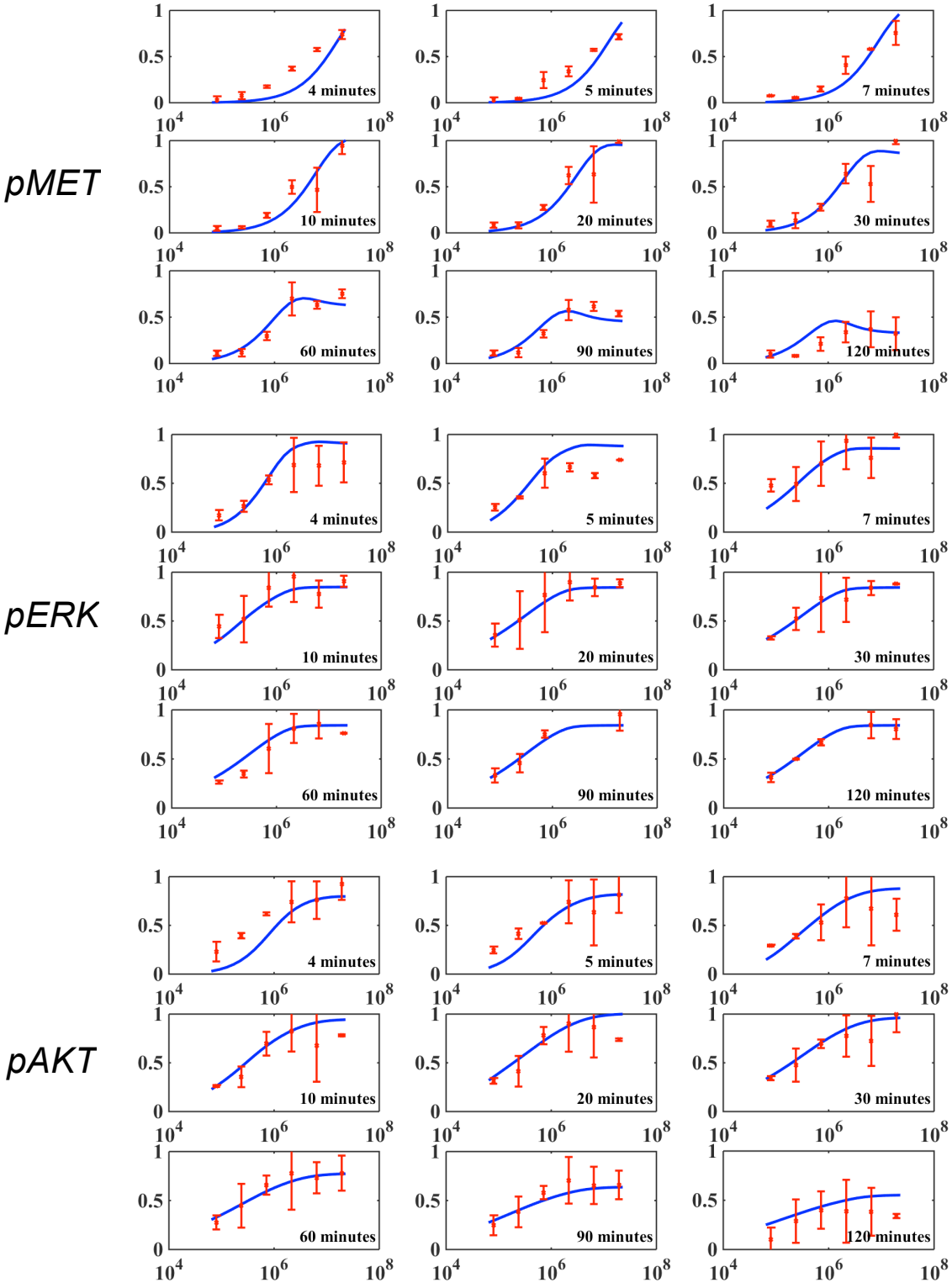


Figure 5. Plotted optimisation results of c-MET signalling model after HGF stimulation at concentrations from 0.037nM to 9nM over the course of two hours in dose-response representation. Error-bars depict experimental data and solid-lines depict model simulations.

As shown in Figure 4 and Figure 5, we identified a set of parameters that fit well to the training set of time-course experimental data which comprises phosphorylated c-MET (pMET), phosphorylated ERK1,2 (pERK) and phosphorylated AKT (pAKT) at different HGF concentrations. In addition, we checked that the prior knowledge from literature in the form of soft constraints were all fulfilled. For instance, after stimulating the cells with 9nM HGF, the fraction of internalised c-MET to total c-MET is predicted to 17.05% after 2 hours which is in a good agreement with literature which suggested a fraction between 15-30%. Also, we found that the predicted fraction of degraded internalised c-MET to total c-MET is 68.22% in the same experimental setting while literature suggest a fraction of 70-80% (see visualised comparisons in Figure 6). Hence, we are ensured that we obtained a good dynamical model of c-MET signalling with integrated prior knowledge where it can be further used for the discovery of therapeutic target, e.g. by performing a sensitivity analysis, and for designing the optimal properties of therapeutic targets by performing *in silico* simulations (Schoeberl et al., 2009).

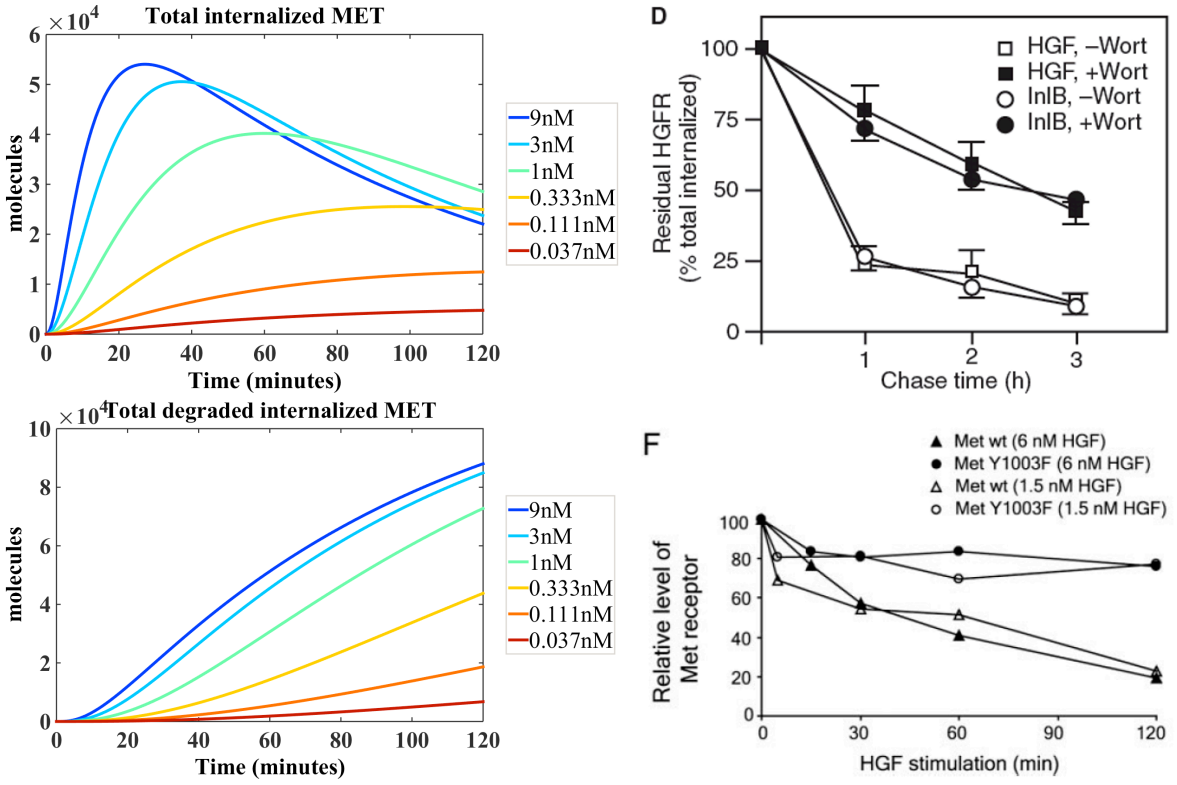


Figure 6. Internalisation and degradation rates of c-MET from model simulation compared against prior knowledge in literature. Note: The total number of c-MET receptor in the model is 129,000 molecules per cell. Upper row: total number of internalised c-MET (N. Li, Xiang, Dokainish, Ireton, & Elferink, 2005), Lower row: total number of degraded internalised c-MET (Abella et al., 2005).

In the next step, we investigated how a detailed quantitative ODE-based model can be applied to analyse the efficiency of a therapeutic agent. For the treatment of cancers that arose from the deregulation of c-MET receptor and its ligand, multiple strategies could be applied, see (Cecchi, Rabe, & Bottaro, 2010). One of the treatment options is to use a target-specific antibody against c-MET to block the binding between receptors and ligands. This mechanism of action prevents the auto-phosphorylation of the receptor which in turn reduces (or eliminates) the signals towards downstream signalling molecules. In this demonstrative example, we introduced *OA-5D5*, a one-armed humanized IgG1 monoclonal antibody against c-MET receptor (formerly under development by Roche/Genentech as *MetMAb*) (Merchant et al., 2013) as a subject of study to analyse how the properties of this antibody could affect the change of signalling patterns in downstream signalling pathways by applying the optimised c-MET signalling model that we previously built.

First, we examined if the optimised model of c-Met signalling is predictive for *OA-5D5* to ensure that the model will be compatible for the analysis of this antibody. To investigate this, we performed an experiment by pre-incubating *OA-5D5* at different concentrations from 1pM to 1µM for 2 hours on 3 human cancer cell lines which are A549 (adenocarcinomic alveolar epithelial cells), H2170 (squamous cell carcinoma of lung) and SW-620 (metastasised colorectal adenocarcinoma). Then, the cells were treated with HGF at 1nM for 10 minutes and pAKT was measured as a read-out. On the modelling part, the inhibitory mechanism and measured kinetic parameters of *OA-5D5* were incorporated into the optimised c-MET signalling model. Subsequently, a target engagement prediction was performed by simulating the model in the experimental conditions as performed in wet-lab experiment and compare it to the experimental data. The results of model prediction are shown in Figure 7.

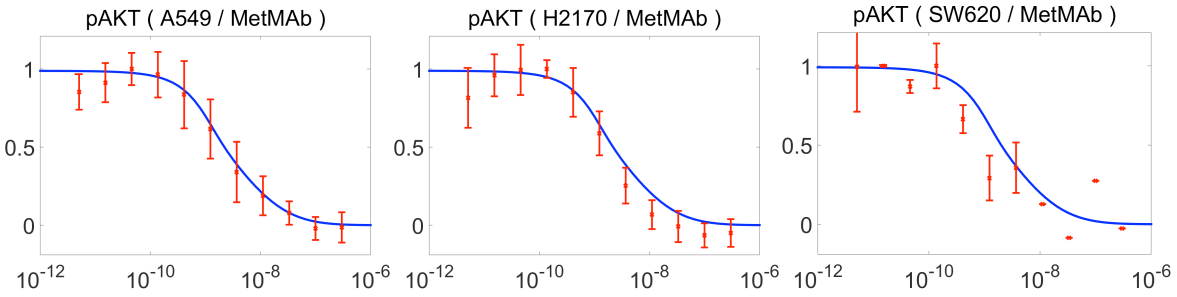


Figure 7: Model predictions of OA-5D5 (MetMAb) inhibitory effects on phosphorylated AKT (pAKT) at different concentrations in A549, H2170 and SW620 cell lines (from left to right). Red error bars depict the experimental data and blue lines depict model simulations.

As shown in Figure 7, the simulated results matched very well to the experimental data. These results confirmed that the integrated c-Met signalling model with *OA-5D5* inhibitory mechanism can accurately predict the changes of downstream signalling molecules. Therefore, this predictive c-MET model could then be further used to perform *in silico* predictions in order to investigate biological phenomena once the systems properties or the experimental conditions are altered. To demonstrate an example, we performed *in silico* simulations in two experimental conditions which are 1) Pre-incubate *OA-5D5* (1pM-1μM) for 120 minutes and stimulate with 1nM of HGF for 10 minutes, and 2) Co-incubate *OA-5D5* (1pM-1μM) together with 1nM of HGF for 120 minutes, with the levels of c-MET varied from 10,000 to 1,000,000 molecules per cell. The predicted results on 3 signalling molecules including pMET, pERK1,2 and pAKT are shown in Figure 8.

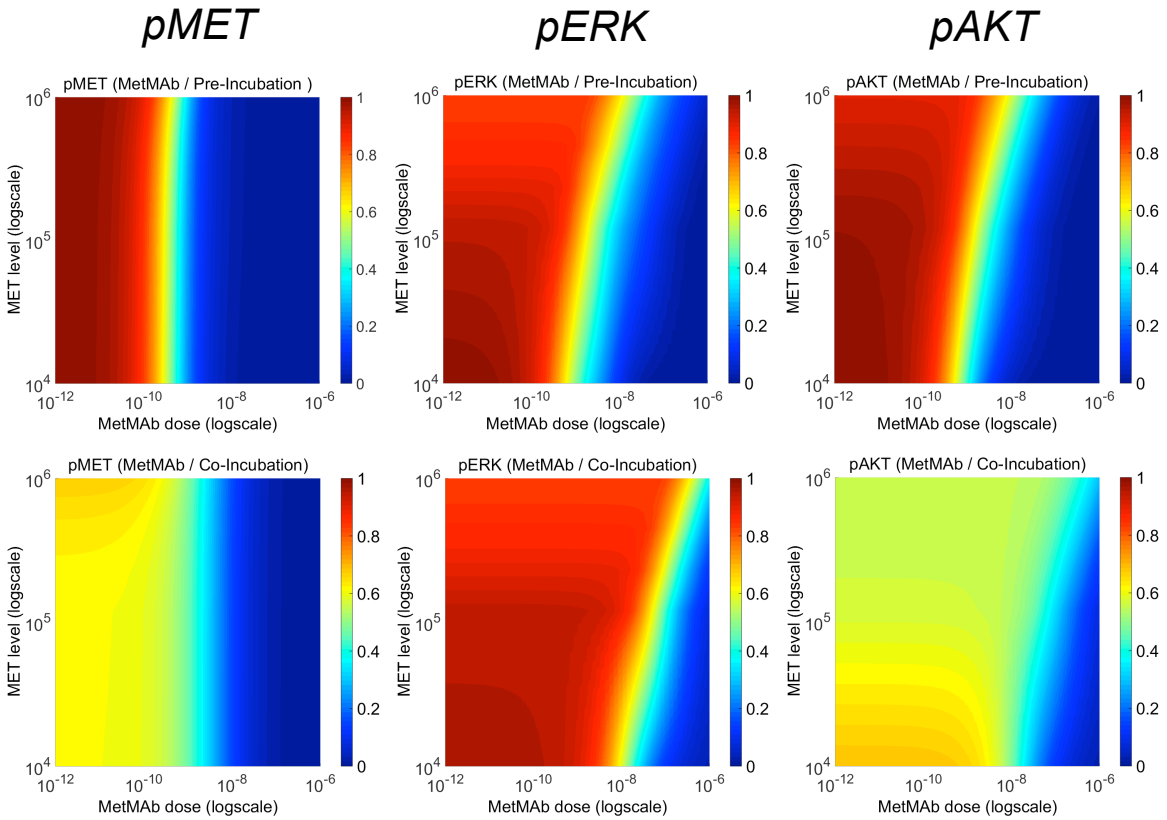


Figure 8. *In silico* predictions of molecular activities in the c-MET signalling pathway when c-MET is varied from 10^4 to 10^6 molecules per cell. The predictions were performed by pre-incubating OA-5D5 (MetMAb) for 2 hours followed by 1nM HGF stimulation for 10 minutes (top row) or by co-incubation of OA-5D5 (MetMAb) and 1nM HGF (bottom row) for 2 hours.

As demonstrated in Figure 8, the inhibitory effects of *OA-5D5* are relatively different comparing the two experimental conditions. Low concentration of *OA-5D5* at 1nM can totally inhibit the pMET signal when having pre-incubated the cells with antibodies while up to 50% of pMET were still active at this concentration once *OA-5D5* and HGF

were co-incubated. In contrary, pMET and pAKT signals could be decreased up to 40% once HGF was co-incubated with *OA-5D5* at a very low concentration, e.g. at 1pM, while there was no inhibitory effect if the antibody was pre-incubated at this concentration. These results demonstrate that the detailed mechanistic model of c-MET signalling can serve as a tool to explore the effectiveness of therapeutic targets in various experimental conditions which in turn facilitates future experimental planning to obtain desirable outcomes.

From this modelling study, there is one important remark that should be mentioned, i.e. the set of kinetic parameters which were used to assess the effectiveness of *OA-5D5* was identified from only one round of model optimisation. In a practical way, multiple rounds of optimisation should be performed and a large set of kinetic parameter values from the fitted models should be analysed for their distributions. These values will indicate whether the obtained parameter values are identifiable. In the presenting case, the optimisation has to be performed on an in-house computational cluster due to the large size and the high complexity of the model. However, the cluster is a shared resource for a group of modellers who work on different projects. Therefore, only a few rounds of optimisation were performed on the c-MET signalling model according to the availability of the cluster. In parallel, *in silico* simulations to assess the effectiveness of *OA-5D5* (Figure 8) which were performed on a local standard computer also took a few hours per condition which is considered to be impractical. To summarise, the large models' size, the high complexity of the models, and the limited computational resources, are the main limiting factors which impede the advancement of modelling work in a pharmaceutical setting that need to be resolved.

Apart from the application of a quantitative mechanistic model on analysing the properties of therapeutic targets as demonstrated, different types of quantitative models can also be applied in a pharmaceutical setting. For instance, the bioavailability of drugs can be analysed in pharmacokinetic/pharmacodynamics (PK/PD) models (Roy et al., 2013). Also, partial differential equation (PDE)-based models can be applied to predict the changes of drug's concentration over time and over the distance from tumour cores (Graff & Wittrup, 2003). These examples therefore additionally highlighted the broad applications of quantitative models in systems biology in a pharmaceutical setting.

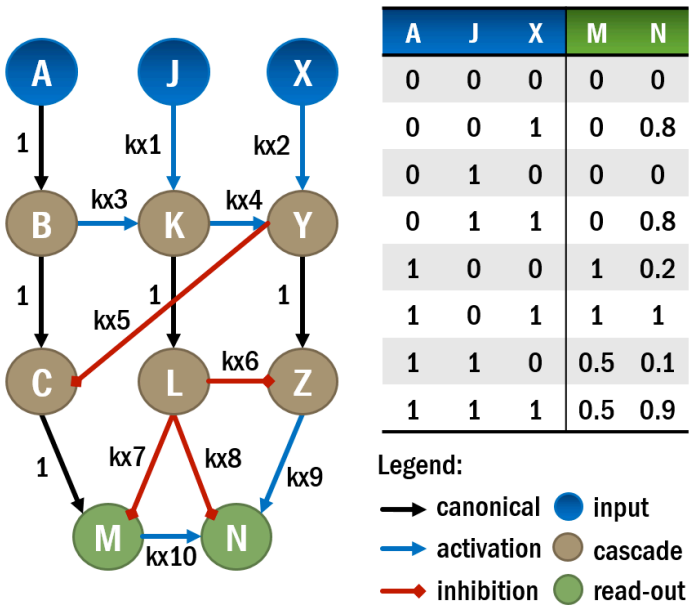
4.4.2 A PBN-based approach for network pre-selection

As demonstrated in Chapter 4.4.1, detailed mechanistic ordinary differential equation (ODE)-based and partial differential equation (PDE)-based models are mainly applied for the discovery and assessment of therapeutic targets in a pharmaceutical setting. Nevertheless, the basic ground of these advanced mathematical models still relied on the model of cellular signal transduction network which is a pre-requisite element for understanding the signal flow at the molecular level. In general, the required computational resources and optimisation time for ODE-based models of signalling networks are considerably high as mentioned in Chapter 4.4.1. One of the major causes for such high computational time is the large size of networks. The number and complexity of ODEs normally grows according to the number of included molecules and interactions which are derived from evidence in literature. However, these interactions are typically not exclusive for the specific experimental settings of interested, e.g. for specific cell lines or for specific intracellular signalling pathways downstream of certain receptor tyrosine kinases. To select a suitable network topology with a "brute-force" approach by examining all combinations of potential interactions in the ODE-based framework will take unbearable time to perform. This issue therefore calls for a fast and reliable approach to pre-select the suitable network topology prior to building detailed mechanistic models which could in turn boost the speed of pharmaceutical discovery and development.

Based on the performance of the PBN framework on the study of signal transduction networks as shown in Chapter 4.1, 4.2 and 4.3, we foresee that PBN could also be used to pre-select a suitable network topology based on the experimental data at steady-state. By analysing the selection probabilities of an optimised PBN model, one could determine if certain reactions can be removed. The closer the value gets close to zero, the more likely the corresponding interaction does not influence the signal flow in the network. Following this concept, one could perform an iterative optimisation pipeline to successively remove unnecessary interactions until only core interactions which are necessary to explain experimental data remain in the model.

To explore if the PBN approach can actually be applied for pre-selecting a suitable model topology before starting with detailed mechanistic modelling of a network. We built a small example model comprising 11 nodes and 15 edges with artificial steady-state measurement data as shown in Figure 9. Among the 15 edges, 5 are assigned as canonical interactions, i.e. core interactions, while the other 10 interactions (with the prefix 'kx') remained uncertain if they are necessary to be included. Then, we applied *optPBN* to

perform the optimisation and it shows that only five out of ten uncertain interactions have optimised weights which are not close to zero (Table 1). These interactions with the corresponding weights are essential to be integrated into the model in order to explain the measurement data. The other five interactions with identified weights close to zero could be excluded from the model without any impact on the quality of model fitting. The low standard deviations from multiple parameter sets (best 500 sets were used) also ensured that the optimised parameters are clustered together. The contextualised model topology by *optPBN* with optimised weights is shown in Figure 10.



A	J	X	M	N
0	0	0	0	0
0	0	1	0	0.8
0	1	0	0	0
0	1	1	0	0.8
1	0	0	1	0.2
1	0	1	1	1
1	1	0	0.5	0.1
1	1	1	0.5	0.9

Figure 9: The network topology of a small example model with steady-state measurements.

Table 1: Fitting cost, optimisation time and parameter distributions generated by *optPBN*. The interactions with optimised weights close to zero are shown in red, otherwise, in green.

Mean fitting costs	0.0062	
Optimisation time	Stand-alone (1 core): 4051 seconds Grid'5000 (40 cores): 166 seconds	
Parameters	**Mean**	**SD**
kx1	1.000	0.002
kx2	0.996	0.012
kx3	0.000	0.002
kx4	0.004	0.012
kx5	0.000	0.000
kx6	0.000	0.004
kx7	0.499	0.005
kx8	0.016	0.028
kx9	0.791	0.031
kx10	0.193	0.037

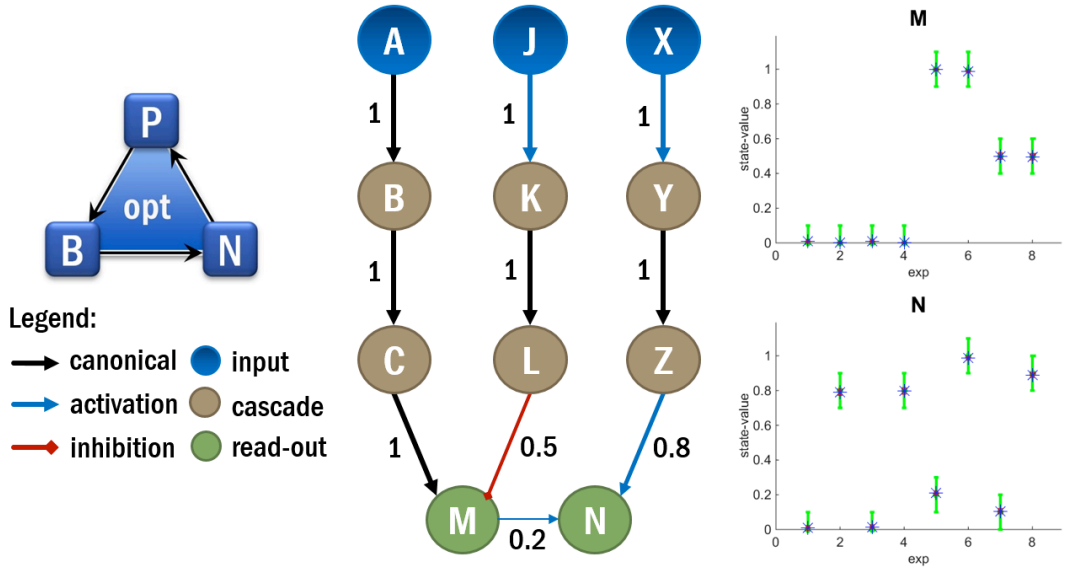


Figure 10: [Left] The contextualised network topology of the small example model with optimised weights as identified by *optPBN*. The sizes of arrows are adjusted according to the optimised weights. [Right] Plotted simulation results. The experimental data are presented as green error bars and the simulated values of the optimised model are shown as blue stars with red arrow bars.

Next, we evaluate if the core network topology pre-selected by *optPBN* can be used for the building of a detailed mechanistic ODE-based model that fits to the measurement data. Based on the initial network topology with only 5 canonical (core) interactions, we added combinations of 0 to 10 interactions picked from the pool of uncertain 'kx' interactions and applied the law of mass action to derive a total of 1024 ODE-based model variants. Then, we performed the optimisation of these model variants applying the Systems Biology Toolbox 2, *SBtoolbox2* (Schmidt & Jirstrand, 2006). The results from the optimisations in terms of fitting costs and computational time are shown in Figure 11A and 11B, respectively.

In Figure 11A, it is shown that the fitting costs vary in the range of 0 to 0.6. Among them, the initial model variant which contains only canonical interactions without any additional uncertain 'kx' interaction (depicted as Pick0) could not fit to the data and has the highest fitting cost. In parallel, adding more uncertain 'kx' interactions into the initial model variant (depicted as Pick1, Pick2, …, depending on the number of added interactions) improved the fitting costs as some essential interactions are added to the model. Interestingly, we found that the model variant with 5 uncertain 'kx' interactions (Pick5) which exactly overlaps to the list of essential interactions as identified by *optPBN* (see Table 1 and Figure 10) has the lowest fitting cost. Hence, this result demonstrates that the contextualised network topology identified by *optPBN* matches the minimal topology of the ODE-based model that fits well to the same set of measurement data.

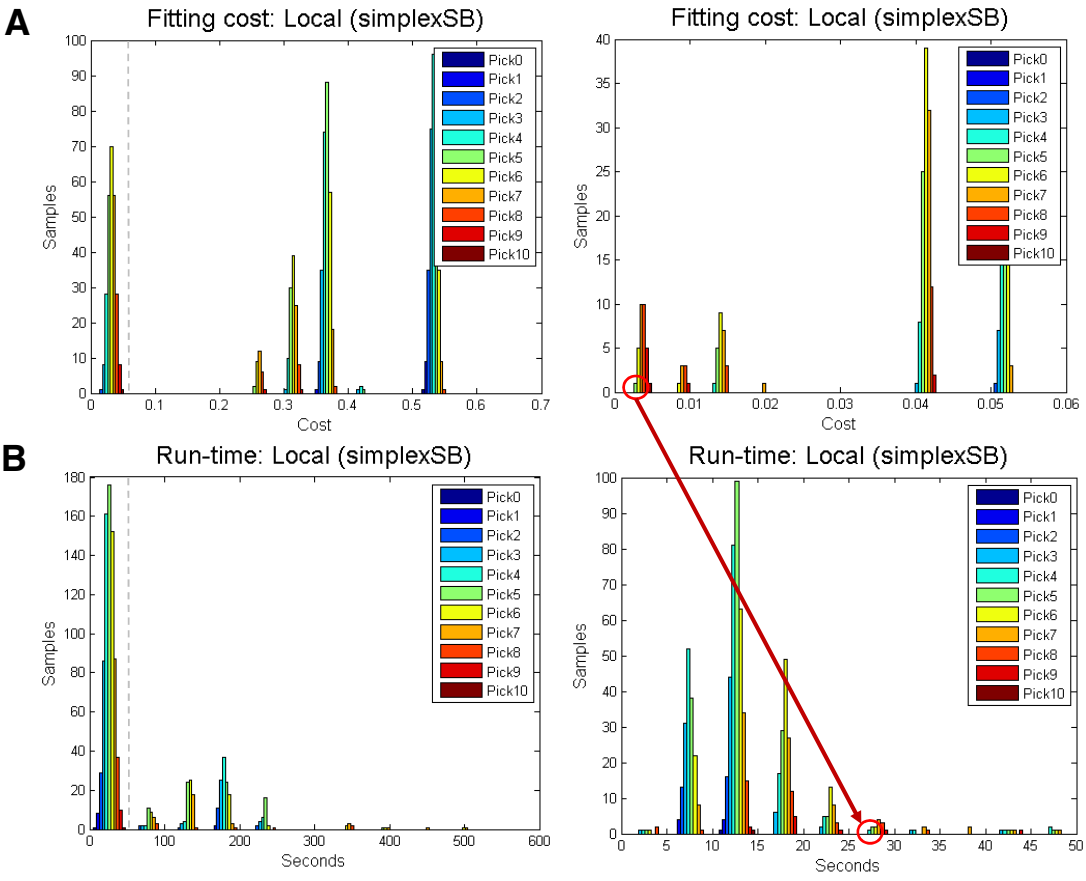


Figure 11. Fitting costs and computational time of the 1024 ODE-based model variants. A) Fitting cost (left: overview, right: zoomed-in with cost <0.06 [left part of the dotted line]). B) Computational time (left: overview, right: zoomed-in with time <50 seconds [left part of the dotted line]). The red circles highlight the fitting cost and computational time of the ODE-based model variant with the same network topology as the one contextualised by *optPBN*.

In terms of computational time as shown in Figure 11B, we observed that the model variants required from a few seconds up to 500 seconds to perform parameter identifications in the ODE framework. We also found that the model variant with 5 selected interactions which overlaps with the *optPBN* results returned the best fitting cost in just only 27 seconds while the total optimisation time for all model variants is 57840 seconds (about 16 hours). This finding demonstrates that the network pre-selection by *optPBN* could guide modellers to directly identify the essential set of reactions for building a detailed mechanism model in a reasonable timeframe. Such pipeline might be able to replace the trail-and-error selection of uncertain interactions in the ODE framework which requires much higher computational time once steady-state data is available. This will be further explored in follow-up studies.

Chapter 5
DISCUSSION AND PERSPECTIVES

Molecular signatures of deregulated signal transduction networks at steady-state reflect sustained aberrant cellular phenotypes in diseases.

One of the principal goals of every living organism is to survive and to maintain its well-being. After exposures to internal and external changes in the cellular environment, e.g. changes in temperature, in nutritional sources, or in genetic integrity, the corresponding regulatory mechanisms have to operate in an efficient and effective manner in order to maintain homeostasis or 'balanced state' of the cell. Failures of these regulations range from systemic to molecular levels and often lead to various types of diseases, such as neurodegenerative diseases, metabolic diseases, and cancers.

Upon perturbations, 'signals' are transduced via intracellular signalling cascades towards multiple cellular components in order to adapt cellular phenotypes corresponding to the type and intensity of perturbators. At the molecular level, cells receive signals through the binding of ligands to signalling receptors which in turn transduce, integrate, modulate, and spread signals towards successive functional units including genes, mRNA, proteins, and metabolites. Even if there exist various kinds of cellular processes downstream of the signal transduction networks such as gene transcription, mRNA translation, or enzymatic conversion of metabolites, these processes as well as their regulatory mechanisms can also be viewed as a single integrated signalling network that transfers signals to communicate and regulate multiple cellular components in an orchestrated fashion in order to achieve the same ultimate goal, to maintain cellular integrity and homeostasis (Gonçalves et al., 2013). The schematic of interconnections and regulations among cellular components is depicted in Figure 12.

On pathological perturbations in cancer biology, molecular perturbations which lead to carcinogenic transformation are usually under a tight control by multiple types of cellular regulation, e.g. DNA excision and repair, cell cycle arrest or senescence, as well as apoptosis. Generally, a short-term perturbation, e.g. an acquisition of a thymine-dimer after strong UV light exposure will be rapidly corrected by the photolyase or excision enzyme with no observable change in the phenotype (Kao, Saxena, Wang, Sancar, & Zhong, 2005). However, certain types of perturbations such as the acquisition of the D842V point mutation on PDGFRα does not have specified repairing mechanism and

could eventually lead to a permanent change in cellular responses, e.g. the over-proliferation and increased survival of mesenchymal cells as found in GISTs. The signatures of these permanents changes can be detected at the molecular level such as the changes of gene expression or an increase in the pool of active signalling proteins and/or enzymes in metabolism.

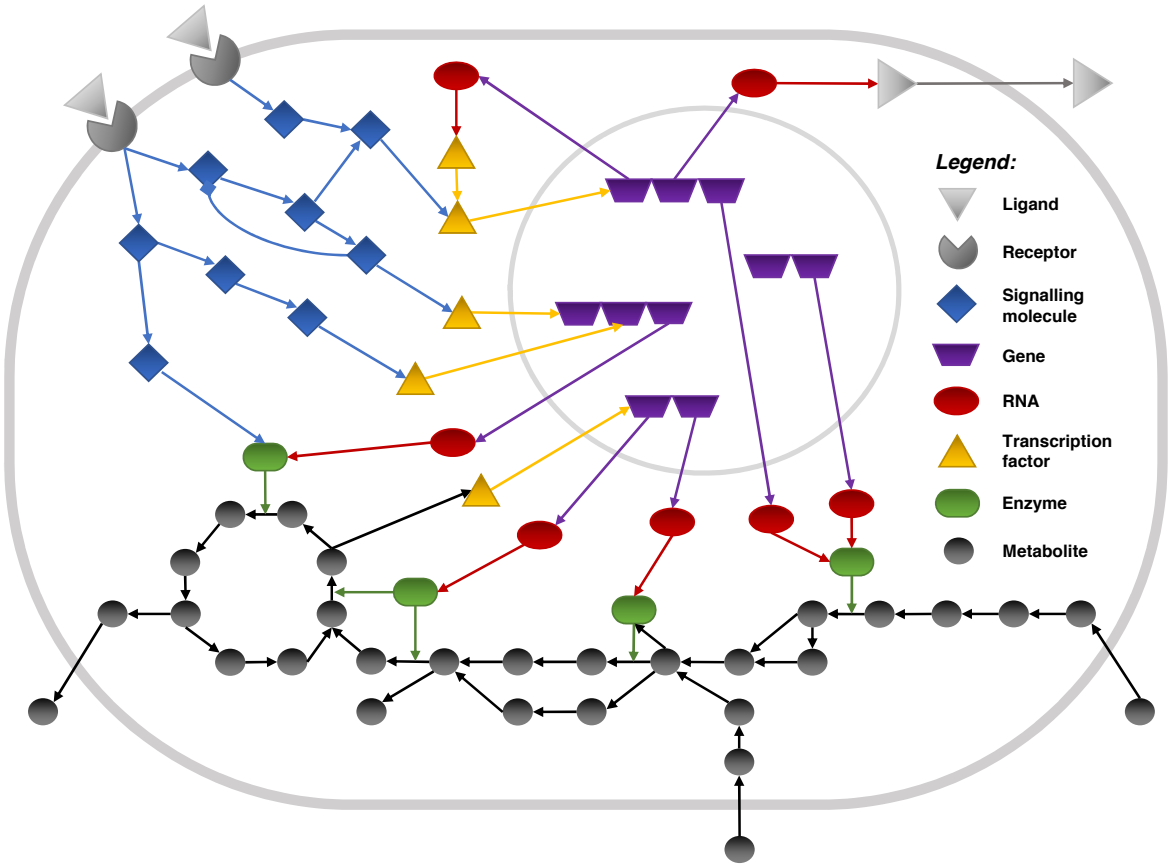


Figure 11. Schematic of the interconnections and regulations among signalling, gene regulation and metabolism, adapted from (Gonçalves et al., 2013).

As the regulations of cells upon perturbations are dynamic, the signatures of cellular responses will evolve overtime until the perturbations are subsided or until the cells reach a new equilibrium or so-called 'steady-state'. Measuring the changes of molecular signatures at an early time point might be beneficial once the precise time point of perturbation is known, e.g. in an *in vitro* experiment. However, in the context of long-term developing diseases such as cancers, it might be more sensible to observe the changes of molecular signatures at steady-state which correlate more to the expressed and sustained pathological phenotypes. Even if some dynamical changes such as oscillation might be observed at steady-state, it is more likely that the averaged measured molecular profiles will still be higher (or lower) compared to the physiological condition

without perturbation. In addition, clinical data from patients' records are also likely to be steady-state measurements in nature. Hence, studying deregulated signal transduction networks based on steady-state measurements does not only reveal the aberrant regulatory mechanisms at the molecular level but it also elucidates the corresponding changes of phenotypes as clinically observed in patients.

PBN as a suitable framework for modelling signal transduction networks with steady-state measurement.

To get a better understanding of the dynamics and other properties of signal transduction networks which are often deregulated in many types of disease, various modelling approaches in systems biology were increasingly applied in recent years. As previously stated in the Chapter 1.3.4., each modelling approach has its own advantages and disadvantages. Therefore, one has to balance and choose the suitable approach according to the research question and to the type and amount of available data.

For the study of signal transduction networks with steady-state measurements, we demonstrated that PBN is one of the suitable modelling frameworks to study and analyse the properties of the system at the equilibrium state. PBN can be applied to depict interactions within signalling networks with simplified Boolean rules while it could still express molecular state-values in a quantitative scale at steady-state. Modelling signal transduction networks with strictly qualitative modelling approaches such as the original Boolean networks or Petri nets does not require extensive prior knowledge but these models could not express quantitative measures neither because of their qualitative nature. On the other hand, modelling the systems with more quantitative approaches such as fuzzy logic models or ODE-based models is considered to be too detailed for fitting to steady-state data. More importantly, the system dynamics as obtained from quantitative models might not always be accurate as there might exist multiple model variants that are able to fit the same set of steady-state measurements. Also, the problem with unidentifiability of kinetic parameters commonly occurs due to the limited amount of data. In parallel, the fitting time for detailed mechanistic models is usually higher than the one of PBN. Hence, in terms of application and efficiency, PBNs seem to be a more suitable modelling approach to fit steady-state data where quantitative details are expected to be derived from the model.

In a mathematical view, both BN and PBN were generally designed to study the properties of networks at steady-state, e.g. in the form of attractors or steady-state distributions. The mathematical foundation to calculate or approximate steady-state

distributions are well-established which allows for a comparison of these steady-state properties to steady-state data from biological experiments (Shmulevich et al., 2003). Even if it is also possible to study network properties at steady-state with ODE-based models, the simulated trajectories of molecular states have to be long enough to ensure that there will be no change in the state-value at the time point which is considered to reach steady-state. When there exist multiple steady-states, a high number of repeated simulations with different initial conditions are also required to discover all possible steady-states. These situations often require longer simulation time and it is considered to be impractical for the study of large-scale networks

PBN models of signal transduction networks with simplified Boolean rules are able to fit steady-state data proteomic data and return quantitative biological insights.

In this dissertation, we demonstrated the applications of PBN on the study of 3 signalling networks including the apoptotic network in primary hepatocytes, deregulated PDGF signalling in GIST, and comparative L-plastin signalling in non-metastatic and metastatic breast cancer cell lines. An important remark from the 3 studies is that many types of interactions within signalling networks can be modelled with simplified Boolean rules. It was known that different types of biochemical reactions take place in signalling networks, e.g. phosphorylation, dephosphorylation as well as proteasomal degradation via ubiquitination. Nevertheless, we demonstrated that all of these reactions can be simply depicted with the AND, OR and NOT gates in a set of Boolean rules within PBN which in turn allows for the approximation of steady-state distributions that fit to normalised experimental data. This observation also indicates that detailed mechanistic models are not always required for the quantitative study of signal transduction networks.

In the first PBN study on an apoptotic network in primary hepatocytes, we showed that we can derive a PBN from an established Boolean network which in turn allows us to obtain additional quantitative information from the existing set of experimental data. Besides breaking multiple mechanisms of regulation on signalling molecules from one original Boolean rule into multiple Boolean rules with the corresponding selection probabilities, we only modified the Boolean rules of two additional signalling molecules in order to obtain a PBN model that fitted well to the observed measurement data of two downstream signalling molecules and of apoptotic phenotype. This finding demonstrated that the original apoptotic network derived from literature can also be used as a starting model topology for PBN modelling. In addition, this study also showed that the PBN approach is applicable for the modelling and optimisation of large-scale networks, e.g. with 96 nodes and 105 interactions (Trairatphisan et al., 2014). By applying the grid-based

pipeline of *optPBN*, the optimisation task can be performed in less than an hour while a network of similar size, e.g. the c-MET signalling network in the ODE framework required at least several hours to identify a desirable set of kinetic parameters.

In the second example, we demonstrated that PBN can be used to study a deregulated PDGF signalling in GIST. Similar to the study of the apoptotic signalling network in a physiological condition in primary hepatocytes, PBN can also be applied to study an aberrant signalling networks in a pathological context such as in cancers. In this study, we showed once again that we can derive a network topology from literature as a set of simplified Boolean rules in a PBN model which fitted well to a large set of steady-state experimental data. In addition, we could observe in this example that the core model topology of the PDGF signalling network in PBN formalism with only linear interactions in each intracellular signalling pathway could fit and predict the measurement data relatively well. However, it was also shown that an essential crosstalk interaction is still required to fit the data points where the influences from the cross regulations between downstream signalling cascades take place. Such finding gives a clue that most of the signals in aberrant PDGF signalling networks in GIST might actually be linearly transduced with only minimal fine-tuning between downstream signalling cascades. Hence, targeting the signalling molecules in the canonical pathways might have a higher impact on the treatment of GIST compared to focusing on shutting down crosstalk interactions.

In the last study, we demonstrated that we could apply the PBN approach to perform a comparative study of L-plastin signalling profiles in multiple non-metastatic and metastatic breast cancer cell lines. Similar to the previous examples, we could fit a literature-derived PBN model of L-plastin signalling with simplified Boolean rules to an extensive set of measurement data at steady-state. The interesting remark on this study is that we could apply a single PBN model topology to fit up to 4 breast cancer cell lines where a different set of selection probabilities, which are considered as interaction weights, was identified for each cell line. These results allowed us to directly compare how the signals flow in each cell line and also to identify the regulatory mechanisms that these cancer cell lines have in common, e.g. the positive crosstalk from PKC towards PKA and the positive connectivity from RSK towards L-plastin. In addition, signal flow comparison can be performed based on the phenotypes of cell lines such as the role of 8-bromo-cAMP on the activation of PKA is slightly stronger in the invasive cancer cell lines (BT-20 and MDA-MB-435S) while the basal activities of PKA in these cell lines are slightly weaker than the ones in non-invasive cell lines (MCF7 and SK-BR-3). The comparative observations as derived from the weights of interactions might be useful to guide towards

the targets of intervention which control the metastatic process in breast cancers. The detections and inhibitions of these targets might subsequently be further developed into diagnostic approaches or therapeutic agents such as drugs or biomarkers which are beneficial for breast cancer patients in the future.

PBN can cope with uncertainty in a literature-derived network topology in an effective and efficient manner.

In the 3 PBN models as demonstrated previously, each of them has a substantial amount of interactions from literature where their optimised selection probabilities are close to zero. This implies that these interactions can be excluded without decreasing the fit quality, or in another word, these interactions are not necessary to be included in the models in order to explain the measurement data within the context of the study. For instance, the crosstalk interaction from deubiquitinated receptor associated receptor kinase-1 (RIP-deubi) towards the formation of TNF-receptor-α1 signalling complex 2 was shown to be inessential in the context of apoptotic signalling in primary hepatocyte with single ligand or ultra-violet-B (UVB) stimulation. In the second PBN study, we demonstrated that a majority of crosstalk interactions which were proposed in literature within the context of PDGF signalling were not required to explain the steady-state proteomics data that we investigated. Lastly, we showed that the crosstalk interactions from PKC and PKA towards L-plastin which were proposed and validated in different experimental settings can indeed be excluded within the context of our four breast cancer cell lines. The results from these 3 PBN studies point to the same trend that the network topologies derived from comprehensive literature reviews tend to be larger than the minimal topology that are required to fit the data.

Even if the starting model topologies of the PBN examples that we explored contain a certain number of unnecessary interactions, we demonstrated that they could still fit well to the data without sacrificing a major portion of optimisation time. In return, the optimised PBN model with unnecessary interactions integrated also pointed out whether the interactions in question are indeed essential or not by expressing the importance in terms of selection probabilities. Such findings allow us to rapidly identify the core regulatory mechanisms that govern the investigated networks in a contextualised manner.

Applying PBN for studying multiple signalling networks in cancers.

According to the concept of cancer signalling pathways which elicits the acquisition of the cancer hallmarks as introduced by Hanahan and Weinberg (Hanahan & Weinberg, 2000, 2011), it was demonstrated that the deregulation of different intracellular signalling

pathways contributes to distinct cancer phenotypes. For instance, deregulated MAPK pathway usually leads to cell over-proliferation while aberrant PI3K/AKT/mTOR pathway often promotes cell survival via inhibiting pro-apoptotic proteins. These two intracellular signalling cascades can be activated by a wide range of signalling receptors such as cytokine receptors, integrins, as well as a large family of receptor tyrosine kinases. To comprehensively understand the signal transduction process in cancers, one should therefore aim to model and analyse a large-scale signalling network comprising multiple signalling pathways as depicted in Figure 1 which will subsequently allow for designing therapeutic agents that could suppress multiple cancer hallmarks at once.

In recent years, there has been some effort to build mathematical models of multiple signalling pathways for the discovery and assessment of therapeutic targets. The aim of these models is to facilitate the design of therapeutic agents which inhibit multiple receptor tyrosine kinases (RTKs) and prevent resistance mechanisms at the same time. Some examples of these therapeutic agents include MM-141, a bi-specific monoclonal antibody which targets both ErbB3 and IGF-1 receptors to shut down the signals in PI3K/AKT/mTOR pathway effectively (Fitzgerald et al., 2014), and MM-111, a bi-specific monoclonal antibody against HER2 and ErbB3 receptors that inhibits the signals in both MAPK and PI3K/AKT/mTOR axes (Kirouac et al., 2013). Given that the downstream signalling cascades of these RTKs and other RTKs such as PDGFR and EGFR are similar, the model topology of RTK signalling could be shared where it can be used to fit experimental data across various RTKs and to explore the therapeutic effects of different antibodies.

In this dissertation, we showed that PBN can be used to model and analyse signalling networks in cancer, i.e. the deregulated PDGF signalling in GIST and the deregulated L-plastin signalling in breast cancer cell lines. Hence, our PBN models could also be applied for the study of other cancer-related signalling pathways as stated above. In addition, it should also be possible to apply PBN to integrate the signals from multiple RTKs into a single model which will provide a broader picture on how different RTKs interplay with each other to integrate intracellular signals towards cancer phenotypes. Such integrative models will open up an opportunity to predict the effects of combinatorial treatments from therapeutic targets which is similar to the prediction of effects from combined perturbation conditions as shown in the PDGF signalling study.

Complete the loop of regulations: multi-scale modelling with PBN.

As previously illustrated in Figure 12, multiple cellular processes such as gene regulation, signal transduction, and metabolism, work together in an orchestrated fashion to regain homeostasis upon perturbations. For instance, signals are transduced from extracellular ligands towards transcription factors to regulate gene expression which in turn controls the translation of signalling proteins and metabolic enzymes. Therefore, modelling only one cellular process at a time might be insufficient to understand the global regulatory processes which are all connected. To date, there are several systems biology studies which proposed novel integrative modelling approaches to combine the mathematical descriptions of multiple cellular processes into a single multi-scale model. Some examples include the whole cell model of *Mycobacterium Genitalium* (Karr et al., 2012) and the integrated generalised Boolean-based signalling and/or gene regulatory network with a kinetic ODE-based model of metabolic network (Ryll et al., 2014). These two examples illustrate an initiative for building a mechanistic multi-scale model which comprehensively represents cellular regulations and responses at different functional layers as well as at the whole cell level.

Following the initiative on multi-scale modelling, we also consider PBN as an alternative modelling approach for building a multi-scale model. In a simplified view, PBN could serve as the mainstream modelling approach. First of all, PBN is well-studied in the context of gene regulatory networks. As we demonstrated in this dissertation that PBN could also be applied to model signal transduction networks, the integration of these two functional layers within the PBN framework should not pose any issue. In parallel, the connection between the gene regulatory network in Boolean formalism and the metabolic networks in the constrained-based modelling (CBM) formalism with probabilities was already established in PROM, a probabilistic regulation of metabolism framework (Chandrasekaran & Price, 2010). Hence, connecting the first two functional layers in the PBN framework with a metabolic network in CBM format should also be possible. As a final outcome, we expect to see an integrated PBN model of gene regulatory and signal transduction networks connecting to a constraint-based model of metabolism which yields quantitative prediction at the level of metabolites based on the changes in gene expression or in the proteomic profile.

On a contrast view, we anticipate that PBN models can also be generalised into a detailed kinetic model which could subsequently be integrated into a quantitative multi-scale model following a similar methodology as proposed by (Ryll et al., 2014). Currently, an approach to transform PBN into a detailed mechanistic model, e.g. into an ODE-based framework, was not yet established. Nevertheless, we envisage that the Boolean logic

gates within PBN can be generalised into the fuzzy logic or ODE-based frameworks as proposed in (Wittmann et al., 2009) so the conversion of PBNs to ODE-based models seems to be feasible. Once the Boolean logic gates were transformed into algebraic equations, the integration of generalised PBN-based gene regulatory and signalling models into a mechanistic ODE-based model of metabolic network should be relatively straight-forward. Even if this pipeline of model integration is not entirely new, we expect that the final multi-scale model derived from generalised PBNs should still provide more quantitative information comparing to one derived from the generalised Boolean networks as proposed by (Ryll et al., 2014).

Potential applications of the PBN approach in a pharmaceutical setting.

Nowadays, a number of pharmaceutical companies apply a systems biology-based approach to discover potential therapeutic targets and to design the properties of therapeutic agents with mathematical models. Among them, Merrimack Pharmaceuticals (Boston, MA, USA) is one of the leading pharmaceutical companies which applies a quantitative systems biology approach called *Network biology* to analyse the dynamics and other properties of deregulated signal transduction networks of different RTKs that are involved in the pathogenesis of cancers. Generally, quantitative mechanistic models of signal transduction networks, e.g. the ones in an ODE-based framework, are required to decipher quantitative signal transfers within signalling cascades which allow for the assessment of therapeutic targets as demonstrate in Chapter 4.4.1. With respect to this requirement, fine details of molecular interactions in multiple steps were derived into a large set of differential equations to precisely depict the mechanisms of signal transduction processes within the signalling networks of interest. Such detailed assignment often leads to an explosion of the model size and of the number of kinetic parameters which consequently requires more computational time and resources for the optimisation as well as more data to precisely infer kinetic parameters.

As we demonstrated and discussed earlier, we discovered in our PBN examples that large and comprehensive network topologies derived from literature often contain unnecessary interactions which are not required to fit the observed data. The large size of networks is one of the factors that leads to a high computational demand and this is a common problem across all modelling frameworks including the ODE-based models. To account for this problem, we believe that the optimisation of large ODE-based models can be performed much faster once the interactions in the network which do not fit to the context of study are removed. In this regard, we need an efficient approach to contextualise the network topology based on the experimental data prior to the modelling

of detailed mechanistic models. As demonstrated in Chapter 4.1, 4.2 and 4.3, the analyses of the optimisation results from *optPBN* could help to identify non-essential interactions within the network in a fast and effective manner. Therefore, we consider that PBN is one of the modelling frameworks that has a high potential to be applied for such efficient network contextualisation tasks.

We demonstrated in a small example model that the network topology pre-selected by *optPBN* indeed corresponded to the minimal ODE-based model counterpart that still fits well to experimental data (see Chapter 4.4.2). This is a proof-of-concept example to demonstrate the application of PBN on network pre-selection which would save time in pharmaceutical research compared to directly performed trail-and-error combinatorial tests within detailed modelling frameworks. Nevertheless, it should be noted that this still needs to be further explored in an actual pharmaceutical setting.

Another potential application of PBN in a pharmaceutical setting is a quick screening of combinatorial effects from multiple therapeutic agents. We showed in Chapter 4.2 that we can predict the signalling profiles of combined perturbation experiments based on the PBN model with optimised parameters inferred from a set of single perturbation experiments within the PBN framework. We envisage that a similar concept can be applied in a pharmaceutical setting. Once the individual effects of multiple therapeutic agents were depicted as a set of Boolean rules, these mechanisms of action can be translated into the model descriptions in *optPBN* which allows for an efficient *in silico* simulation in the setting where the individual effect from each therapeutic agent is integrated. The results from *optPBN* might thereby allow to predict the combinatorial effects of multiple therapeutic agents in a fast and effective manner.

Challenges and outlooks:

Applying the PBN approach to depict non-linear relationships within signal transduction networks.

Even if we demonstrated that we could apply PBN models to fit the proteomic profiles of several signalling networks where linear relationships and additive effects were shown, we should not exclude from our consideration the evidences of non-linear relationships in signal transduction networks such as non-competitive inhibitions or synergistic effects as previously shown in literature. An example of non-linear relationships in literature includes a synergistic effect on the expression of Intercellular adhesion molecule-1 (ICAM-1) in response to a combinatorial treatment of stem cell factor (SCF) and TNFα on human leukemic mast cell-1 (Tsang, Wong, Ip, & Lam, 2005). To account for non-linear relationships in the BN framework, a combination of fundamental Boolean logic gates AND, OR and NOT was proposed to represent the mechanisms of complex interactions (R.-S. Wang et al., 2012). It was also shown that a large combination of Boolean operators can be applied to depict non-linear behaviours such as complex patterns in cryptography (Sarkar & Subhamoy, 2000) which was applied in computer sciences and engineering. Since PBN also applies Boolean logic gates to model the interactions within the framework, we therefore anticipate that complex biochemical interactions with non-linear relationships could also be represented by combinations of Boolean operators as shown in the examples from computer sciences.

In the original concept of Boolean networks, the AND gate can be applied to connect multiple components which are absolutely required together in a reaction such as the formation of protein complexes from subunits. In parallel, the OR gate can be applied when only one from multiple components is sufficient for the execution of a reaction such as the activation of proteins that have multiple catalytic sites. Combined with the concept of selection probabilities in PBN, we could demonstrate in Chapter 4.1 and 4.2 that many complex mechanisms of interaction within the networks can be depicted by the combination of Boolean logic gates which operate on their targets at certain probabilities. Nevertheless, certain non-linear relationships such as synergy is still beyond the capacity of the original PBN framework to account for this positive connectivity. Such problem could be accounted by normalising all signals by the maximal synergised signal to obtain the maximal probability of one, then introduce additional parameters to specifically add the synergistic effects towards to the target nodes. This proposed method still needs to be tested if it will work properly on a real dataset from biological experiments.

A perspective on the optimisation of PBN models with time-course data

So far, we have demonstrated that PBN is a suitable approach for modelling of signal transduction networks with steady-state measurements. However, in many biological studies, time-course data are generated to observe dynamical changes of signalling molecules over the course of experiments. Generally, the original Boolean network and its variants including PBN are not meant to be studied in the context of dynamical time because of their discrete-time nature. Nevertheless, we would like to discuss here the possibilities to take advantage of time-course data for the modelling and analyses of signal transduction networks within the PBN framework.

In previous studies, there were some efforts to correlate time steps in Boolean networks to the "snapshot" dynamics of signalling networks at several time points. For instance, Saez-Rodriguez *et al*. built a logical T-cell receptor signalling network considering two different time-scales to represent the early and late processes of T-cell receptor activation (Saez-Rodriguez et al., 2007). Similarly, the model topology of the large-scale Boolean model of apoptosis by Schlatter *et al.* was separated into several sub-models at different time points which corresponded to the expected dynamics of signalling networks and feedback regulations in an actual time-scale (Schlatter et al., 2009). Both of these examples demonstrated that it is possible to build Boolean networks that fit to cellular phenotypes at different time-scales. A similar methodology could be applied in the PBN framework as well.

Optimisation of large-scale models in the PBN framework.

In Chapter 4.1, we demonstrated that *optPBN* was successfully applied to optimise a large-scale PBN model of apoptosis with 96 nodes in a timely manner. However, the optimisation of more comprehensive networks such as an integrated signal transduction network in cancer, e.g. as shown in Figure 1, or a multi-scale model of comprehensive cellular processes, e.g. as shown in Figure 12, would require either more computational resources or faster algorithms to solve the optimisation problems efficiently. As the cluster-based or grid-based computational resources are generally shared by a large number of users in the community, these resources might not always be available in time of need. Therefore, more efforts should be emphasized on advancing the efficiency and effectiveness of the optimisation pipelines and algorithms rather than relying only on a "brute-force" approach with the usage of large computational resources.

Currently, we foresee that the problem on limited computational resources and an inefficient optimisation pipeline can be tackled by two different approaches. The first

approach is to improve the implementation of the already well-established PBN framework, e.g. in an alternative programming language. In general, Matlab is an efficient modelling language to process computational tasks in matrix form. However, the simulation task in *optPBN* which is based on the calculation of molecular state trajectories stored in vectors did not take full advantage of Matlab. In addition, the *BN/PBN* toolbox, on which the *optPBN* toolbox is based on, comprises a bundle of connected Matlab scripts which are inefficient in terms of computation. Based on such limitation, our collaborators at the Computer Science and Communications Research Unit at the University of Luxembourg developed *ASSA-PBN*, a Java-based toolbox which allows for the simulation of PBN models in a highly efficient manner (Mizera, Pang, & Yuan, 2015). In the current implementation, *ASSA-PBN* could perform PBN simulations with a trajectory length of more than one million time steps on a standard computer in a timely manner while the longest length that *optPBN* can simulate in the same timeframe is no longer than a few ten thousands time steps.

From a technical view, as the calculation of influences from parent nodes to a target node which was introduced in (Shmulevich et al., 2002) might require the calculation of joint distributions, the required lengths of a trajectory to reach steady-state with a desired accuracy are generally much longer than the ones computed for the independent marginal distribution of each node as assumed in the *optPBN* study. Also, decreasing the error range of approximated steady-state distributions on the accuracy parameter 'r' also increases the required lengths of the trajectory in a substantial proportion. As previously stated, the increased trajectory lengths in PBNs could be handled efficiently by *ASSA-PBN*. Therefore, the implementation of *ASSA-PBN* opens up a possibility to obtain more accurate PBN results and to obtain more detailed features such as influence values which in turn describe the impact of the changes on parent nodes towards their target nodes.

In terms of limitation, *ASSA-PBN* still needs an input generated from the *optPBN* toolbox which combines the model descriptions in Boolean rules and experimental data into an optimisation problem. Hence, *ASSA-PBN* still partly relies on the functionality of the *optPBN* toolbox. Future work will be to establish a PBN construction pipeline from Boolean rules within the *ASSA-PBN* framework and also to provide a user-friendly interface for the Java-based tool. Regarding the computation, the implementation of parallel computing in the *ASSA-PBN* framework on a cluster-based or grid-based infrastructure is yet to be implemented. A parallelisation on a general-proposed graphic processing unit (GPU) as introduced by (Bosnacki, Odenbrett, Wijs, Ligtenberg, & Hilbers,

2012) is also an interesting approach to maximise the computational power on a standard local machine.

The second approach to account for the problem of limited computational resources and an inefficient optimisation algorithm is to explore an alternative form of PBN representation where the optimisation problems in the respective form can be solved more efficiently. Recently, we developed a new efficient optimisation tool called *FALCON*, a Fast ALgorithm for COntextualisation of Networks, based on the concept of *optPBN*. *FALCON* takes two input files, i.e. a simple list of network interactions and a list of experimental conditions with the corresponding measurement data. Only two types of simplified interactions were considered in *FALCON*, i.e. activation and inhibition. The Boolean logic gates 'AND' and 'OR' can be assigned to combine multiple inputs (optional). Such simple input format should eliminate the inconsistency when deriving network interactions either from a list of biochemical interactions or from a network graph. Once the model descriptions were derived, we applied a Matlab solver to fit models to experimental data and to optimise the parameter values in an efficient way. Please note that we refrain from providing detailed descriptions on the methodology of *FALCON* in this dissertation as we would like to explore the possibility to commercialise this software in the near future.

To evaluate the performance of the new tool, we applied *FALCON* to perform the optimisation of the deregulated PDGF signalling model in GIST and of the L-plastin signalling model in the BT-20 cell line on a standard PC (1 core) and we compared the results against the ones obtained from the grid-based pipeline of *optPBN* on Grid'5000 (160 cores). The comparative results are shown in Table 2.

Table 2: Compared fitting costs and optimisation time between *optPBN* versus *FALCON*

Model	*optPBN* On Grid'5000 (160 cores)		*FALCON* On standard PC (1 core)		Time Improvement per core (folds)
	Fitting cost	Time	Fitting cost	Time	
PDGF	0.1806	937s	0.1566	17s	8819
L-plastin	0.2765	2395s	0.2996	34s	11271

As shown in Table 2, the time improvement of *FALCON* over *optPBN* is up to 10,000 fold for realistic case studies. Not only that the computational time is significantly improved, we showed that it is also possible to run *FALCON* on a standard PC to solve

the complex optimisation problems of the PDGF and L-plastin signalling pathways while this task is not practical for the *optPBN* pipeline.

In summary, we demonstrated that the PBN approach can be applied to study the steady-state properties of signal transduction networks. Also, it allows for network contextualisation prior to the building of mechanistic models based on steady-state data with promising future applications in pharmaceutical industry. In the near future, once the integrated *optPBN*/*ASSA-PBN* and *FALCON* pipelines are mature enough both in terms of mathematical foundation and applications, it might be worth to perform a benchmarking to compare the efficiency and effectiveness of the two approaches and to explore large-scale applications in pharmaceutical industry. Once these tools will indeed operate effectively and efficiently in an actual pharmaceutical setting, we hope that they will be valuable computational tools which contribute to the advancement of translational sciences from molecular biology towards clinical medicine.

Chapter 6

REFERENCES

Abella, J. V, Peschard, P., Naujokas, M. a, Lin, T., Saucier, C., Urbé, S., & Park, M. (2005). Met / Hepatocyte Growth Factor Receptor Ubiquitination Suppresses Transformation and Is Required for Hrs Phosphorylation Met / Hepatocyte Growth Factor Receptor Ubiquitination Suppresses Transformation and Is Required for Hrs Phosphorylation. *Molecular and Cellular Biology*, *25*, 9632–9645. http://doi.org/10.1128/MCB.25.21.9632

Albert, I., Thakar, J., Li, S., Zhang, R., & Albert, R. (2008). Boolean network simulations for life scientists. *Source Code for Biology and Medicine*, *3*, 16. http://doi.org/10.1186/1751-0473-3-16

Aldridge, B. B., Saez-Rodriguez, J., Muhlich, J. L., Sorger, P. K., & Lauffenburger, D. a. (2009). Fuzzy Logic Analysis of Kinase Pathway Crosstalk in TNF/EGF/Insulin-Induced Signaling. *PLoS Computational Biology*, *5*(4). http://doi.org/10.1371/journal.pcbi.1000340

Altomare, D. a, & Testa, J. R. (2005). Perturbations of the AKT signaling pathway in human cancer. *Oncogene*, *24*, 7455–7464. http://doi.org/10.1038/sj.onc.1209085

Altomare, D. a, You, H., Xiao, G.-H., Ramos-Nino, M. E., Skele, K. L., De Rienzo, A., … Testa, J. R. (2005). Human and mouse mesotheliomas exhibit elevated AKT/PKB activity, which can be targeted pharmacologically to inhibit tumor cell growth. *Oncogene*, *24*(May), 6080–6089. http://doi.org/10.1038/sj.onc.1208744

Bahlawane, C., Eulenfeld, R., Wiesinger, M. Y., Wang, J., Muller, A., Girod, A., … Haan, S. (2015). Constitutive activation of oncogenic PDGFRα-mutant proteins occurring in GIST patients induces receptor mislocalisation and alters PDGFRα signalling characteristics. *Cell Communication and Signaling*, *13*(1). http://doi.org/10.1186/s12964-015-0096-8

Balmanno, K., & Cook, S. J. (2009). Tumour cell survival signalling by the ERK1/2 pathway. *Cell Death and Differentiation*, *16*, 368–377. http://doi.org/10.1038/cdd.2008.148

Benchimol, S. (2001). P53-Dependent Pathways of Apoptosis. *Cell Death and Differentiation*, *8*(11), 1049–1051. http://doi.org/10.1038/sj.cdd.4400918

Blair, R. H., Trichler, D. L., & Gaille, D. P. (2012). Mathematical and statistical modeling in cancer systems biology. *Frontiers in Physiology*, *3 JUN*(June), 1–8. http://doi.org/10.3389/fphys.2012.00227

Bos, J. L. (2000). *Molecular Mechanisms of Signal Transduction*. IOS Press, NATO Science Series, 316, 287pp.

Bosnacki, D., Odenbrett, M. R., Wijs, A., Ligtenberg, W., & Hilbers, P. (2012). Efficient reconstruction of biological networks via transitive reduction on general purpose graphics processors. *BMC Bioinformatics*, *13*, 281. http://doi.org/10.1186/1471-2105-13-281

Braun, S., Heumos, I., Schaller, G., Riethdorf, L., Riethmu, G., & Pantel, K. (2001). erbB2 Overexpression on Occult Metastatic Cells in Bone Marrow Predicts Poor Clinical Outcome of Stage I – III Breast Cancer Patients 1, 1890–1895.

Burness, M. L., Grushko, T. a, & Olopade, O. I. (2010). Epidermal growth factor receptor in triple-negative and basal-like breast cancer: promising clinical target or only a marker? *Cancer Journal (Sudbury, Mass.)*, *16*(1), 23–32. http://doi.org/10.1097/PPO.0b013e3181d24fc1

Calcagno, S. R., Li, S., Colon, M., Kreinest, P. a., Thompson, E. A., Fields, A. P., & Murray, N. R. (2008). Oncogenic K-ras promotes early carcinogenesis in the mouse proximal colon. *International Journal of Cancer*, *122*(August 2007), 2462–2470. http://doi.org/10.1002/ijc.23383

Cecchi, F., Rabe, D. C., & Bottaro, D. P. (2010). Targeting the HGF/Met signalling pathway in cancer. *European Journal of Cancer*, *46*(7), 1260–1270. http://doi.org/10.1016/j.ejca.2010.02.028

Chai, L. E., Loh, S. K., Low, S. T., Mohamad, M. S., Deris, S., & Zakaria, Z. (2014). A review on the computational approaches for gene regulatory network construction. *Computers in Biology and Medicine*, *48*(1), 55–65. http://doi.org/10.1016/j.compbiomed.2014.02.011

Chakraborti, S., Mandal, M., & Das, S. (2003). Regulation of matrix metalloproteinases: an overview. *Molecular and Cellular …*, 269–285. Retrieved from http://link.springer.com/article/10.1023/A:1026028303196

Chandrasekaran, S., & Price, N. D. (2010). Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(41), 17845–17850. http://doi.org/10.1073/pnas.1005139107

Chassagnole, C., Noisommit-Rizzi, N., Schmid, J. W., Mauch, K., & Reuss, M. (2002). Dynamic modeling of the central carbon metabolism of Escherichia coli. *Biotechnology and Bioengineering*, *79*(1), 53–73. http://doi.org/10.1002/bit.10288

Chen, H., Guo, J., Mishra, S. K., Robson, P., Niranjan, M., & Zheng, J. (2014). Single-cell transcriptional analysis to uncover regulatory circuits driving cell fate decisions in early mouse development. *Bioinformatics*.

Chen, X., Akutsu, T., Tamura, T., & Ching, W.-K. (2013). Finding optimal control policy in Probabilistic Boolean Networks with hard constraints by using integer programming and dynamic programming. *Internal Journal of Data Mining and Bioinformatics*, *7*(3), 321–343. http://doi.org/10.1109/BIBM.2010.5706570

Cheng, X., Qiu, Y., Hou, W., Ching, W., Modeling, A., & Kong, H. (2014). A Semi-tensor Product Approach for Probabilistic Boolean Networks. In *The 8th International Conference on Systems Biology (ISB)* (pp. 85–90).

Choi, J. H., Ryu, S. H., & Suh, P. G. (2007). On/Off-regulation of phospholipase C-gamma1-mediated signal transduction. *Advances in Enzyme Regulation*, *47*, 104–116. http://doi.org/10.1016/j.advenzreg.2006.12.010

Clevers, H., & Nusse, R. (2012). Wnt/beta-catenin signaling and disease. *Cell*, *149*(6), 1192–1205. http://doi.org/10.1016/j.cell.2012.05.012

Crick, F. H. C. (1970). Central Dogma of Molecular Biology. *Nature*. *227*, 561-563.

DeNardo, D. G., Andreu, P., & Coussens, L. M. (2010). Interactions between lymphocytes and myeloid cells regulate pro-versus anti-tumor immunity. *Cancer and Metastasis Reviews*, *29*, 309–316. http://doi.org/10.1007/s10555-010-9223-6

Dümcke, S., Bräuer, J., Anchang, B., Spang, R., Beerenwinkel, N., & Tresch, A. (2014). Exact likelihood computation in Boolean networks with probabilistic time delays, and its application in signal network reconstruction. *Bioinformatics*, *30*(3), 414–419. http://doi.org/10.1093/bioinformatics/btt696

Dunn, G. P., Old, L. J., & Schreiber, R. D. (2004). The three Es of cancer immunoediting. *Annual Review of Immunology*, *22*(4), 329–360. http://doi.org/10.1146/annurev.immunol.22.012703.104803

Egeblad, M., Nakasone, E. S., & Werb, Z. (2010). Tumors as organs: Complex tissues that interface with the entire organism. *Developmental Cell*, *18*(6), 884–901. http://doi.org/10.1016/j.devcel.2010.05.012

Elmore, S. (2007). Apoptosis: a review of programmed cell death. *Toxicologic Pathology*, *35*(4), 495–516. http://doi.org/10.1080/01926230701320337

Feron, O. (2009). Pyruvate into lactate and back: From the Warburg effect to symbiotic energy fuel exchange in cancer cells. *Radiotherapy and Oncology*, *92*(3), 329–333. http://doi.org/10.1016/j.radonc.2009.06.025

Ferrara, N., & Davis-Smyth, T. (1997). The Biology of Vascular Endothelial Growth Factor, *18*(1), 4–25. Retrieved from http://press.endocrine.org/doi/full/10.1210/edrv.18.1.0287

Fitzgerald, J. B., Johnson, B. W., Baum, J., Adams, S., Iadevaia, S., Tang, J., … Lugovskoy, A. a. (2014). MM-141, an IGF-IR- and ErbB3-directed bispecific antibody, overcomes network adaptations that limit activity of IGF-IR inhibitors. *Molecular Cancer Therapeutics*, *13*(2), 410–25. http://doi.org/10.1158/1535-7163.MCT-13-0255

Foley, J., Nickerson, N. K., Nam, S., Allen, K. T., Gilmore, J. L., Nephew, K. P., & Riese, D. J. (2010). EGFR signaling in breast cancer: Bad to the bone. *Seminars in Cell and Developmental Biology*, *21*(9), 951–960. http://doi.org/10.1016/j.semcdb.2010.08.009

Fresno Vara, J. A., Casado, E., de Castro, J., Cejas, P., Belda-Iniesta, C., & González-Barón, M. (2004). PI3K/Akt signalling pathway and cancer. *Cancer Treatment Reviews*, *30*, 193–204. http://doi.org/10.1016/j.ctrv.2003.07.007

Friedman, N. (2004). Inferring Cellular Networks Using Probabilistic Graphical Models. *Sciences*, *303*, 799–805. http://doi.org/10.1186/1471-2105-8-S6-S5

Furqan, M., Mukhi, N., Lee, B., & Liu, D. (2013). Dysregulation of JAK-STAT pathway in hematological malignancies and JAK inhibitors for clinical application. *Biomarker Research*, *1*, 1–5. http://doi.org/10.1186/2050-7771-1-5

Gao, Y. M., Xu, P., Wang, X. H., & Liu, W. Bin. (2013). The complex fluctuations of probabilistic Boolean networks. *BioSystems*, *114*(1), 78–84. http://doi.org/10.1016/j.biosystems.2013.07.008

Geva-Zatorsky, N., Rosenfeld, N., Itzkovitz, S., Milo, R., Sigal, A., Dekel, E., … Alon, U. (2006). Oscillations and variability in the p53 system. *Molecular Systems Biology*, *2*, 2006.0033. http://doi.org/10.1038/msb4100068

Gilmore, T. D. (2006). Introduction to NF-kappaB: players, pathways, perspectives. *Oncogene*, *25*(51), 6680–6684. http://doi.org/10.1038/sj.onc.1209954

Gonçalves, E., Bucher, J., Ryll, A., Niklas, J., Mauch, K., Klamt, S., … Saez-Rodriguez, J. (2013). Bridging the layers: towards integration of signal transduction, regulation and metabolism into mathematical models. *Molecular bioSystems*, *9*, 1576–83. http://doi.org/10.1039/c3mb25489e

Graff, C. P., & Wittrup, K. D. (2003). Theoretical analysis of antibody targeting of tumor spheroids: Importance of dosage for penetration, and affinity for retention1. *Cancer Research*, *63*(6), 1288–1296.

Grammer, T. C., & Blenis, J. (1997). Evidence for MEK-independent pathways regulating the prolonged activation of the ERK-MAP kinases. *Oncogene*, *14*(14), 1635–1642. http://doi.org/10.1038/sj.onc.1201000

Gramza, A. W., Corless, C. L., & Heinrich, M. C. (2009). Resistance to tyrosine kinase inhibitors in gastrointestinal stromal tumors. *Clinical Cancer Research*, *15*(24), 7510–7518. http://doi.org/10.1158/1078-0432.CCR-09-0190

Hanahan, D., & Weinberg, R. a. (2000). The Hallmarks of Cancer. *Cell*, *100*(1), 57–70. http://doi.org/10.1016/S0092-8674(00)81683-9

Hanahan, D., & Weinberg, R. a. (2011). Hallmarks of cancer: The next generation. *Cell*, *144*(5), 646–674. http://doi.org/10.1016/j.cell.2011.02.013

Harburger, D. S., & Calderwood, D. a. (2009). Integrin signalling at a glance. *Journal of Cell Science*, *122*(Pt 2), 159–163. http://doi.org/10.1242/jcs.018093

Hardy, S., & Robillard, P. N. (2008). Petri net-based method for the analysis of the dynamics of signal propagation in signaling pathways. *Bioinformatics*, *24*(2), 209–217. http://doi.org/10.1093/bioinformatics/btm560

Heldin, C. H. (2014). Targeting the PDGF signaling pathway in the treatment of non-malignant diseases. *Journal of Neuroimmune Pharmacology*, *9*(2), 69–79. http://doi.org/10.1007/s11481-013-9484-2

Hemmings, B. A., & Restuccia, D. F. (2012). PI3K-PKB / Akt Pathway. *Cold String Harb Perspect Biol*, (4), a011189.

Hoesel, B., & Schmid, J. a. (2013). The complexity of NF-κB signaling in inflammation and cancer. *Molecular Cancer*, *12*(1), 86. http://doi.org/10.1186/1476-4598-12-86

Hoffmann, A., Levchenko, A., Scott, M. L., & Baltimore, D. (2002). The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science (New York, N.Y.)*, *298*(5596), 1241–1245. http://doi.org/10.1126/science.1071914

Hombría, J. C., & Brown, S. (2002). The Fertile Field of Drosophila JAK / STAT Signalling, *12*(02), 569–575.

Huber, M. a, Azoitei, N., Baumann, B., Grünert, S., Sommer, A., Pehamberger, H., … Wirth, T. (2004). NF- κ B is essential for epithelial- mesenchymal transition and metastasis in a model of breast cancer progression. *The Journal of Clinical Investigation*, *114*(4), 569–581. http://doi.org/10.1172/JCI200421358.The

Hwang, J., Oh, Y., Shin, K., Kim, H., Ryu, S. H., & Suh, P. (2005). Molecular cloning and characterization of a novel phospholipase C , PLC-η, *186*, 181–186.

Hynes, N. E., & MacDonald, G. (2009). ErbB receptors and signaling pathways in cancer. *Current Opinion in Cell Biology*, *21*, 177–184. http://doi.org/10.1016/j.ceb.2008.12.010

Imamura, R., Konaka, K., Matsumoto, N., Hasegawa, M., Fukui, M., Mukaida, N., … Suda, T. (2004). Fas ligand induces cell-autonomous NF-κB activation and interleukin-8 production by a mechanism distinct from that of tumor necrosis factor-α. *Journal of Biological Chemistry*, *279*(45), 46415–46423. http://doi.org/10.1074/jbc.M403226200

Iskar, M., Zeller, G., Zhao, X. M., van Noort, V., & Bork, P. (2012). Drug discovery in the age of systems biology: The rise of computational approaches for data integration. *Current Opinion in Biotechnology*, *23*(4), 609–616. http://doi.org/10.1016/j.copbio.2011.11.010

Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., & Forman, D. (2011). Global cancer statistics. *CA Cancer J Clin*, *61*(2), 69–90. http://doi.org/10.3322/caac.20107.Available

Jucker, M., Sudel, K., Horn, S., Sickel, M., Wegner, W., Fiedler, W., & Feldman, R. a. (2002). Expression of a mutated form of the p85 □ regulatory subunit of phosphatidylinositol 3-kinase in a Hodgkin ' s lymphoma-derived cell line (CO). *Leukemia*, *16*, 894–901. http://doi.org/10.1038/sj/leu/2402484

Kaderali, L., Dazert, E., Zeuge, U., Frese, M., & Bartenschlager, R. (2009). Reconstructing signaling pathways from RNAi data using probabilistic boolean threshold networks. *Bioinformatics*, *25*(17), 2229–2235. http://doi.org/10.1093/bioinformatics/btp375

Kao, Y.-T., Saxena, C., Wang, L., Sancar, A., & Zhong, D. (2005). Direct observation of thymine dimer repair in DNA by photolyase. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(45), 16128–16132. http://doi.org/10.1073/pnas.0506586102

Karin, M., & Greten, F. R. (2005). NF-kappaB: linking inflammation and immunity to cancer development and progression. *Nature Reviews. Immunology*, *5*(10), 749–759. http://doi.org/10.1038/nri1703

Karr, J. R., Sanghvi, J. C., MacKlin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B., … Covert, M. W. (2012). A whole-cell computational model predicts phenotype from genotype. *Cell*, *150*, 389–401. http://doi.org/10.1016/j.cell.2012.05.044

Kassis, J., Moellinger, J., Lo, H., Greenberg, N. M., Kim, H. G., & Wells, a. (1999). A role for phospholipase C-gamma-mediated signaling in tumor cell invasion. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, *5*(August), 2251–2260.

Kauffman, S. a. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, *22*(3), 437–467. http://doi.org/10.1016/0022-5193(69)90015-0

Kiani, N. a, & Kaderali, L. (2014). Dynamic Probabilistic Threshold Networks to Infer Signaling Pathways from Time-Course Perturbation Data. *BMC Bioinformatics*, *15*(1), 250. http://doi.org/10.1186/1471-2105-15-250

Kim, E. K., & Choi, E.-J. (2010). Pathological roles of MAPK signaling pathways in human diseases. *Biochimica et Biophysica Acta*, *1802*(4), 396–405. http://doi.org/10.1016/j.bbadis.2009.12.009

Kim, M. J., Chang, J. S., Park, S. K., Hwang, J. I., Ryu, S. H., & Suh, P. G. (2000). Direct interaction of SOS1 Ras exchange protein with the SH3 domain of phospholipase C-γ1. *Biochemistry*, *39*, 8674–8682. http://doi.org/10.1021/bi992558t

Kirouac, D. C., Du, J. Y., Lahdenranta, J., Overland, R., Yarar, D., Paragas, V., … Onsum, M. D. (2013). Computational modeling of ERBB2-amplified breast cancer identifies combined ErbB2/3 blockade as superior to the combination of MEK and AKT inhibitors. *Science Signaling*, *6*(288), ra68. http://doi.org/10.1126/scisignal.2004008

Kitano, H. (2002). Systems biology: a brief overview. *Science (New York, N.Y.)*, *295*(5560), 1662–4. http://doi.org/10.1126/science.1069492

Klymkowsky, M. W., & Savagner, P. (2009). Epithelial-mesenchymal transition: a cancer researcher's conceptual friend and foe. *The American Journal of Pathology*, *174*(5), 1588–1593. http://doi.org/10.2353/ajpath.2009.080545

Kobayashi, K., & Hiraishi, K. (2014). Verification and optimal control of context-sensitive probabilistic Boolean networks using model checking and polynomial optimization. *The Scientific World Journal*, *2014*. http://doi.org/10.1155/2014/968341

Kolch, W., Calder, M., & Gilbert, D. (2005). When kinases meet mathematics: The systems biology of MAPK signalling. *FEBS Letters*, *579*(8), 1891–1895. http://doi.org/10.1016/j.febslet.2005.02.002

Land, H., Parada, L. F., & Weinberg, R. a. (1983). Cellular oncogenes and multistep carcinogenesis. *Science (New York, N.Y.)*, *222*, 771–778. http://doi.org/10.1136/bmj.287.6399.1084

Lawrence, M. C., Jivan, A., Shao, C., Duan, L., Goad, D., Zaganjor, E., … Cobb, M. H. (2008). The roles of MAPKs in disease. *Cell Research*, *18*(4), 436–442. http://doi.org/10.1038/cr.2008.37

Levy, D. E., & Lee, C. K. (2002). What does Stat3 do? *Journal of Clinical Investigation*, *109*(9), 1143–1148. http://doi.org/10.1172/JCI200215650

Li, C., Ge, Q. W., Nakata, M., Matsuno, H., & Miyano, S. (2007). Modelling and simulation of signal transductions in an apoptosis pathway by using timed Petri nets. *Journal of Biosciences*, *32*(1), 113–127. http://doi.org/10.1007/s12038-007-0011-6

Li, F., Long, T., Lu, Y., Ouyang, Q., & Tang, C. (2004). The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(14), 4781–4786. http://doi.org/10.1073/pnas.0305937101

Li, N., Xiang, G. S., Dokainish, H., Ireton, K., & Elferink, L. a. (2005). The Listeria protein internalin B mimics hepatocyte growth factor-induced receptor trafficking. *Traffic*, *6*, 459–473. http://doi.org/10.1111/j.1600-0854.2005.00290.x

Li, Z., Song, J., & Yang, J. (2014). Partial Stability of Probabilistic Boolean network. In *Control and Decision Conference (2014 CCDC), The 26th Chinese* (pp. 1952–1956).

Lo, H. W., Hsu, S. C., & Hung, M. C. (2006). EGFR signaling pathway in breast cancers: From traditional signal transduction to direct nuclear translocalization. *Breast Cancer Research and Treatment*, *95*(3), 211–218. http://doi.org/10.1007/s10549-005-9011-0

Lynch, D. K., & Daly, R. J. (2002). PKB-mediated negative feedback tightly regulates mitogenic signalling via Gab2. *EMBO Journal*, *21*(1-2), 72–82. http://doi.org/10.1093/emboj/21.1.72

Machado, D., Costa, R. S., Rocha, M., Ferreira, E. C., Tidor, B., & Rocha, I. (2011). Modeling formalisms in Systems Biology. *AMB Express*, *1*(1), 45. http://doi.org/10.1186/2191-0855-1-45

Malumbres, M., & Barbacid, M. (2009). Cell cycle, CDKs and cancer: a changing paradigm. *Nature Reviews. Cancer*, *9*(3), 153–166. http://doi.org/10.1038/nrc2602

Manning, B. D., & Cantley, L. C. (2007). AKT/PKB Signaling: Navigating Downstream. *Cell*, *129*, 1261–1274. http://doi.org/10.1016/j.cell.2007.06.009

Markevich, N., Moehren, G., Demin, O., Kiyatkin, A., Hoek, J., & Kholodenko, B. (2004). Signal processing at the Ras circuit: what shapes Ras activation patterns? *Systems Biology*, *1*(1), 104–113. http://doi.org/10.1049/sb

Markku, M., & Jerzy, L. (2006). Gastrointestinal Stromal Tumors Review on Morphology, Molecular Pathology, Prognosis, and Differential Diagnosis. *Virchows Archiv : An International Journal of Pathology*, *130*, 1466–1478. http://doi.org/10.1016/B978-1-4160-5544-0.00081-7

Mendoza, M. C., Er, E. E., & Blenis, J. (2011). The Ras-ERK and PI3K-mTOR pathways: Cross-talk and compensation. *Trends in Biochemical Sciences*, *36*(6), 320–328. http://doi.org/10.1016/j.tibs.2011.03.006

Merchant, M., Ma, X., Maun, H. R., Zheng, Z., Peng, J., Romero, M., … Yansura, D. G. (2013). Monovalent antibody design and mechanism of action of onartuzumab, a MET antagonist with anti-tumor activity as a therapeutic agent. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, E2987–96. http://doi.org/10.1073/pnas.1302725110

Miranda, E. N., & Parga, N. (2007). Noise Effects in the Kauffman Model. *Europhysics Letters (EPL)*, *10*(4), 293–298. http://doi.org/10.1209/0295-5075/10/4/002

Mizera, A., Pang, J., & Yuan, Q. (2015). Reviving the Two-state Markov Chain Approach (Technical Report).

Mura, I., & Csikász-Nagy, A. (2008). Stochastic Petri Net extension of a yeast cell cycle model. *Journal of Theoretical Biology*, *254*(4), 850–860. http://doi.org/10.1016/j.jtbi.2008.07.019

Nanney, L. B., Gates, R. E., King, L. E., & Carpenter, G. (1992). Altered Distribution of Phospholipase C-g1 in Benign Hyperproliferative Epidermal Diseases. *Cell Growth & Differentiation*, *3*(April), 233–239.

Ouyang, L., Shi, Z., Zhao, S., Wang, F. T., Zhou, T. T., Liu, B., & Bao, J. K. (2012). Programmed cell death pathways in cancer: A review of apoptosis, autophagy and programmed necrosis. *Cell Proliferation*, *45*(6), 487–498. http://doi.org/10.1111/j.1365-2184.2012.00845.x

Parsons, S. J., & Parsons, J. T. (2004). Src family kinases, key regulators of signal transduction. *Oncogene*, *23*(48), 7906–7909. http://doi.org/10.1038/sj.onc.1208160

Petri, C. A. (1962). Kommunikation mit Automaten. *Fakultät Für Mathematik Und Physik*, *Doktor*, 128.

Putney, J. W. (2002). PLC-gamma: an old player has a new role. *Nature Cell Biology*, *4*(December), E280–E281. http://doi.org/10.1038/ncb1202-e280

Qian, X., & Dougherty, E. R. (2013). Validation of gene regulatory network inference based on controllability. *Frontiers in Genetics*, *4*(DEC), 1–13. http://doi.org/10.3389/fgene.2013.00272

Rammohan, A., Sathyanesan, J., Rajendran, K., Pitchaimuthu, A., Perumal, S.-K., Srinivasan, U., … Govindan, M. (2013). A gist of gastrointestinal stromal tumors: A review. *World Journal of Gastrointestinal Oncology*, *5*(6), 102–12. http://doi.org/10.4251/wjgo.v5.i6.102

Rangamani, P., & Iyengar, R. (2007). Modelling spatio-temporal interactions within the cell. *Journal of Biosciences*, *32*(1), 157–167. http://doi.org/10.1007/s12038-007-0014-3

Rawlings, J. S., Rosler, K. M., & Harrison, D. a. (2004). The JAK/STAT signaling pathway. *Journal of Cell Science*, *117*, 1281–1283. http://doi.org/10.1242/jcs.00963

Rodriguez-Viciana, P., Warne, P. H., Dhand, R., Vanhaesebroeck, B., Gout, I., Fry, M. J., … Downward, J. (1994). Phosphatidylinositol-3-OH kinase as a direct target of Ras. *Nature*, *370*(6490), 527–532. http://doi.org/10.1038/370527a0

Roy, a. C., Park, S. R., Cunningham, D., Kang, Y. K., Chao, Y., Chen, L. T., … de Gramont, a. (2013). A randomized phase II study of PEP02 (MM-398), irinotecan or docetaxel as a second-line therapy in patients with locally advanced or metastatic gastric or gastro-oesophageal junction adenocarcinoma. *Annals of Oncology*, *24*(6), 1567–1573. http://doi.org/10.1093/annonc/mdt002

Ryll, a, Bucher, J., Bonin, a, Bongard, S., Gonçalves, E., Saez-Rodriguez, J., … Klamt, S. (2014). A model integration approach linking signalling and gene-regulatory logic with kinetic metabolic models. *Bio Systems*. http://doi.org/10.1016/j.biosystems.2014.07.002

Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D. a, & Nolan, G. P. (2005). Causal protein-signaling networks derived from multiparameter single-cell data. *Science (New York, N.Y.)*, *308*(5721), 523–529. http://doi.org/10.1126/science.1105809

Saez-Rodriguez, J., Simeoni, L., Lindquist, J. a., Hemenway, R., Bommhardt, U., Arndt, B., … Schraven, B. (2007). A logical model provides insights into T cell receptor signaling. *PLoS Computational Biology*, *3*(8), 1580–1590. http://doi.org/10.1371/journal.pcbi.0030163

Samstag, Y., Eibert, S. M., Klemke, M., & Wabnitz, G. H. (2002). Actin cytoskeletal dynamics in T lymphocyte activation and migration of the immunological synapse at the interface be-. http://doi.org/10.1189/jlb.0602272.http

Sansal, I., & Sellers, W. R. (2004). The biology and clinical relevance of the PTEN tumor suppressor pathway. *Journal of Clinical Oncology*, *22*, 2954–2963. http://doi.org/10.1200/JCO.2004.02.141

Sarkar, P., & Subhamoy, M. (2000). Construction of Nonlinear Boolean Functions with Important Cryptographic Properties. *Advances in Cryptology—EUROCRYPT 2000*, *1807*(2), 491–511. http://doi.org/10.1007/3-540-45539-6

Scaltriti, M. (2006). The Epidermal Growth Factor Receptor Pathway : A Model for T argeted Therapy, *12*(18), 5268–5273. http://doi.org/10.1158/1078-0432.CCR-06-1554

Schindler, C. W. (2002). JAK-STAT signaling in human disease JAK-STAT signaling in human disease. *New York*, *109*(9), 1133–1137. http://doi.org/10.1172/JCI200215644.Cytokines

Schlatter, R., Schmich, K., Vizcarra, I. A., Scheurich, P., Sauter, T., Borner, C., … Sawodny, O. (2009). ON/OFF and beyond - A Boolean model of apoptosis. *PLoS Computational Biology*, *5*(12), e1000595. http://doi.org/10.1371/journal.pcbi.1000595

Schmich, K., Schlatter, R., Corazza, N., Ferreira, K. S., Ederer, M., Brunner, T., … Merfort, I. (2011). Tumor necrosis factor α sensitizes primary murine hepatocytes to Fas/CD95-induced apoptosis in a Bim- and Bid-dependent manner. *Hepatology*, *53*(1), 282–292. http://doi.org/10.1002/hep.23987

Schmidt, H., & Jirstrand, M. (2006). Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology. *Bioinformatics (Oxford, England)*, *22*(4), 514–5. http://doi.org/10.1093/bioinformatics/bti799

Schoeberl, B., Eichler-Jonsson, C., Gilles, E. D., & Müller, G. (2002). Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nature Biotechnology*, *20*(4), 370–375. http://doi.org/10.1038/nbt0402-370

Schoeberl, B., Pace, E. a, Fitzgerald, J. B., Harms, B. D., Xu, L., Nie, L., … Nielsen, U. B. (2009). Therapeutically targeting ErbB3: a key node in ligand-induced activation of the ErbB receptor-PI3K axis. *Science Signaling*, *2*(77), ra31. http://doi.org/10.1126/scisignal.2000352

Schubbert, S., Shannon, K., & Bollag, G. (2007). Hyperactive Ras in developmental disorders and cancer. *Nature Reviews. Cancer*, *7*(April), 295–308. http://doi.org/10.1038/nrc2109

Shayesteh, L., Lu, Y., Kuo, W. L., Baldocchi, R., Godfrey, T., Collins, C., … Gray, J. W. (1999). PIK3CA is implicated as an oncogene in ovarian cancer. *Nature Genetics*, *21*(january), 99–102. http://doi.org/10.1038/5042

Shmulevich, I., Dougherty, E. R., Kim, S., & Zhang, W. (2002). Probabilistic Boolean Networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics (Oxford, England)*, *18*(2), 261–274. http://doi.org/10.1093/bioinformatics/18.2.261

Shmulevich, I., Gluhovsky, I., Hashimoto, R. F., Dougherty, E. R., & Zhang, W. (2003). Steady-state analysis of genetic regulatory networks modelled by probabilistic Boolean networks. *Comparative and Functional Genomics*, *4*(6), 601–608. http://doi.org/10.1002/cfg.342

Su, M., Mei, Y., & Sinha, S. (2013). Role of the crosstalk between autophagy and apoptosis in cancer. *Journal of Oncology*, *2013*. http://doi.org/10.1155/2013/102735

Sullivan, J. C., Kalaitzidis, D., Gilmore, T. D., & Finnerty, J. R. (2007). Rel homology domain-containing transcription factors in the cnidarian Nematostella vectensis. *Development Genes and Evolution*, *217*(1), 63–72. http://doi.org/10.1007/s00427-006-0111-6

Sun, M., Wang, G., Paciga, JE., Feldman, RI., Yuan, ZQ., ..., Cheng, JQ. (2001). AKT1/PKBalpha kinase is frequently elevanted in human cancers and its consitutive action is required for oncogenic transformation in NIH3T3 cells. *The American Journal of Pathology*, *159*(2), 431–437.

Swinney, D. C., & Anthony, J. (2011). How were new medicines discovered? *Nature Reviews. Drug Discovery*, *10*(7), 507–519. http://doi.org/10.1038/nrd3480

Tan, C. B., Zhi, W., Shahzad, G., & Mustacchia, P. (2012). Gastrointestinal Stromal Tumors: A Review of Case Reports, Diagnosis, Treatment, and Future Directions. *ISRN Gastroenterology*, *2012*, 1–16. http://doi.org/10.5402/2012/595968

Tinoco, G., Warsch, S., Glück, S., Avancha, K., & Montero, A. J. (2013). Treating breast cancer in the 21st century: Emerging biological therapies. *Journal of Cancer*, *4*(2), 117–132. http://doi.org/10.7150/jca.4925

Trairatphisan, P., Mizera, A., Pang, J., Tantar, A. A., & Sauter, T. (2014). optPBN: An optimisation toolbox for probabilistic Boolean networks. *PLoS ONE*, *9*(7). http://doi.org/10.1371/journal.pone.0098001

Trairatphisan, P., Mizera, A., Pang, J., Tantar, A. A., Schneider, J., & Sauter, T. (2013). Recent development and biomedical applications of probabilistic Boolean networks. *Cell Communication and Signaling : CCS*, *11*(1), 46. http://doi.org/10.1186/1478-811X-11-46

Tsang, C.-M., Wong, C.-K., Ip, W.-K., & Lam, C. W.-K. (2005). Synergistic effect of SCF and TNF-alpha on the up-regulation of cell-surface expression of ICAM-1 on human leukemic mast cell line (HMC)-1 cells. *Journal of Leukocyte Biology*, *78*(1), 239–247. http://doi.org/10.1189/jlb.0704400

Turing, a. M. (1952). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London*, *237*(641), 37–72. http://doi.org/10.1016/S0092-8240(05)80008-4

Vander Heiden, M. G., Cantley, L. C., & Thompson, C. B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science (New York, N.Y.)*, *324*(May), 1029–1033. http://doi.org/10.1126/science.1160809

Vermeulen, K., Van Bockstaele, D. R., & Berneman, Z. N. (2003). The cell cycle: A review of regulation, deregulation and therapeutic targets in cancer. *Cell Proliferation*, *36*(3), 131–149. http://doi.org/10.1046/j.1365-2184.2003.00266.x

Vivanco, I., & Sawyers, C. L. (2002). The phosphatidylinositol 3-Kinase AKT pathway in human cancer. *Nature Reviews. Cancer*, *2*(7), 489–501. http://doi.org/10.1038/nrc839

Wang, C.-C., Cirit, M., & Haugh, J. M. (2009). PI3K-dependent cross-talk interactions converge with Ras as quantifiable inputs integrated by Erk. *Molecular Systems Biology*, *5*(246), 246. http://doi.org/10.1038/msb.2009.4

Wang, R.-S., Saadatpour, A., & Albert, R. (2012). Boolean modeling in systems biology: an overview of methodology and applications. *Physical Biology*, *9*(5), 055001. http://doi.org/10.1088/1478-3975/9/5/055001

Wang, Z., & Moran, M. F. (2002). Phospholipase C-gamma1: a phospholipase and guanine nucleotide exchange factor. *Molecular Interventions*, *2*(6), 352–355,338. http://doi.org/10.1124/mi.2.6.352

Warburg, O. (1956). On the Origin of Cancer Cells. *Science*, *123*(3191), 309–14. http://doi.org/10.1126/science.123.3191.309

Wittmann, D. M., Krumsiek, J., Saez-Rodriguez, J., Lauffenburger, D. a, Klamt, S., & Theis, F. J. (2009). Transforming Boolean models to continuous models: methodology and application to T-cell receptor signaling. *BMC Systems Biology*, *3*, 98. http://doi.org/10.1186/1752-0509-3-98

Xia, Z., & Storm, D. R. (2012). Role of signal transduction crosstalk between adenylyl cyclase and MAP kinase in hippocampus-dependent memory. *Learning & Memory*, *19*(9), 369–374. http://doi.org/10.1101/lm.027128.112

Xie, T., Xia, Z., Zhang, N., Gong, W., & Huang, S. (2010). Constitutive NF-κB activity regulates the expression of VEGF and IL-8 and tumor angiogenesis of human glioblastoma. *Oncology Reports*, *23*, 725–732. http://doi.org/10.3892/or

Yang, M., Li, R., & Chu, T. (2014). A New Method and Application for Controlling the Steady-state Probability Distributions of Probabilistic Boolean Networks. In *IEEE Congress on Evolutionary Computation (CEC)* (pp. 1490–1495). http://doi.org/10.1109/CEC.2014.6900436

Yuan, Q., Trairatphisan, P., Pang, J., Mauw, S., Wiesinger, M., & Sauter, T. (2012). Probabilistic Model Checking of the PDGF Signaling Pathway. In *Transactions on Computational Systems Biology XIV* (pp. 151–180). Retrieved from http://www.springerlink.com/index/N01M1GR5H55615V7.pdf

Zhao, Y., & Cheng, D. Z. (2014). On controllability and stabilizability of probabilistic Boolean control networks. *Science China Information Sciences*, *57*(1), 1–14. http://doi.org/10.1007/s11432-013-4851-4

Zhu, P., & Han, J. (2014). Stochastic multiple-valued gene networks. *IEEE Transactions on Biomedical Circuits and Systems*, *8*(1), 42–53. http://doi.org/10.1109/TBCAS.2013.2291398

Zhu, P., Liang, J., & Han, J. (2014). Gene perturbation and intervention in context-sensitive stochastic Boolean networks. *BMC Systems Biology*, *8*(1), 60. http://doi.org/10.1186/1752-0509-8-60

Zimmermann, S., & Moelling, K. (1999). Phosphorylation and regulation of Raf by Akt (protein kinase B). *Science (New York, N.Y.)*, *286*(5445), 1741–1744. http://doi.org/10.1126/science.286.5445.1741

Chapter 7

APPENDICES

7.1 Selected supplementary information of Chapter 4.1

Results from the optPBN pipeline for four case studies

**Note: Disc. = Discrete mode, Cont. = Continuous mode, Ext. = Extended model structure, Init. = Initial model tructure iter. = iterations, Apo. = Apotosis, Delta = difference between steady-state distributions and measurement data nsteps = length of simulated two-state Markov chain trajectory used to estimate the steady-state probablities*

Stand-alone version (Optimiser: Particle swarm algorithm, PSO)

Case study (Model)	Approach	ergodic PBN and two-state Markov chain						
Model 1 (Disc.)	Pswarm iter.	r	Time (s)	nsteps	$C_2^{(NFkB)}$	Delta		
	1000	0.05	129	101	1 (&)	0		
		0.025	305	101	1 (&)	0		
		0.01	1578	382	1 (&)	0		
	5000	0.05	627	101	1 (&)	0		
		0.025	1093	101	1 (&)	0		
		0.01	5841	382	1 (&)	0		
Model 3 (Disc.)	Pswarm iter.	r	Time (s)	nsteps	$C_2^{(NFkB)}$	Delta	$C_1^{(ERK)}$, $C_3^{(ERK)}$	Delta
	1000	0.05	357	101	1 (&)	0	0.5 (|), 0.5 (Lt)	0
		0.025	2194	101	1 (&)	0	0.5 (|), 0.5 (Lt)	0
		0.01	5451	382	1 (&)	0	0.5 (|), 0.5 (Lt)	0
	5000	0.05	1761	101	1 (&)	0	0.5 (|), 0.5 (Lt)	0
		0.025	3011	101	1 (&)	0	0.5 (|), 0.5 (Lt)	0
		0.01	25064	382	1 (&)	0	0.5 (|), 0.5 (Lt)	0
Model 2 (Cont.)	Pswarm iter.	r	Time (s)	nsteps	$C_1^{(N3)}$, $C_3^{(N3)}$	Delta		
	1000	0.05	150	408	0.6045, 0.3955	0.0045		
		0.025	321	1514	0.6087, 0.3913	0.0087		
		0.01	1351	10612	0.6090, 0.3910	0.009		
	5000	0.05	659	438	0.6237, 0.3763	0.0237		
		0.025	1205	1444	0.6041, 0.3959	0.0041		
		0.01	7968	10357	0.5979, 0.4021	0.0021		
Model 3 (Cont.)	Pswarm iter.	r	Time (s)	nsteps	$C_1^{(NFkB)}$, $C_2^{(NFkB)}$	Delta	$C_1^{(ERK)}$, $C_2^{(ERK)}$	Delta
	1000	0.02	906	433	0.6934, 0.3066	0.0066	0.5477, 0.4523	0.0523
		0.01	1302	1861	0.6990, 0.3010	0.001	0.6122, 0.3878	0.0122
		0.005	8802	10534	0.6968, 0.3032	0.0032	0.5868, 0.4132	0.0132
	5000	0.02	4384	543	0.6992, 0.3008	0.0008	0.6532, 0.3468	0.0532
		0.01	6151	1861	0.7078, 0.2922	0.0078	0.6028, 0.3972	0.0028
		0.005	48747	10403	0.7016, 0.2984	0.0016	0.6046, 0.3954	0.0046
Model 4 (Cont./Ext.)	Pswarm iter.	r	Time (s)	nsteps	Simulated (Apo.)	Delta	Simulated (NFkB)	Delta
	1000	0.025	16783	376	0.0027 (vs 0)	0.0027	0.0053 (vs 0)	0.0053
				1505	0.4133 (vs 0.63)	0.2167	0.2970 (vs 0.32)	0.023
				1725	0.0017 (vs 0)	0.0017	0.5843 (vs 1)	0.4157
				1667	0.0012 (vs 0.1)	0.0988	0.6029 (vs 0.98)	0.3771
				815	0.7963 (vs 1)	0.2037	0.0025 (vs 0.26)	0.2575
				2029	0.2617 (vs 0.49)	0.2283	0.4421 (vs 0.46)	0.0179
				Sum Delta (2 nodes) = 1.8484 ; SSE (3 nodes) = 0.7527				
	5000	0.025	45503	457	0.0066 (vs 0)	0.0066	0.0022 (vs 0)	0.0022
				1346	0.6218 (vs 0.63)	0.0082	0.2117 (vs 0.32)	0.1083
				549	0 (vs 0)	0	0.9271 (vs 1)	0.0729
				1630	0.0006 (vs 0.1)	0.0994	0.7571 (vs 0.98)	0.2229
				101	0.9406 (vs 1)	0.0594	0.0099 (vs 0.26)	0.2501
				1819	0.4766 (vs 0.49)	0.0134	0.3090 (vs 0.46)	0.151
				Sum Delta (2 nodes) = 0.9944 ; SSE (3 nodes) = 0.2138				
	Cij	Best run		Cost ->	0.1939	Value		
	$C_1^{(NFkB)}$	NF_kB = (~I_kBa | ~I_kBe)				0.7982		
	$C_2^{(NFkB)}$	NF_kB = (~I_kBa | ~I_kBe) | (C8a | C8a_2)				0.2018		
	$C_1^{(complex2)}$	complex2 = (complex1&FADD)				0.7517		
	$C_2^{(complex2)}$	complex2 = (complex1&FADD&RIP_deubi)				0.2483		
	Cij	Mean (500/5000 runs)		Cost ->	0.2681	Mean	S.D.	
	$C_1^{(NFkB)}$	NF_kB = (~I_kBa | ~I_kBe)				0.8036	0.0433	
	$C_2^{(NFkB)}$	NF_kB = (~I_kBa | ~I_kBe) | (C8a | C8a_2)				0.1964	0.0433	
	$C_1^{(complex2)}$	complex2 = (complex1&FADD)				0.669	0.0739	
	$C_2^{(complex2)}$	complex2 = (complex1&FADD&RIP_deubi)				0.331	0.0739	

Grid-based version (Optimisers: Differential evolution [DE] + Evolutionary algorithm [EA], 80 cores)

Run-time analysis (Table 4)

Case study (model)	G5K iter.	r	Stand-alone (1 core)	Grid-based (80 cores)	Improvement (folds)
Model 1	1000	0.025	305s (5m5s)	8s	38.1
	5000		1093s (18m13s)	16s	68.3
Model 2	1000		321s (5m21s)	9s	35.7
	5000		1205s (20m5s)	16s	75.3
Model 3 (Cont.)	1000		1302s (21m42s)	18s	72.3
	5000		6151s (1h42m31s)	39s	157.7
Model 4 (Cont./Ext.)	1000		16783s (4h39m43s)	99s (1m39s)	169.5
	5000		45503s (12h38m23s)	259s (4m19s)	175.7

Optimisation results of case study 4

Case study (model)	G5K iter.							
Model 4 (Cont./Ext.)	15000	**r**	**Time (s)**	**nsteps**	**Simulated (Apo.)**	**Delta**	**Simulated (NFkB)**	**Delta**
		0.025	634	880	0.0054 (vs 0)	0.0054	0.0027 (vs 0)	0.0027
				919	0.717 (vs 0.63)	0.087	0.1501 (vs 0.32)	0.1699
				817	0.0063 (vs 0)	0.0063	0.8481 (vs 1)	0.1519
				1667	0.0018 (vs 0.1)	0.0982	0.8723 (vs 0.98)	0.1077
				119	0.9868 (vs 1)	0.0132	0.0033 (vs 0.26)	0.2567
				1366	0.4914 (vs 0.49)	0.0014	0.2932 (vs 0.46)	0.1668
			80 cores	**Sum Delta (2 nodes) = 1.0672 ; SSE (3 nodes) = 0.1987**				
	Cij	**Best run**		Cost ->	0.176	Value		
	$C_1^{(NFkB)}$	NF_kB = (~I_kBa | ~I_kBe)				0.8544		
	$C_2^{(NFkB)}$	NF_kB = (~I_kBa | ~I_kBe) | (C8a | C8a_2)				0.1456		
	$C_1^{(complex2)}$	complex2 = (complex1&FADD)				1		
	$C_2^{(complex2)}$	complex2 = (complex1&FADD&RIP_deubi)				0		
	Cij	Mean (500/7500 runs)		Cost ->	0.1994	Mean	S.D.	
	$C_1^{(NFkB)}$	NF_kB = (~I_kBa | ~I_kBe)				0.8354	0.0209	
	$C_2^{(NFkB)}$	NF_kB = (~I_kBa | ~I_kBe) | (C8a | C8a_2)				0.1646	0.0209	
	$C_1^{(complex2)}$	complex2 = (complex1&FADD)				0.848	0.1105	
	$C_2^{(complex2)}$	complex2 = (complex1&FADD&RIP_deubi)				0.152	0.1105	
Model 4 (Cont./Init.)	15000	**r**	**Time (s)**	**nsteps**	**Simulated (Apo.)**	**Delta**	**Simulated (NFkB)**	**Delta**
		0.025	580	880	0 (vs 0)	0	0.0168 (vs 0)	0.0168
				608	0.9208 (vs 0.63)	0.2908	0.0198 (vs 0.32)	0.3002
				815	0.0028 (vs 0)	0.0028	0.8873 (vs 1)	0.1127
				1723	0.002 (vs 0.1)	0.098	0.7889 (vs 0.98)	0.1911
				101	0.992 (vs 1)	0.008	0.0023 (vs 0.26)	0.2577
				1494	0.4681 (vs 0.49)	0.0219	0.3083 (vs 0.46)	0.1517
			80 cores	**Sum Delta (2 nodes) = 1.4517 ; SSE (3 nodes) = 0.3275**				
	Cij	**Best run**		Cost ->	0.2668	Value		
	$C_1^{(complex2)}$	complex2 = (complex1&FADD)				0.7524		
	$C_2^{(complex2)}$	complex2 = (complex1&FADD&RIP_deubi)				0.2476		
	Cij	Mean (500/7500 runs)		Cost ->	0.2869	Mean	S.D.	
	$C_1^{(complex2)}$	complex2 = (complex1&FADD)				0.8351	0.1114	
	$C_2^{(complex2)}$	complex2 = (complex1&FADD&RIP_deubi)				0.1649	0.1114	

Pooled Grid-based results from 3 separated runs (Extended structure of Model 4)

	Cij	Mean (**500*3** runs)	Cost ->	0.2043	Mean	S.D.
Pooled Model 4 (Ext)	$C_1^{(NFkB)}$	NF_kB = (~I_kBa | ~I_kBe)			0.831	0.0304
	$C_2^{(NFkB)}$	NF_kB = (~I_kBa | ~I_kBe) | (C8a | C8a_2)			0.169	0.0304
	$C_1^{(complex2)}$	complex2 = (complex1&FADD)			0.7899	0.1716
	$C_2^{(complex2)}$	complex2 = (complex1&FADD&RIP_deubi)			0.2101	0.1716

7.2 Selected supplementary information of Chapter 4.2

Supporting Information File S1:

Model description of the PDGF signalling network in PBN format

---

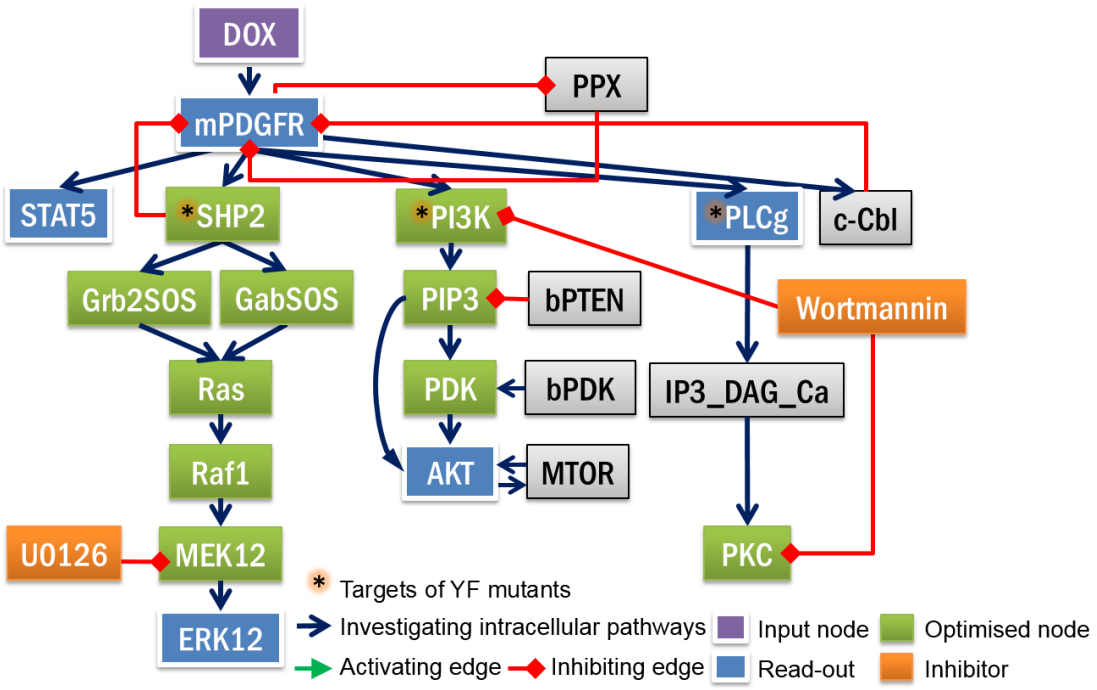
This document provides a detailed description of the PBN models of PDGF signalling which are presented in the manuscript. The model descriptions are separated into 3 parts as follows:

1) The core model structure with major intracellular signalling pathways
2) The integration of crosstalk interactions which were proposed in literature
3) The refined model structure after integrating data from additional experiment

Please note that we applied the annotation '->' to depict an activating interaction and '-|' to depict an inhibitory interaction.

Part 1: The core model structure with major intracellular signalling pathways

We built a core model structure of PDGF signalling comprising PDGFRα and its regulatory mechanisms together with the major downstream signalling pathways including MAPK, PI3K/AKT/mTOR, PLCγ/PKC pathways as well as STAT5. The core model topology is shown below:



From the top, we depicted the induction of PDGFRα-transgene by doxycycline (DOX) in our cell system. The regulatory mechanisms of mutated PDGFRα (mPDGFR, represented as 'PDGFR' in the

model) are governed by cytoplasmic phosphatase (PPX), protein tyrosine phosphatase shp2 (SHP2) and proteosomal degradation via ubiquitination promoted by c-Cbl (cCbl). The network topology of downstream signalling pathways comprises as follows:

- MAPK pathway (SHP2 -> Grb2SOS/GabSOS -> Ras -> Raf1 -> MEK1,2 (MEK12) -> ERK1,2 (ERK12))
- PI3K-AKT-mTOR pathway (PI3K -> PIP3 -> PDK -> AKT <-> mTOR) with basal activity of PTEN (bPTEN) inhibiting PIP3 and basal activity of PDK (bPDK) activating PDK. Note that PIP3 can also directly activate Akt.
- PLCg-PKC pathway (PLCg -> IP3_DAG_Ca -> PKC)
- STAT5

According to this model topology, we formulated the model descriptions with Boolean rules as follows:

```
=== PDGFR activation and regulatory mechanisms ===
% DOX:              rules=[rules; {'DOX = 1','1','C'}];
% PPX:              rules=[rules; {'PPX = ~PDGFR','1','C'}];
% cCbl:             rules=[rules; {'cCbl = PDGFR','1','C'}];
% PDGFR:            rules=[rules; {'PDGFR = DOX','1','D'}];
                    rules=[rules; {'PDGFR = DOX & ~cCbl','1','D'}];
                    rules=[rules; {'PDGFR = DOX & ~PPX','1','D'}];
                    rules=[rules; {'PDGFR = DOX & ~SHP2','1','D'}];

=== MAPK pathway ===
% SHP2:             rules=[rules; {'SHP2 = PDGFR','1','C'}];
% Grb2SOS:          rules=[rules; {'Grb2SOS = SHP2','1','C'}];
% GabSOS:           rules=[rules; {'GabSOS = SHP2','1','C'}];
% Ras:              rules=[rules; {'Ras = Grb2SOS | GabSOS','1','C'}];
% Raf1:             rules=[rules; {'Raf1 = Ras','1','C'}];
% MEK12:            rules=[rules; {'MEK12 = MEK12_induce','1','C'}];
% MEK12_induce:     rules=[rules; {'MEK12_induce = Raf1','1','C'}];
% ERK12:            rules=[rules; {'ERK12 = MEK12','1','C'}];

=== PI3K-AKT-mTOR pathway ===
% PI3K_PDGFR:       rules=[rules; {'PI3K_PDGFR = PDGFR','1','C'}];
% prePI3K:          rules=[rules; {'prePI3K = PI3K_PDGFR','1','C'}];
% PI3K:             rules=[rules; {'PI3K = prePI3K','1','C'}];
% bPTEN:            rules=[rules; {'bPTEN = 1','1','C'}];
% PIP3:             rules=[rules; {'PIP3 = PI3K','1','D'}];
                    rules=[rules; {'PIP3 = PI3K & ~bPTEN','1','D'}];
% bPDK:             rules=[rules; {'bPDK = 1','1','C'}];
% PDK:              rules=[rules; {'PDK = PIP3','1','D'}];
                    rules=[rules; {'PDK = bPDK','1','D'}];
% Akt:              rules=[rules; {'Akt = PIP3','1','D'}];
                    rules=[rules; {'Akt = PDK','1','D'}];
                    rules=[rules; {'Akt = MTOR','1','D'}];
% MTOR:             rules=[rules; {'MTOR = Akt','1','C'}];

=== PLCg-PKC pathway ===
% PLCg:             rules=[rules; {'PLCg = PDGFR','1','C'}];
% IP3_CaIon_DAG:    rules=[rules; {'IP3_CaIon_DAG = PLCg','1','C'}];
```

```
% PKC_induce:        rules=[rules; {'PKC_induce = IP3_CaIon_DAG','1','C'}];
% PKC:               rules=[rules; {'PKC = PKC_induce','1','C'}];

  === STAT5 ===
% STAT5:             rules=[rules; {'STAT5 = PDGFR','1','C'}];
```

Based on these assigned Boolean rules, there are a few important observations. First, the flags 'C' and 'D' at the end of the Boolean rules are defined to correlate with the terms 'Constant' and 'Default' respectively. Once the flag 'C' is used, it means that the selection probability of that Boolean rule will be constant and the corresponding node (molecule) will not be optimised so all rules for this node need to have the flag 'C' then. This flag is exclusively applied for the nodes with only one input as the probability of activating or inhibiting such nodes have to be 1. In parallel, when a node has more than one incoming interaction, multiple Boolean rules can be assigned for such node and the respective node is typically optimised. The flag 'D' can be assigned throughout multiple Boolean rules on a single node to give a full boundary of the selection probability from 0 to 1 during the optimisation.[1] With this flag assigned, each regulatory mechanism as coded in different Boolean rule will have an equal chance to dominate the other. In general, this assignment is defined once the influence from each interaction is unclear. For instance, the activation of Akt can be mediated by PIP3, PDK and mTOR but there is no prior information nor assumption whether one specific source of signal will be stronger than the others.

Another observation that should be noted is some molecules, i.e. MEK1,2, PI3K and PKC have intermediate nodes. These nodes represent the intermediate steps of signal transduction processes where multiple activating and inhibiting interactions take place. On the last observation on the inhibitory mechanisms, we applied 2 types of single perturbation to the system, i.e. additional YF point mutations to abrogate recruitment sites on the PDGFRα (dMAPK and dPI3K) and signalling inhibitors (Wortmannin and U0126) in the initial experiment. Based on our experimental observations and literature information, we found that the abrogation of tyrosine residue Y720F in dMAPK mutant also affects the binding of pPLCg to PDGFRα while Wortmannin has an additional off-target on PKC apart from inhibiting PI3K. To account for these inhibitory mechanisms without adding more nodes, we repeated the same Boolean rule for the activating interaction once and we set the second Boolean rule as 0 once the targeted molecule is inhibited. To illustrate our methodology, we provide an example of Boolean rules for the D842V-PDGFRα mutant without YF point mutation or signalling inhibitor (DV-WT) compared to the Boolean rules of D842V-PDGFRα-dMAPK mutant and Wortmannin treatment (DV-dMAPK-Wort.) as follows:

```
=== DV-WT ===
% SHP2:             rules=[rules; {'SHP2 = PDGFR','1','D'}];
                    rules=[rules; {'SHP2 = PDGFR','1','D'}];
% MEK12:            rules=[rules; {'MEK12 = MEK12_induce','1','D'}];
                    rules=[rules; {'MEK12 = MEK12_induce','1','D'}];
% PI3K_PDGFR:       rules=[rules; {'PI3K_PDGFR = PDGFR','1','D'}];
                    rules=[rules; {'PI3K_PDGFR = PDGFR','1','D'}];
% prePI3K:          rules=[rules; {'prePI3K = PI3K_PDGFR','1','C'}];
```

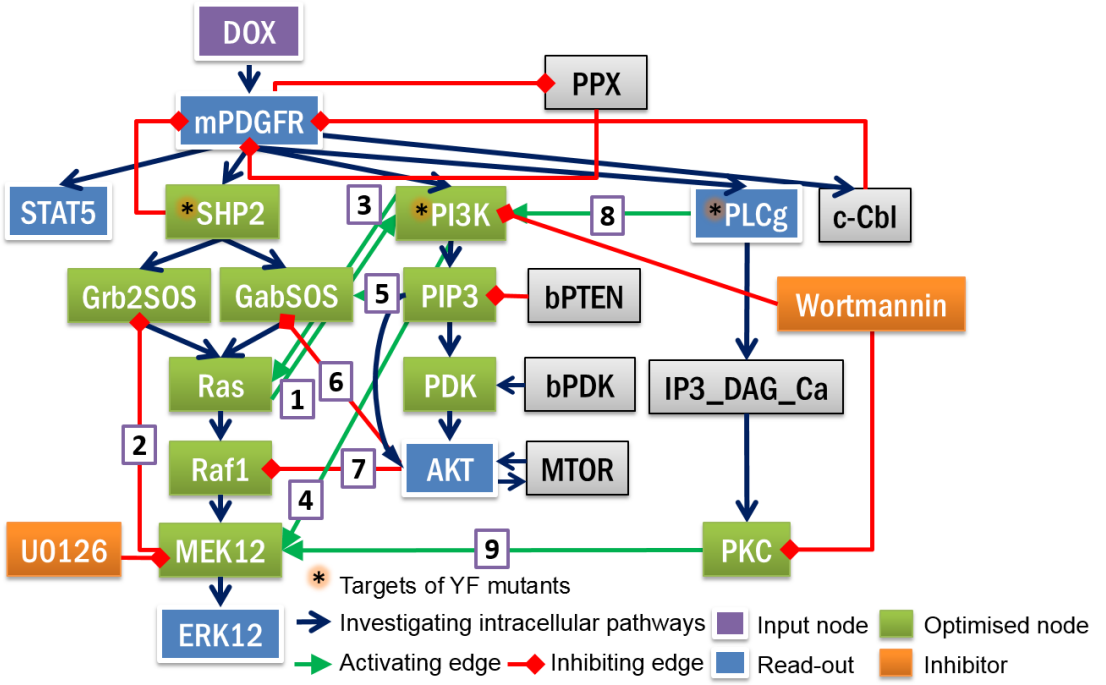
[1] If the sum of optimised selection probabilities from multiple Boolean rules is larger than 1, the selection probability of each Boolean rule will be normalised to the sum of all probabilities in order to ensure that the sum of the selection probabilities will be 1.

```
% PI3K:          rules=[rules; {'PI3K = prePI3K','1','D'}];
                 rules=[rules; {'PI3K = prePI3K','1','D'}];
% PLCg:          rules=[rules; {'PLCg = PDGFR','1','D'}];
                 rules=[rules; {'PLCg = PDGFR','1','D'}];
% PKC:           rules=[rules; {'PKC = PKC_induce','1','D'}];
                 rules=[rules; {'PKC = PKC_induce','1','D'}];

=== DV-dMAPK-Wort. ===

% SHP2:          rules=[rules; {'SHP2 = PDGFR','1','D'}];
                 rules=[rules; {'SHP2 = 0','1','D'}]; % effect of dMAPK
% MEK12:         rules=[rules; {'MEK12 = MEK12_induce','1','D'}];
                 rules=[rules; {'MEK12 = MEK12_induce','1','D'}];
% PI3K_PDGFR:    rules=[rules; {'PI3K_PDGFR = PDGFR','1','D'}];
                 rules=[rules; {'PI3K_PDGFR = PDGFR','1','D'}];
% prePI3K:       rules=[rules; {'prePI3K = PI3K_PDGFR','1','C'}];
% PI3K:          rules=[rules; {'PI3K = prePI3K','1','D'}];
                 rules=[rules; {'PI3K = 0','1','D'}]; % effect of Wortmannin
% PLCg:          rules=[rules; {'PLCg = PDGFR','1','D'}];
                 rules=[rules; {'PLCg = 0','1','D'}]; % effect of dMAPK
% PKC:           rules=[rules; {'PKC = PKC_induce','1','D'}];
                 rules=[rules; {'PKC = 0','1','D'}]; % effect of Wortmannin
```

As demonstrated, the second rules of SHP2 and PLCg were changed to 0 to account for the inhibitory effect of dMAPK while the second rules of PI3K and PKC were also set to 0 to take the inhibitory effect of Wortmannin into consideration. A similar assignment was applied to the second rule of PI3K_PDGFR in order to account for the dPI3K effect, and it was also applied to the second rule of MEK12 in order to account for U0126 inhibition. With the methodology for the assignment of Boolean rules as demonstrated, we were able to generate the model structures of PDGF signalling which could represent the six experimental conditions within the training dataset and we subsequently applied them to perform model fitting accordingly.

Part 2: The integration of crosstalk interactions which were proposed in literature

After we tried to fit the core model topology to the training dataset, we discovered that the model fitted to the data relatively well but there were still a few data points where could not be fitted. We hypothesise that a certain number of crosstalk interactions is required to explain all data points. Therefore, we built 9 additional model variants by adding the crosstalk interactions which were proposed in the literature to the core model topology one-at-a-time and we re-performed the optimisation on the model variants to the training dataset. The types and targets of the crosstalk interactions in the 9 model variants are shown as follows:



In terms of the implementation in the PBN model, one interaction was added to the core model topology (Part 1) for each model variant. For instance, the model variant with crosstalk number 9 (PKC -> MEK12) has the following Boolean rules:

```
% MEK12_induce:    rules=[rules; {'MEK12_induce = Raf1','1','H'}];
                   rules=[rules; {'MEK12_induce = PKC','1','L'}];
```

Note that another set of flags 'H' and 'L' were used. As previous mentioned in the manuscript, we consider that the signals flow through canonical pathways stronger than flowing pass crosstalk interactions. Hence, we assign the flag 'H' (High) for the canonical interaction Raf1 -> MEK12 and the flag 'L' (Low) for the crosstalk interaction PKC -> MEK12. This assignment ensure that the boundary of the interaction(s) with the flag 'H' will always have higher selection probability compared to the interaction(s) with the flag 'L'.

Among the pool of all model variants, one of the complex model variants that should be mentioned is the model variant with crosstalk number 1 (Ras -> PI3K). In this model variant, we introduce a new node called 'PI3K_crosstalk' coded as "rules=[rules; {'PI3K_crosstalk = Ras','1','C'}];" to account for this crosstalk interaction. Previously, we had the node 'PI3K_PDGFR' which represents PI3K signal after the activation by PDGFR. To integrate the signals, the successive node 'prePI3K' was assigned to represent PI3K signals which receive the inputs both from PDGFR and from crosstalk interaction(s) as depicted in following Boolean rules:

```
% prePI3K:    rules=[rules; {'prePI3K = PI3K_PDGFR','1','H'}];
              rules=[rules; {'prePI3K = PI3K_crosstalk','1','L'}];
```

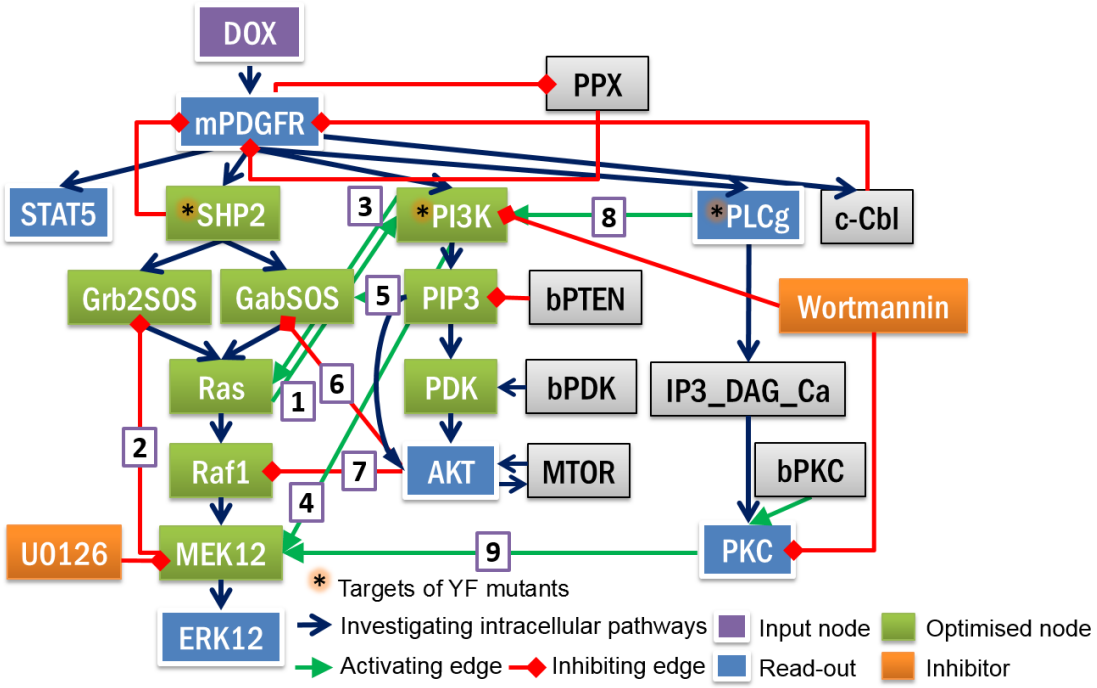
Note that the information on canonical pathway versus crosstalk interaction is also applied here with the flags 'H' and 'L'. Subsequently, the integrated signals of PI3K in 'prePI3K' node can further be inhibited by Wortmannin by setting the second rule of 'PI3K' node to 0 as explained above.

Part 3: The refined model structure after integrating data from the additional experiment

After we performed the additional experiment (Figure 3) and we found that PKC activity is independent of the activation from PDGFRα, we added the node 'bPKC' to account for the basal activity of PKC with the following Boolean rules:

```
% PKC_induce:        rules=[rules; {'PKC_induce = IP3_CaIon_DAG','1','D'}];
                     rules=[rules; {'PKC_induce = bPKC','1','D'}];
```

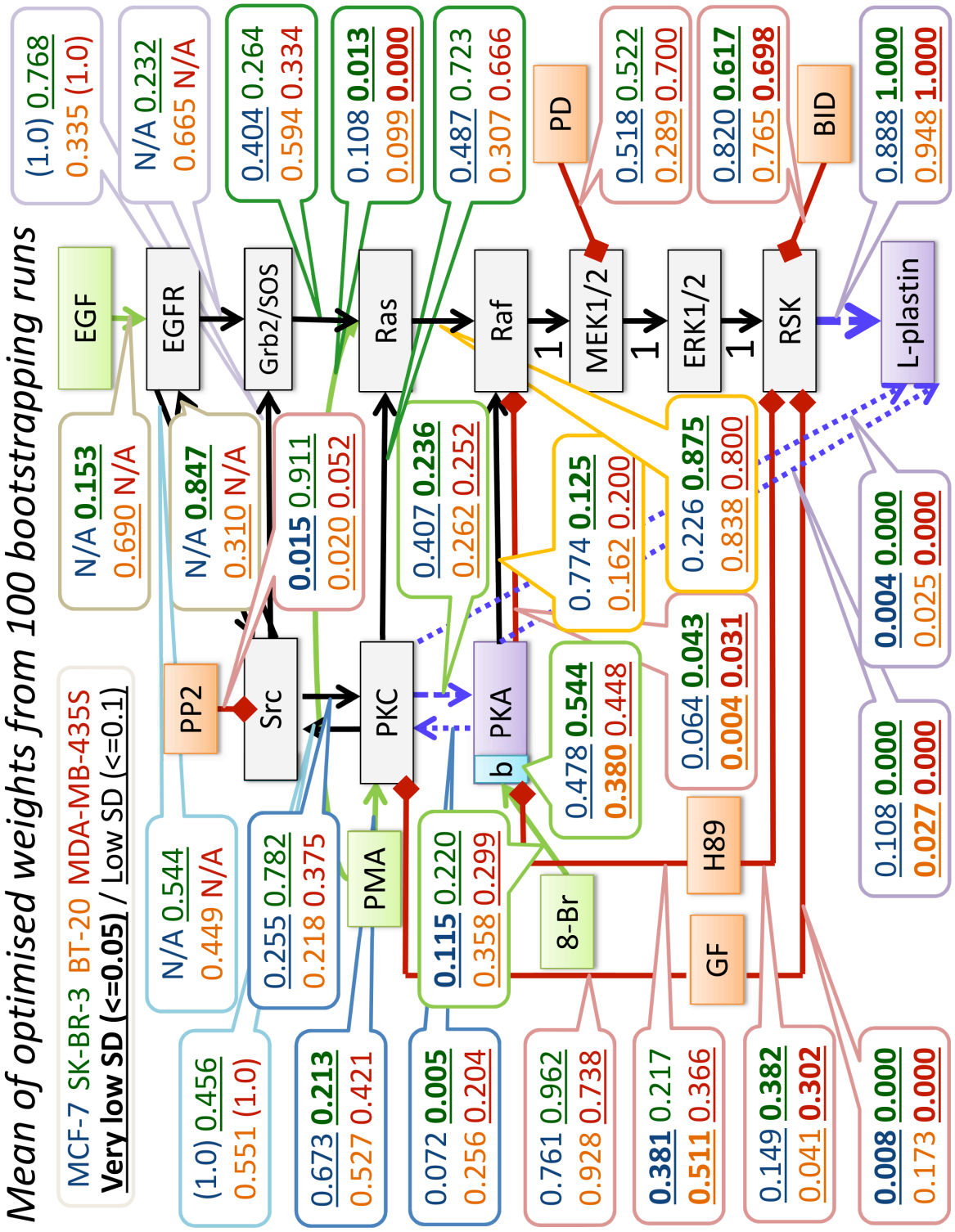
The updated model topology with the additional information on two new PKC inhibitors (GF109 and Gö6976) is shown in the figure below:



Note that even if we performed an additional experiment with the two new PKC inhibitors, we only took the information from the experimental conditions which overlapped with the ones in the training dataset for modelling. With this regard, we only included pPKC substrates data on negative control, positive control (DV-WT) and DV-WT-Wortmannin conditions into the training dataset for final model refinement.

After we fitted the refined model to the updated training dataset, we also performed an *in silico* analysis to investigate the importance of crosstalk interactions which were described in Step 2. This analysis revealed that only crosstalk interactions number 3 (PI3K -> Ras) or number 4 (PI3K -> MEK1,2) are important to fit the training dataset. Then, we proceeded with the final PBN model, i.e. the refined model plus the crosstalk interaction number 4 (PI3K -> MEK1,2) to evaluate the predictive power of the model. We found that the final model could predict the signalling profiles in the combined perturbation conditions within the validation dataset accurately. Hence, we concluded that our final PBN model is highly predictive.

7.3 Selected supplementary information of Chapter 4.3



7.4 Joint publication: Probabilistic model checking of the PDGF signalling pathway (Yuan et al., 2012)

Preface: In this research article, we applied the probabilistic model checker PRISM to analyse the platelet-derived growth factor (PDGF) signalling pathway. We converted a literature-derived PDGF signalling network from an interaction graph into a reaction-based ODE model using the law of mass action and we shortly discuss on the translation from kinetic rates into stochastic rates. Parameter sensitivity analysis was performed to ensure that the qualitative system dynamics are robust against perturbations on parameter values.

Once biochemical interactions and stochastic parameters were derived, they were represented as a continuous-time Markov chain (CTMC) model which is compatible to be analysed in PRISM, a probabilistic model checker for stochastic models. Several research objectives were defined to be analysed thereby, i.e. to analyse the dynamics of PDGF-induced signalling, to analyse the influence of crosstalk interactions, and to analyse the importance of individual reactions. These objectives were practically specified as observing properties in PRISM where several properties were subsequently derived, e.g. the dynamical probabilities of certain molecules to be activated over certain time points or the long-run (steady-state) probabilities of certain molecules to be in an active state. The 'node' and 'edge' knock-out model variants were generated and analysed where the changing activities of downstream signalling molecules (MEK1,2 & AKT) were observed.

Lastly, we also compared the dynamical and steady-state results from stochastic verification versus ODE simulation. Surprisingly, we found that the long-run probability of AKT can be largely differed between the two approaches in an isolated reciprocal positive feedback structure. We identified that the state values of molecules in the ODE framework will always be multiplied up within a positive-feedback loop which leads to very high steady-state values even if the parameter values are small. In contrast, the long-run probabilities in the CTMC model are proportionally reduced according to the stochastic parameter values as analysed by PRISM (see Figure 14 in the article). This finding highlights the importance of choosing an appropriate modelling framework which should correspond to the observations on the biological systems of interest.

Remark: The building and analysis of the ODE-based model and the comparison between the results from CTMC and ODE-based models were performed by P. Trairatphisan.

Probabilistic Model Checking of the PDGF Signaling Pathway*

Qixia Yuan[1,2], Panuwat Trairatphisan[3], Jun Pang[1]**,
Sjouke Mauw[1], Monique Wiesinger[3], and Thomas Sauter[3]

[1] Computer Science and Communications, University of Luxembourg, Luxembourg
[2] School of Computer Science and Technology, Shandong University, China
[3] Life Sciences Research Unit, University of Luxembourg, Luxembourg

Abstract. In this paper, we apply the probabilistic symbolic model checker PRISM to the analysis of a biological system – the Platelet-Derived Growth Factor (PDGF) signaling pathway, demonstrating in detail how this pathway can be analyzed in PRISM. Moreover, we compare the results from verification and ODE simulation on the PDGF pathway and demonstrate by examples the influence of model structure, parameter values and pathway length on the two analysis methods.

1 Introduction

Biological systems consist of components, which interact to influence each other and therefore the whole system's behavior. The field of systems biology aims to understand such complex interactions. Due to the similarity between biological systems and complex distributed/reactive systems studied in computer science [2], modeling and analyzing techniques developed in the field of formal methods can be applied to biological systems as well [3]. Due to efficient verification techniques, formal methods can analyze large systems exhibiting complex behaviors – this process is typically supported by automatic computer tools. This potentially gives formal methods an advantage, as *in silico* experiments are much easier to perform than *in vitro* experiments for the aim of analyzing and understanding biological systems. During the last decade, there has been a rapid and successful development in applying formal methods to systems biology – new formalisms are developed for systems biology to create models for biological phenomena, new algorithms and tools are specially designed and tailored for the analysis of such models (e.g., see [4–6]).

In this paper, we explore the usage of model checking for biological systems. Model checking is referred to as the automatic process of checking whether a system model satisfies a given specification (expressed as a temporal logic formula), by exhaustively exploring all possible executions of the system. This differs from simulation-based techniques, which only study a subset of the executions. More specifically, we focus on the probabilistic (or stochastic) model checking approach [7, 8], first introduced by Hart,

* An extended abstract appears in the proceedings of CompMod 2011 [1]. The first two authors made equal contributions to this work.

** Corresponding author.

Sharir and Pnueli [9], as biological systems usually have complicated stochastic behaviors. This technique is well-established and widely used for ascertaining the correctness of real-life systems, including distributed systems and communication protocols. In probabilistic (or stochastic) model checking, systems are normally represented by Markov chains or Markov decision processes. Properties of the models are expressed in quantitative extensions of temporal logics. In the literature, depending on the models used, usually probabilistic model checking has its focus on discrete-time Markov chains (DTMCs), while stochastic model checking deals with continuous-time Markov chains (CTMCs). Stochastic verification, in particular, has gained notable success in analyzing probabilistic systems including biological signaling pathways (e.g., see [10, 11]).

The stochasticity which occurs in biological signaling pathways can considerably affect the changes of biological processes. For instance, the stochasticity of initial conditions of caspases enzymes in the separatrix region can influence the cells to escape or enter apoptotic process [12]. In this paper, we use the probabilistic model checker PRISM [11] to yield a better understanding of the Platelet-Derived Growth Factor (PDGF) signaling pathway. PDGF, described approximately 30 years ago as a major mitogenic component of whole blood [13], is a growth factor that regulates cell growth and division. It promotes angiogenesis and also preserves vascular integrity through the recruitment of pericytes to endothelial tubes. Clinical studies reveal that aberrant expression of PDGF and its receptor is often associated with a variety of disorders such as atherosclerosis, fibroproliferative diseases and most importantly, neoplasia [13]. Deregulation of the PDGF signaling pathway plays a critical role in the development of many types of human diseases such as gastrointestinal stromal tumor and hypereosinophilic syndrome [14, 15, 13, 16, 17]. Based on intensive literature review, we have built the PDGF signal transduction model in ODE (Ordinary Differential Equation) format. The essential part of the PDGF signaling pathway contains the coupling of PDGF ligand to its receptor `PDGFR`, the negative regulatory mechanism on `PDGFR` and the activation of two main downstream signaling pathways, i.e., MAPK (Mitogen-Activated Protein Kinase) and PI3K/Akt pathways. In addition, there also exist positive and negative crosstalk interactions between different downstream signaling pathways (more details on the PDGF signaling pathway can be found in Sect. 3). In our study of the PDGF pathway, there are three main goals: (1) to analyze the dynamics of PDGF induced signaling, (2) to analyze the influence of the crosstalk reactions and (3) to analyze the importance of individual reactions/molecules on downstream signaling molecules. The first two can be used to check whether the constructed signaling pathway is consistent with respect to biological data, while the last one can lead us to some prediction. We have achieved these goals by stochastic verification using PRISM.

Moreover, we present the differences of the results obtained from ODE simulation and stochastic verification on the PDGF pathway, and demonstrate by examples the influence of model structure, parameter values and pathway length on the two analysis methods. In particular, we show that the two methods can predict the results differently, especially when parameter values are small.

Related work. There exists a large body of work on applying formal techniques to the analysis of biological systems. We focus on the use of PRISM and probabilistic model

checking in the literature, and other studies on the modeling and analysis of the PDGF pathway.

Calder et al. [18] perform a case study on the the RKIP inhibited ERK pathway using PRISM. Interestingly, they present a result stating that with a small number of molecules simulation results of their stochastic CTMC model and the corresponding ODE model are comparable. In this paper, we show that this result holds for the PDGF signaling pathway even with only one instance for each molecule (see Sect. 6). In [19], PRISM is used to study the MAPK cascade where a small subset of the MAPK pathway was modeled. The authors explain how the biological pathway can be modeled in PRISM and how this enables the analysis of a set of quantitative properties. In principle, these studies are correlated to our work as both ERK pathway and MAPK cascade are among the main components in the PDGF signaling pathway. However, the work [18] focuses on the molecules in the ERK pathway, which is a part of the MAPK pathway, and the results from [19] only cover the analysis for a subset of the long MAPK pathway. In our study, we investigate a more general representation of the MAPK pathway in PDGF signaling. Thus, these make the direct comparison of these studies to the results from our PDGF signaling pathway's analysis infeasible.

Pronk et al. [20] apply PRISM to the biological problem of codon bias. They show that the results obtained from the quantitative analysis in PRISM agree with the biological literature. Ribosome kinetics and aa-tRNA competition are modeled as CTMCS analyzed in PRISM [21]. In [22], Kwiatkowska et al. use PRISM to analyze the FGF (Fibroblast Growth Factor) signaling pathway. Although only a model corresponding to a single instance of the pathway is built, it is still rich enough to explain the roles of the components in the pathway and how they interact. The tunable activation threshold hypothesis of T Cells is studied through computational modeling of T cell signaling pathways in PRISM [23], and the authors demonstrate tuning and synergy. Jha *et al.* [24] present the first algorithm for performing statistical model checking using Bayesian sequential hypothesis testing and test the algorithm on the FGF signaling pathway and several others. More recently, Liò et al. use PRISM to diagnose the emerging of bone pathologies [25].

In addition, there are a few papers which apply the traditional methods in Systems Biology to study the PDGF signaling pathway. Zhang et al. [26] model the survival signaling in large granular lymphocyte leukemia, which is partly related to the PDGF signaling, using a Boolean model of the network's dynamics. Wang et al. [27] model the crosstalk interaction between MAPK and PI3K/Akt pathways in ODE format. The experimental data and the evidence of crosstalk reaction from [27] also partly contribute to the model structure and the justification of the reactions in our work.

Outline of the paper. In Sect. 2, we give an overview of probabilistic model checking and the tool PRISM. Sect. 3 describes the PDGF signaling pathway. In Sect. 4, we build a model in PRISM for the PDGF signaling pathway and describe several properties of the model that we are interested in. Our verification results are given in Sect. 5. In Sect. 6, we compare stochastic verification and ODE simulation by investigating the influence of model structure, parameter values and pathway length. Finally, we draw the conclusions of this paper and discuss some future work in Sect. 7.

2 Probabilistic Model Checking and PRISM

We briefly introduce stochastic verification and the model checker – PRISM [11].

2.1 CTMC and CSL

Probabilistic model checking is a variant of model checking, which aims at analyzing the correctness of finite state systems with a focus on quantitative aspects. Model checking of a system requires two inputs: a formal description of the system, which is usually given in a high-level modeling formalism (e.g., Petri nets or process algebra) and a specification of the system properties, which is usually given as temporal logic (e.g., CTL or LTL) formulas. After accepting the two inputs, a model checking tool then can verify whether the system satisfies the desired properties and give counter-examples if the system does not satisfy a certain property, by exploring all possible behaviors of the system exhaustively. As the word "probabilistic" indicates, probabilistic model checking focuses on systems with stochastic behaviors. Instead of asking the model checker "will the molecule become active in the end?", we can ask "what is the probability of the molecule being active at the steady state?" or "what is the probability of the molecule being active at time instant t?". In probabilistic model checking, systems are normally represented by Markov chains or Markov decision processes. In this paper, we use continuous-time Markov chains (CTMCs) to build the signaling pathway models and stochastic verification for thesis anlyses.

A CTMC can model both (continuous) real time and probabilistic choice by assigning rates at transitions between states. The formal definition of a CTMC is given as follows.

Definition 1. *Let $\mathbb{R}_{\geq 0}$ denote the set of non-negative reals and AP be a fixed finite set of atomic propositions. A CTMC is a tuple (S, R, L) where:*

- *S is a finite set of states;*
- *$R : S \times S \to \mathbb{R}_{\geq 0}$ is a transition rate matrix;*
- *$L : S \to 2^{AP}$ is a labeling function which associates each state with a set of atomic propositions.*

The transition rate matrix R assigns rates to each pair of states, which are used as parameters of the exponential distribution. A transition can occur between two states s and s' if $R(s, s') > 0$, and the probability of the transition being triggered within t time-units equals to $1 - e^{R(s,s')\cdot t}$. If $R(s, s') > 0$ for more than one state s', the first transition to be triggered determines the next state. Therefore, the choice of the successor state of s is probabilistic. The time spent in state s before any such transition occurs is exponentially distributed with $E(s) = \sum_{s' \in S} R(s, s')$. Hence, the probability of moving to state s' is $\frac{R(s,s')}{E(s)}$, i.e., the probability that the delay of going from s to s' "finishes before" the delays of any other outgoing transition from s. A path in a CTMC is a sequence σ in the form of $s_0 t_0 s_1 t_1 \cdots$ with $R(s_i, s_{i+1}) > 0$ and $t_i \in \mathbb{R}_{\geq 0}$ for all $i \geq 0$. The amount of time spent in s_i is denoted by t_i.

For a CTMC, we consider two types of state probabilities: *transient* probability is related to a state in the CTMC at a particular time instant, and *steady* probability

describes the state of the CTMC in a long run. If we denote the state of the CTMC at time t as $X(t)$, the transient probability at time t is the probability that the CTMC is in state s at time t, i.e., $p_s(t) = Pr\{X(t) = s\}$. Intuitively, the steady-state probability for being at state s is then defined as

$$p_s = \lim_{t \to \infty} p_s(t).$$

Corresponding to CTMC models, we use Continuous Stochastic Logic (CSL) to specify properties of built models. CSL, originally introduced by Aziz *et al.* [28], provides a powerful means to specify both path-based and traditional state-based performance measures on CTMCs.

Definition 2. *The syntax of CSL is given as follows:*

$$\phi \ ::= \ true \mid a \mid \neg\phi \mid \phi \wedge \phi \mid P_{\sim p}[\phi U^I \phi] \mid S_{\sim p}[\phi]$$

where a is an atomic proposition, $\sim \in \{<, \leq, \geq, >\}$, $p \in [0, 1]$, and I is an interval of $\mathbb{R}_{\geqslant 0}$.

CSL formulas are evaluated over the states of a CTMC. CSL includes the standard operators from propositional logic: $true$ (satisfied in all states); atomic propositions (a is true in states which are labelled with a); negation ($\neg\phi$ is true if ϕ is not); and conjunction ($\phi_1 \wedge \phi_2$ is true if both ϕ_1 and ϕ_2 are true). Other standard boolean operators such as disjunction ($\phi_1 \vee \phi_2 \equiv \neg(\neg\phi_1 \wedge \neg\phi_2)$) and implication ($\phi_1 \Rightarrow \phi_2 \equiv \neg\phi_1 \vee \phi_2$) can be derived from these in the usual way. CSL also includes two probabilistic operators, P and S, both of which include a probability bound $\sim p$. A formula $P_{\sim p}[\psi]$ is true in a state s if the probability of the path formula ψ being satisfied from state s meets the bound $\sim p$. A path formula evaluates to either true or false for a single path in a model. In this paper, we use a simple type of the path formula, $F^I\phi \equiv true\, U^I \phi$, called an *eventual* formula, which is true for a path σ if ϕ eventually becomes true for some time instant $t \in I$. Particularly, if the time interval is set to zero, e.g. $F^{[t,t]}\phi$, the formula is true for a path σ if ϕ becomes true at time instant t. The S operator is used to specify steady-state behavior of a CTMC, i.e., its behavior in the long-run or equilibrium. More precisely, $S_{\sim p}[\psi]$ asserts that the steady-state probability of being in a state satisfying ψ meets the bound $\sim p$. We refer the reader to the papers [29, 7, 8] for model checking algorithms for computing steady-state probabilities.

2.2 The model checker PRISM

PRISM [11] is a model checking tool developed at the universities of Birmingham and Oxford. It allows one to model and analyze systems containing stochastic behaviors. PRISM supports three kinds of models: discrete-time Markov chains (DTMCs), continuous-time Markov chains (CTMCs) and Markov decision processes (MDPs). Analysis is performed through model checking such systems against properties written in the probabilistic temporal logics PCTL if the model is a DTMC or an MDP, or CSL in the case of a CTMC, as well as their extensions for quantitative specifications and costs/rewards.

In PRISM a model consists of a number of modules that contain variables and can interact with each other. The values of the variables at any given time constitute the state of the module, and the local states of all modules decide the global state of the whole model. The behavior of a module, normally the changes in states which it can undergo, is specified by a set of guarded commands of the form:

$$[a]\ g \rightarrow r : u;$$

a is an action label in the style of process algebra, which introduces synchronization into the model. It can only be performed simultaneously by all modules that have an occurrence of action label a. If a transition does not have to synchronize with other transitions, then no action label needs to be provided for this transition. The symbol g is a predicate over all the variables in the system. A guarded command $g \rightarrow r : u$ means that if the guard g is true, the system is updated according to u with rate r, which is corresponding to the transition rate of CTMC. A transition updates the value of variables by giving their new value of the form $x' = \mathsf{expr}$, where x is a variable and its primed version x' refers to the value of x in the next state, expr is an expression built on the *unprimed* variables. If an update does not contain $x' = ...$, then the value of the variable x remains unchanged.

PRISM models can be augmented with information about rewards (or equivalently, costs). The tool can analyze properties which relate to the expected values of these rewards. A CTMC in PRISM can be augmented with two types of rewards: *state reward* associated with states which are accumulated in proportion to the time spent in the state, and *transition reward* associated with transitions which are accumulated each time the transition is taken. CSL is extended with quantitative costs/rewards as well, which is quite useful in analyzing the quantitative properties of a biological system, by introducing the R operator:

$$R\ ::=\ R_{\sim r}[I^{=t}] \mid R_{\sim r}[C^{\leq t}] \mid R_{\sim r}[F\phi] \mid R_{\sim r}[S]$$

where $\sim \in \{<, \leq, \geq, >\}$, $r, t \in \mathbb{R}_{\geq 0}$ and ϕ is a CSL formula. Intuitively, a state s satisfies $R_{\sim r}[I^{=t}]$ if from s the expected state reward at time instance t meets the bound $\sim r$; a state s satisfies $R_{\sim r}[C^{\leq t}]$ if the expected reward accumulated up until t time units past satisfies $\sim r$; a state s satisfies $R_{\sim r}[F\phi]$ if from s the expected reward accumulated before a state satisfying ϕ is reached meets the bound $\sim r$; and a state s satisfies $R_{\sim r}[S]$ if from s the long-run average expected reward satisfies $\sim r$.

It is often useful to take a quantitative approach – computing the actual probability that some behavior of a model is observed, rather than just verifying whether or not the probability is above or below a given bound. Hence, PRISM allows the P and S operators in CSL to take the following form: $P_{=?}[\psi]$ and $S_{=?}[\psi]$.

3 The PDGF Signaling Pathway

3.1 Biology of the PDGF signaling pathway

Cell signaling is part of a complex system in cellular communication. It allows the cells to activate a large number of signaling molecules and to regulate their activity. In

order to transfer a regulatory signal upon reception of a triggering stimulus, the signal is transformed into a chemical messenger within the signaling cell, e.g., via transfer of a phosphate group (phosphorylation) [30]. For further details on cell signaling see, for example, [31, 30].

Platelet-Derived Growth Factor (PDGF), described approximately 30 years ago as a major mitogenic component of whole blood [13], is a growth factor that regulates cell growth and division. By binding to its receptor (PDGFR), it regulates many biological processes such as migration, survival and proliferation [32]. PDGFR is a receptor tyrosine kinase, which in general transfers upstream signals to many downstream signaling pathways by phosphorylation. Up to now, five pairs of PDGF which can be formed as a molecule that can bind with its receptor to form a complex (or so called 'ligand') are known, PDGF-AA, -AB, -BB, -CC, -DD, interacting with three different types of PDGFR complexes, PDGFR-$\alpha\alpha$, -$\alpha\beta$ and -$\beta\beta$. Each of the PDGFR subtypes has a different affinity to the different PDGF ligands [33].

After PDGFRs couple with their respective ligands, phosphorylation of the receptor at specific tyrosine residues will occur, thus enabling binding of signaling enzymes including Src, phosphatidylinositol 3 kinase (PI3K), phospholipase Cγ (PLCγ) and SHP2 in the MAPK pathway at specific binding sites. The recruitment of these signaling enzymes to PDGFR is mediated via an intrinsic SH2 domain. The translocation of PI3K and PLCγ to the plasma membrane also increases their accessibility to their respective substrates. Moreover, recent findings suggest that PDGFR also has potential binding sites for CrkL [34], which will activate Rap1 to positively influence c−Raf in the MAPK pathway [35], for Signal Transducer and Activator of Transcription (STAT), which might regulate the signal in parallel to the JAK-STAT pathway [36] and also for cCbl, which promotes ubiquitination of PDGFR. cCbl is also considered to be one of the negative regulatory molecules in PDGF signal transduction [37].

3.2 Model structure of the PDGF signaling pathway

Based on intensive literature reviews, we have built a PDGF signal transduction model in ODE format which consists of 17 molecules (see Fig. 1). The model consists of the main parts from the PDGF signaling pathway including the coupling of PDGF ligand to PDGFR, the negative regulatory feedbacks on PDGFR and the activation of two main downstream signaling pathways, the MAPK and PI3K/Akt pathways, with the crosstalk interactions between the two pathways. The scope of our model with 17 nodes is sufficient to capture the main dynamic behavior of the PDGF signaling pathway which is further analyzed in PRISM.

Fig. 1 describes how signals are transduced in the PDGF pathway by activating or deactivating specific downstream pathways signaling molecules. In this model, there are three inputs, viz. PDGFL (PDGF ligand), bPTEN, and bPDK. PDGFL is the node which represents the upstream molecule activating the whole network. PPX, and bPTEN are nodes which represent phosphatase enzymes in the cytoplasm that negatively regulate their targets. Lastly, bPDK, standing for basal activity of PDK, is the node that constantly gives a basal additional input to PDK node. This node is always active in order to activate the survival pathway to counteract apoptotic signaling and keep the cell survive at a basal level. There are three different types of arrows in the network: blue

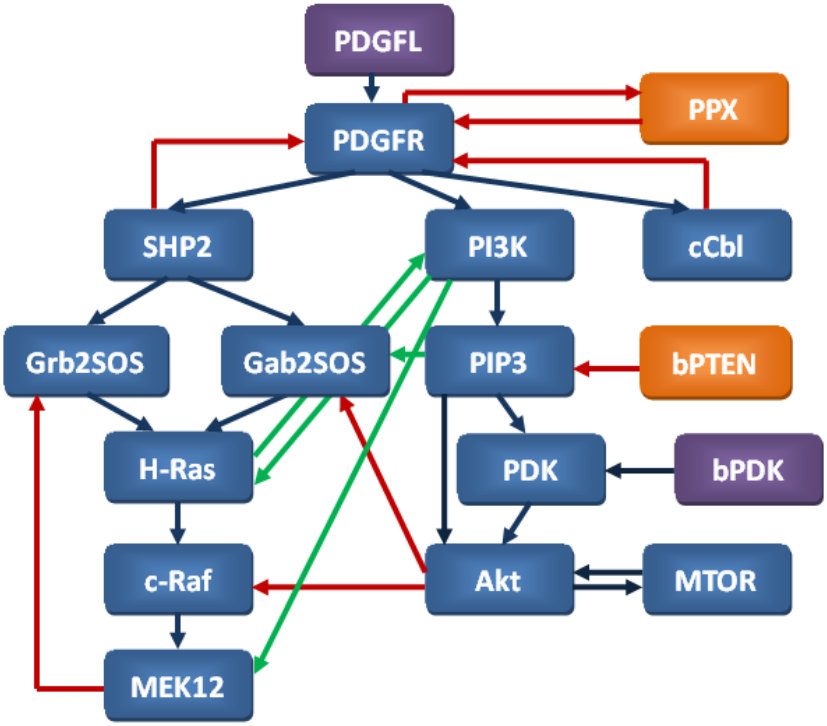


Fig. 1. The extracted PDGF signaling pathway (blue arrows: main pathway, green arrows: positive crosstalk, red arrows: negative regulatory)

arrows, green arrows and red arrows. The blue arrows represent the main activating interactions, indicating the two main downstream signaling pathways in the network. The MAPK pathway covers `SHP2`, `Grb2SOS`, `Gab2SOS`, `Ras`, `c-Raf` and `MEK12`. These molecules play a major role in the cellular proliferative circuit [38]. The PI3K/Akt pathway covers the molecules `PI3K`, `PIP3`, `PDK`, `bPTEN`, `bPDK` and `Akt`. In addition, this pathway closely interacts with the `MTOR` node which represents the mTOR pathway through positive feedback regulation. Both of these pathways play a major role together in the viability circuit of the cells [38]. The green arrows represent positive crosstalk interactions to the molecules the arrows point to. Lastly, the red arrows represent either negative crosstalk interactions or other negative regulatory interactions. The molecules will become active after they have been activated by either blue or green arrows. In contrary, the molecules will become inactive after they have been deactivated by the red arrows or the basal phosphatase activities in the cell (not shown in the figure).

`PDGFR` can be activated by `PDGFL`. The active `PDGFR` in turn activates three downstream molecules which are `SHP2`, `PI3K` and `cCbl`. Both `SHP2` and `cCbl` assert a negative feedback to PDGF making it inactive. The three blue arrows connecting `PDGFR` to these downstream signaling molecules, so called *mutant arrows*, are the targets of system interventions both experimentally and computationally. The experimental intervention can be performed by introducing a point mutation from tyrosine to phenylalanine at the specific recruitment site for the downstream signaling enzyme, for

instance, Y720F for SHP2 recruitment site, YY731/742FF for PI3K recruitment site, and Y1018F for cCbl recruitment site, leading to the loss of signal capacity of the respective signaling pathway [39–41]. Thus, the result of computational simulation such as the relative activities at the steady state of downstream signaling molecules from these respective mutants can be validated experimentally in biological laboratories.

After the upstream signaling molecules, SHP2 and PI3K, have been activated, they transfer the signal via the phosphorylation process to downstream signaling molecules. In the MAPK pathway, the process starts from the transfer of the signal from SHP2 to Grb2SOS and Gab2SOS. Then, the two molecules in turn transfer the signal to H−Ras, c−Raf, and MEK12, respectively. There is also a negative feedback regulation from MEK12 to Grb2SOS to modulate the signal in this pathway. In the PI3K/Akt pathway, the signal transfer starts from PI3K to PIP3 which then in turn activates PDK and Akt. PIP3 can be deactivated by the phosphatase enzyme PTEN (represented as bPTEN) and the node PDK can get additional input from the basal activity of itself (represented as bPDK). At the downstream part of the PI3K/Akt pathway, the node Akt can be activated by either PDK or PIP3 and it also forms a positive feedback loop with the mTOR pathway which was represented as the MTOR node.

In addition to the activation and regulation within each pathway, there are also several crosstalk reactions between the two pathways. These are the positive crosstalk regulations from PI3K to MEK12, from PIP3 to Gab2SOS, and a positive feedback loop between PI3K and H−Ras. In parallel, there are also negative regulations from Akt to Gab2SOS and c−Raf. These crosstalk reactions modulate the signals between the two pathways to generate a robust network.

Fig. 2 contains the list of model reactions. Each molecule is simplified to be in two states, either *inactive* or *active* (indicated by the suffix _act in the figure). All the reactions except for reactions 9, 10 and 11 describe the molecules changing between the two states; while reactions 9, 10 and 11 indicate the basal production, the basal degradation and the internalization following activation of PDGFR. For instance, reaction 2 describes that an active PPX gives a negative feedback to PDGFR, making it inactive.

3.3 Conversion of the interaction graph to a reaction-based ODE model

Generally, we can generate a set of biochemical reactions to represent the interactions for each molecule based on the interaction graph presented in Fig. 1. Nevertheless, the strengths of both activation/phosphorylation and inhibition/dephosphorylation interactions for each molecule are different. Therefore, during the conversion process of an interaction graph to biochemical reactions, a consistent procedure is applied to determine the biochemical reactions as well as the parameter values. In this section we will explain this procedure in some detail.

In our study, we assigned the parameter set based on the knowledge derived from the literature (e.g., [14, 15, 13, 16, 17]) and our own experimental observation of PDGFRα mutant. We also obtained the information that the time constants of the MAPK and PI3K/Akt pathways are comparable as shown in [27]. The experimental data in [27] shows that the time for the MAPK pathway to activate ERK12 (the molecule below MEK12) and the time for the PI3K/Akt pathway to activate Akt are highly similar.

a. PDGFR and PPX	
1) PDGFL + PDGFR → PDGFR_act + PDGFL	kon
2) PDGFR_act + PPX_act → PDGFR + PPX_act	koff1
3) PDGFR_act + SHP2_act → PDGFR + SHP2_act	koff2
4) PDGFR_act + cCbl_act → PDGFR + cCbl_act	kubi
5) PDGFR_act + PPX_act → PPX + PDGFR_act	koffppx
6) PDGFR_act + cCbl → cCbl_act + PDGFR_act	k1
7) PDGFR_act + SHP2 → SHP2_act + PDGFR_act	k5
8) PDGFR_act + PI3K → PI3K_act + PDGFR_act	k6
9) → PDGFR	kbasal
10) PDGFR →	kbasal
11) PDGFR_act →	kdeg
b. SHP2, Grb2SOS and Gab2SOS	
12) SHP2_act + Grb2SOS → Grb2SOS_act + SHP2_act	k52
13) SHP2_act + Gab2SOS → Gab2SOS_act + SHP2_act	k522
14) SHP2_act → SHP2	kp5
15) Grb2SOS_act + H-Ras → H-Ras_act + Grb2SOS_act	k53
16) Grb2SOS_act → Grb2SOS	kp52
17) MEK12_act + Grb2SOS_act → Grb2SOS + MEK12_act	kcross1
18) Gab2SOS_act + H-Ras → H-Ras_act + Gab2SOS_act	k532
19) Gab2SOS_act → Gab2SOS	kp522
20) PIP3_act + Gab2SOS → Gab2SOS_act + PIP3_act	kcross2
21) Akt_act + Gab2SOS_act → Gab2SOS + Akt_act	kcross3
c. H-Ras, c-Raf and MEK12	
22) H-Ras_act + c-Raf → c-Raf_act + H-Ras_act	k54
23) H-Ras_act → H-Ras	kp53
24) PI3K_act + H-Ras → H-Ras_act + PI3K_act	kcross4
25) H-Ras_act + PI3K → PI3K_act + H-Ras_act	kcross9
26) c-Raf_act + MEK12 → MEK12_act + c-Raf_act	k55
27) c-Raf_act → c_Raf	kp54
28) Akt_act + c-Raf_act → c-Raf + Akt_act	kcross6
29) MEK12_act → MEK12	kp55
30) PI3K_act + MEK12 → MEK12_act + PI3K_act	kcross7
d. PI3K, PIP3 and PDK	
31) PI3K_act + PIP3 → PIP3_act + PI3K_act	k62
32) PI3K_act → PI3K	kp6
33) PIP3_act + PDK → PDK_act + PIP3_act	k63
34) PIP3_act + Akt → Akt_act + PIP3_act	k64
35) PIP3_act + bPTEN → PIP3 + bPTEN	kp62
36) bPDK + PDK → PDK_act + bPDK	k632
37) PDK_act + Akt → Akt_act + PDK_act	k642
38) PDK_act → PDK	kp63
e. Akt, cCbl, MTOR, bPTEN and bPDK	
39) MTOR_act + Akt → Akt_act + MTOR_act	k643
40) Akt_act + MTOR → MTOR_act + Akt_act	k644
41) Akt_act → Akt	kp64
42) cCbl_act → cCbl	kp1
43) MTOR_act → MTOR	kp642

Fig. 2. List of biochemical reactions in the PDGF signaling pathway ODE model with respective parameters in tab-separated format

All parameters are assigned as relative values in the range of zero to one, compared to the sum of all positive reaction rates around the respective molecule. We believe that this assignment could still capture the real reaction rates on the same molecule. Moreover, these parameter values should be able to represent the rates which are comparable to another signaling pathway as the time constants between the two signaling pathways are relatively similar. In general, we follow Kwiatkowska et al.'s work [10, 11] on translating kinetic rates to stochastic rates. Namely, for first-order (non-binary) reactions they take the stochastic rate to be the kinetic rate; for binary reactions, assuming that the kinetic rate is given in terms of molar concentrations, the stochastic rate can be obtained by dividing by the product of the volume and Avogadro's number [42]. In the case of bimolecular reactions (second order) modeled by the standard law of mass action as used in our paper (e.g., reactions 2-8), the kinetic rate and the stochastic rate are equal if the states are normalized to the maximal value (as done in our paper), according to the conversion of mole in chemistry.

To assign the parameter set, the reaction rate of that respective reaction is equal to one if only one molecule which activates the respective node is present. For example, the node SHP2 is solely activated by `PDGFR` once `PDGFR` is active (***PDGFR_act***). Thus, the biochemical reaction for the activation of the node `SHP2` is generated as follow:

$$PDGFR\_act + SHP2 \xrightarrow{k5} SHP2\_act + PDGFR\_act$$

`SHP2` will only be activated by `PDGFR_act`. Therefore, the respective parameter ***k5*** is assigned to 1.0. This rule is also applied to all nodes which have only one positive interaction on the interaction graph such as `cCbl` or `MTOR`. Apart from this, the activity of the node `SHP2` is also under the influence of the activities of cytoplasmic phosphatase enzymes which dephosphorylate the active form of `SHP2` (***SHP2_act***) resulting in the dephosphorylated/inactive state (***SHP2***). This is represented in the following biochemical reaction:

$$SHP2\_act \xrightarrow{kp5} SHP2$$

From biological observation, the reaction rate of the phosphorylation process is significantly higher than the rate of the dephosphorylation process [43]. In our study, it is assumed that the sum of the dephosphorylation strength for each molecule by cytoplasmic phosphatase enzymes in general is roughly 10% compared to the maximal phosphorylation strength. Therefore, parameter ***kp5***, as well as all ***kp***-parameters which represent the constant parameters for the dephosphorylation process, are all set to 0.1. In this case, we can derive an ODE for the ***SHP2_act*** node as follows:

$$d/dt(SHP2\_act) = k5 \times PDGFR\_act \times SHP2 - kp5 \times SHP2\_act$$

In another case, if there is more than one positive interaction on a single node in the interaction graph, we assume that the sum of all activation parameters is equal to one. In this situation, we additively consider the strength of each interaction according to their interaction strengths derived from the literature, without considering synergistic effects. These interaction strengths are then considered in term of relative values which have been assigned to each parameter accordingly. In general, the reaction rate of the

canonical pathway (the main pathway of activation) is always higher than the rate of other additional inputs such as crosstalk reactions or basal activities. For instance, node PDK can be activated by node PIP3 with an interaction strength of 90%, while it can also be activated by node bPDK which represents the basal activity of PDK by 10%. In this case, two biochemical reactions of PDK node activation are generated as follows:

$$PIP3\_act + PDK \xrightarrow{k63} PDK\_act + PIP3\_act$$
$$bPDK + PDK \xrightarrow{k632} PDK\_act + bPDK$$

In this case, parameter *k63* has been assigned to 0.9 while parameter *k632* has been assigned to 0.1, according to their interaction strengths as mentioned. When also considering the dephosphorylation rate, we could derive an ODE for the *PDK_act* node as follows:

$$d/dt(PDK\_act) = k63 \times PIP3\_act \times PDK + k632 \times bPDK \times PDK - kp63 \times PDK\_act$$

The same rule is also applied to nodes with the same type of interaction such as H−Ras or Akt.

In the most complex case, in the situation that there are both positive and negative interactions on a single node, special consideration and assignment are applied. First, we separate the positive and negative interactions and we consider only the strength of all positive interactions as a sum value of one. Then, we consider the proportion of negative interaction to deactivate/dephosphorylate the molecule in addition to the activities from cytoplasmic phosphatase enzymes with the parameter values according to their inhibition strengths. To give an example, node Gab2SOS is activated by two nodes which are SHP2 with a strength of 80% and PIP3 with strength 20%, according to the maximal activation strength. Node Gab2SOS is also deactivated by node Akt with strength 20% and by cytoplasmic phosphatase enzymes with strength 10%, according to the maximal activation. All biochemical reactions which are related to the changing states of the Gab2SOS node are shown as follows:

$$SHP2\_act + Gab2SOS \xrightarrow{k522} Gab2SOS\_act + SHP2\_act$$
$$PIP3\_act + Gab2SOS \xrightarrow{kcross2} Gab2SOS\_act + PIP3\_act$$
$$Akt\_act + Gab2SOS\_act \xrightarrow{kcross3} Gab2SOS + Akt\_act$$
$$Gab2SOS\_act \xrightarrow{kp522} Gab2SOS$$

As already mentioned, the parameters for the positive interactions from SHP2 and PIP3 are firstly considered and they have been assigned to each reaction with a sum of 1.0. In this case, parameter *k522* is assigned to 0.8 and parameter *kcross2* is assigned to 0.2. Then, the additional negative interaction from Akt is considered and the respective parameter is assigned. Thus, parameter *kcross3* is assigned to 0.2. Lastly, the basal dephosphorylation activity from cytoplasmic phosphatase enzymes which is represented by parameter *kp522* is assigned to 0.1. In this case, an ODE is derived for *Gab2SOS_act* node as follows:

$$d/dt(Gab2SOS\_act) = k522 \times SHP2\_act \times Gab2SOS - kp522 \times Gab2SOS\_act + kcross2 \times PIP3\_act \times Gab2SOS - kcross3 \times Akt\_act \times Gab2SOS\_act$$

After we applied this procedure to convert the interaction graph to a biochemical reaction-based ODE model, we obtained a set of biochemical reactions as well as the respective parameter set for our modeling analysis in PRISM.

3.4 Sensitivity analysis of the derived ODE model

As mentioned in Sect. 3.3, the parameter set which has been derived from the interaction graph integrated with knowledge from the literature and our own experimental observations might not fully correlate to the actual biological system. Nevertheless, the dynamic behavior of the system is still conserved in the model structure. In this section, we present the results from the sensitivity analysis to confirm the validity of this statement.

After we obtained the ODE model from the interaction graph, we imported the ODE model into Matlab using Systems Biology Toolbox 2 [44] where a global parameter sensitivity analysis with the FAST method is integrated [45]. The analysis was performed by perturbing all parameter values to observe how the state values would differ based on the perturbation. The perturbation scale used in this analysis is one order of magnitude (for instance, from 1 to 10 or to 0.1).

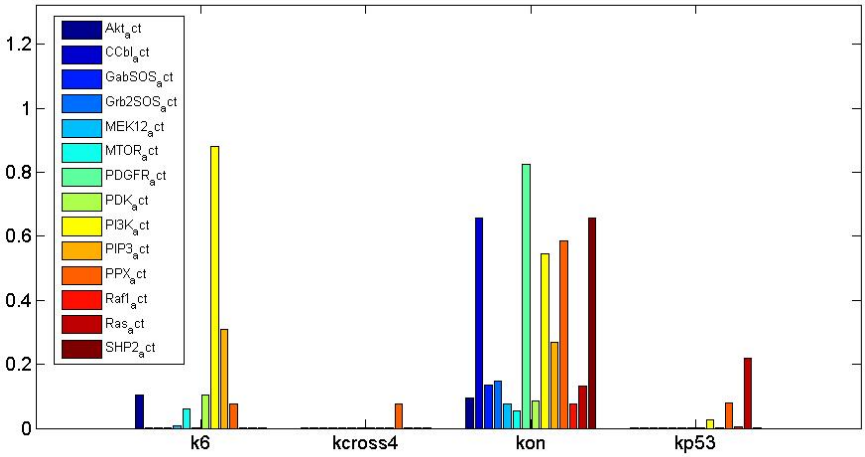
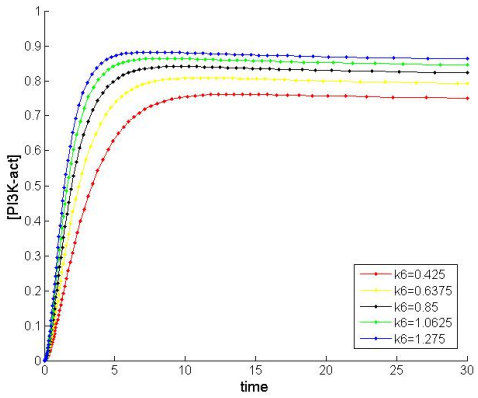


Fig. 3. Sensitivity analysis of the derived ODE model for some selected parameters

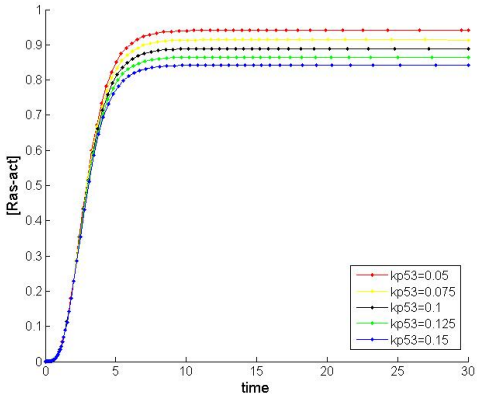
In Fig. 3, we see that some states are sensitive to the change of specific parameters. For instance, *PI3K_act* is sensitive to *k6* and *Ras_act* is sensitive to *kp53*. In parallel, we observe that there is no state, except for *PPX_act*, which is sensitive to the change of *kcross4*. In contrast, there are also many states which are sensitive to a change of a single parameter. For example, *kon*, which is the parameter involved with the activation of PDGFR by PDGF ligand and in turn activates the whole system, contributes to the changes of many activated form of molecules in the model when this parameter value is

perturbed. Based on this study, we identified the sensitive pairs of states and parameters such as *PI3K_act* which is sensitive to *k6*, *Ras_act* which is sensitive to *kp53* and most of the molecules in activated states which are all sensitive to *kon*. In addition, we also identified the insensitive pairs, such as *Ras_act* which is insensitive to *kcross4*.

Next, we challenged the system by increasing and decreasing selected parameter values from each pair up to 50%. For instance, the original *kon* parameter value of 1 was perturbed to 1.5 and 0.5, respectively. Then, the states trajectories of the respective molecules were plotted to observe the change of systems behavior according to the changes of the corresponding parameter values (see Fig. 4, 5 and 6).



(a) *PI3K_act* concentration when *k6* changes from 25% to 50% (original $k6 = 0.85$)



(b) *Ras_act* concentration when *k6* changes fror 25% to 50% (original $kp53 = 0.1$)

Fig. 4. Sensitivity analysis of *PI3K_act* concentration vs. *k6* and *Ras_act* concentration vs. *kp53*

From the results in Fig. 4, we observe the changes of the steady state values of *PI3K_act* and *Ras_act* which are related to changes of parameter values of *k6* and *kp53* accordingly. Nevertheless, we still see that the dynamic change of the molecules which are reflected by the shapes of the figures remained mostly the same. The same observation can be made for Fig. 5 where the *kon* parameter is perturbed and the state changes of *PDGFR_act*, *SHP2_act*, *PI3K_act*, *MEK12_act* and *Akt_act* are observed. Here, we see that *PDGFR_act* is activated slower and weaker (when $kon = 0.5$), or faster and stronger (when $kon = 1.5$), according to the *kon* values. Nevertheless, the dynamic changes (shape of the figures) as well as the steady state values of *PDGFR_act*, upstream molecules (*SHP2_act* and *PI3K_act*) and downstream molecules (*MEK12_act* and *Akt_act*) were not drastically changed. Moreover, the order of steady state values of these molecules is still preserved ($Akt\_act > MEK12\_act > PI3K\_act > SHP2\_act > PDGFR\_act$), even though the parameter value has been perturbed up to 50%. In parallel, when we plotted the state change of *Ras_act* after the perturbation on *kcross4*, little changes are observed as shown in Fig. 6, because *Ras_act* is insensitive to the change of this parameter. According to these results, we find that the decision to normalize the reaction rates in the interval between 0 and 1 is sensible to observe the dynamic changes of the systems which have been conserved within the model structure.

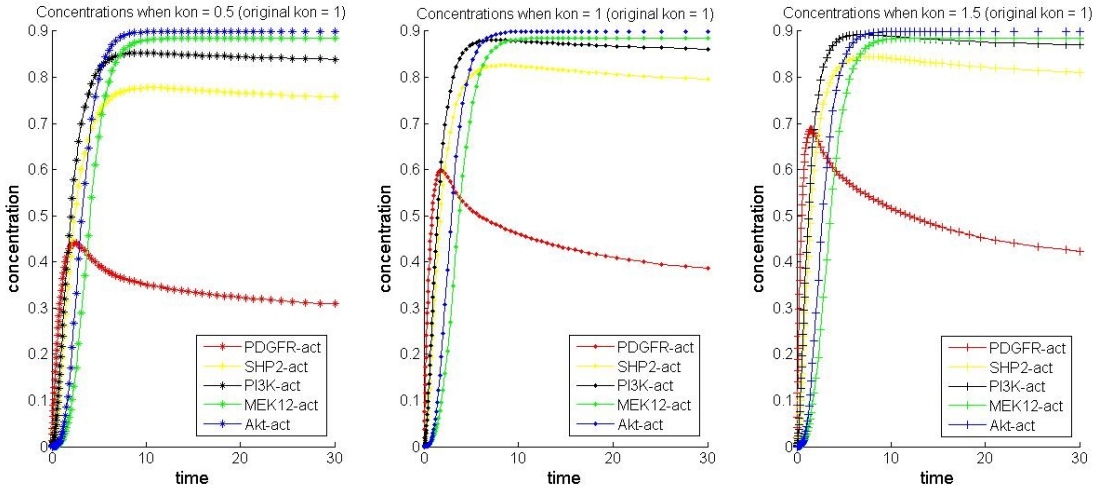


Fig. 5. Sensitivity analysis of the concentrations of *PDGFR_act*, *SHP2_act*, *PI3K_act*, *MEK12_act* and *Akt_act* vs. *kon*

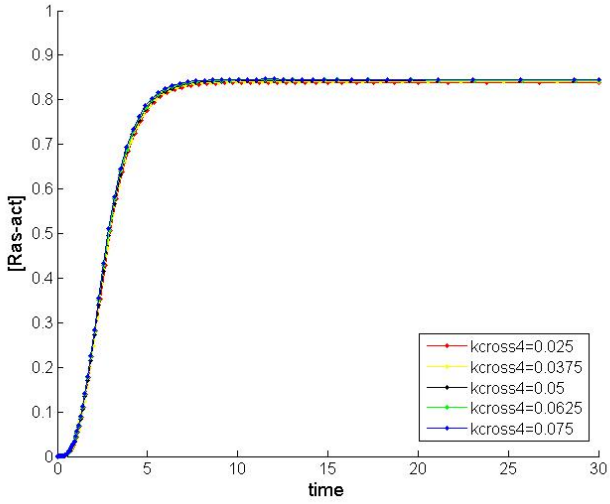


Fig. 6. Sensitivity analysis of *Ras_act* concentration when *kcross4* changes from 25% to 50% (original $kcross4 = 0.05$)

4 Modelling and Property Specifications in PRISM

4.1 PRISM model

We now describe in detail how to build a PRISM model for the PDGF signaling pathway as presented in the previous section. Though our model represents a single instance of the signaling pathway, meaning there can be at most one element of each molecule, we believe it is still rich enough to explain the roles of the molecules in the pathway

and how they interact with each other as shown in the literature [22]. In the single instance model, a molecule's steady state, which is expressed as a probability in PRISM, potentially corresponds to the molecule density in the biological experiments.

```
module PDGFR

    PDGFR : [0..2] init 0; //0 – inactive; 1 – active; 2 – degraded
    [] PDGFL & PDGFR=0 → kon : (PDGFR'=1);
    [bkoff1] PDGFR=1 → koff1 : (PDGFR'=0);
    [bkoff2] PDGFR=1 → koff2 : (PDGFR'=0);
    [bkubi] PDGFR=1 → kubi : (PDGFR'=0);
    [bkoffppx] PDGFR=1 → koffppx : (PDGFR'=1);
    [bk1] PDGFR=1 → k1 : (PDGFR'=1);
    [bk5] PDGFR=1 → k5 : (PDGFR'=1);
    [bk6] PDGFR=1 → k6 : (PDGFR'=1);
    [] PDGFR=0 → kbasal : (PDGFR'=2); //PDGFR degraded
    [] PDGFR=2 → kbasal : (PDGFR'=0);
    [] PDGFR=1 → kdeg : (PDGFR'=2); //PDGFR_act degraded

endmodule
```

(a) PRISM module for PDGFR

```
module PPX

    PPX : [0..1] init 1;
    [bkoff1] PPX=1 → (PPX'=1);
    [bkoffppx] PPX=1 → (PPX'=0);

endmodule
```

(b) PRISM model for PPX

```
rewards
"PDGFRactive"
    PDGFR=1:1;
endrewards
```

(c) PRISM rewards

Fig. 7. PRISM modules and rewards

Each of the nodes (or molecules) of the pathway in Fig. 1, except for the PDGFL, is represented by a separate PRISM module. Since PDGFL, bPTEN and bPDK remain the same after the reactions they are involved in, we set them as constant boolean values true. Fig. 7(a) and Fig. 7(b) show the modules for PDGFR and PPX. In each of the modules, the status of the molecule is represented by a variable with the same name as the module. The variables can have values of either 0 or 1 (PDGFR is an exception, because it can have value 2 since it can degrade), corresponding to the two states, inactive and active, of a molecule. Each command in the PRISM module represents a reaction in Fig. 2. Interactions of multiple molecules are implemented by the synchronization between modules. More precisely, the same label is given to the commands which require synchronization in PRISM modules. For example, in Fig. 7(a) and Fig. 7(b), there

are commands with label `bkoff1` in both of the modules `PDGFR` and `PPX`. The two commands are used to model the reaction (2) in Fig. 2, which involves both `PDGFR` and `PPX`. It guarantees that the two commands (corresponding to one reaction) can only occur when both guards are satisfied. The reaction rate is assigned by the command in module `PDGFR` and hence the reaction rate of the command in module `PPX` is omitted. We have modeled all the 17 molecules in 14 PRISM modules (`PDGFL`, `bPTEN` and `bPDK` are modeled as a constant).

As mentioned in Sect. 2.2, PRISM models can be augmented with information about rewards. We construct rewards to calculate the time for a molecule being active. Fig. 7(c) shows the rewards for calculating the active state of `PDGFR`. Each time `PDGFR` is in its active state, one is added to the total time of `PDGFR` being active. Similarly, we build rewards structures for other molecules as well, including `SHP2`, `Ras`, `MEK12`, `PIP3` and `Akt`.

4.2 Property specifications

As stated in the introduction, the three main goals for this study are: (1) to analyze the dynamics of PDGF induced signaling, (2) to analyze the influence of the crosstalk reactions as defined in Sect. 3, and (3) to analyze the importance of individual reactions on downstream signaling molecules. For the first goal, we study the signal transduction properties of each mutant by removing the mutant arrows one by one and examining how the states of each molecule change accordingly at different time instances. We also examine the total time for each molecule being active. Moreover, it is interesting to study the activities of each molecule at the steady state as well. For the second goal, we do the comparison of probabilities for molecules to be active between different mutants by removing each of the crosstalk reactions. For the last one, we study how the steady state probabilities of molecule `MEK12` and `Akt` change when a certain reaction is removed.

Below we list the properties of the PRISM model that we have analyzed to achieve our goals. Here, we use only the molecule `PDGFR` to illustrate the specification of the properties expressed as CSL formulas.

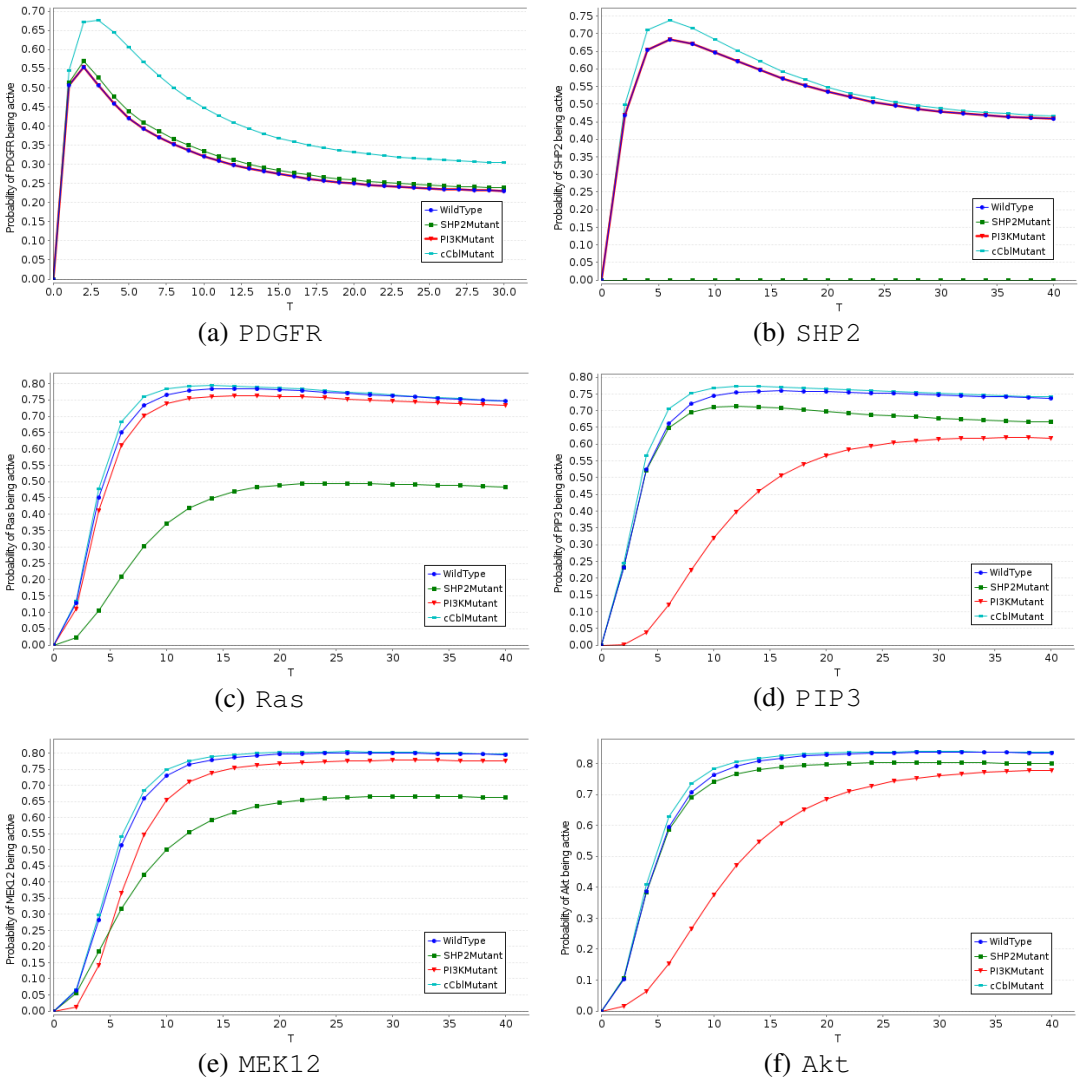
- $P_{=?}[\, F^{[t,t]}\texttt{PDGFR} = 1\,]$
  The probability that the molecule `PDGFR` is active at time instant t.
- $R^{\{\text{“}PDGFRactive\text{”}\}}_{=?}[C <= t]$
  The expected time of `PDGFR` being active by time t. It refers to the reward structure “PDGFRactive” as defined in Fig. 7(c).
- $S_{=?}[\, \texttt{PDGFR} = 1\,]$
  The long-run probability that `PDGFR` is active.

5 Verification Results

We use PRISM to construct the PDGF model described in Sect. 4.1 and analyze the set of properties listed in Sect. 4.2. We describe how to achieve our goals as explained in Sect. 1 in the following subsections.

5.1 Analyzing the dynamics of PDGF induced signaling

For the first goal to analyze the dynamics of PDGF induced signaling, we first develop a base model representing the system, in which all the reactions in Fig. 2 are included. Subsequently, we obtain the mutant models by removing the mutant arrows (as mentioned in Sect. 3) one by one. More precisely, we have developed four models in this experiment including the base model corresponding to the `WildType` condition. The second one, called `SHP2Mutant`, is obtained by removing the mutant arrow pointing to `SHP2`. The third and the fourth ones are `PI3KMutant` and `cCblMutant`. Following the same process, we remove the mutant arrows pointing to `PI3K` and `cCbl` separately in the last two models. The three removed mutant arrows correspond to the reactions (7), (8) and (6) in Fig. 2. For each model, we compute the probability of each molecule being active at time instance t, which is summarized in Fig. 8.



(a) `PDGFR` (b) `SHP2`

(c) `Ras` (d) `PIP3`

(e) `MEK12` (f) `Akt`

Fig. 8. PRISM verification results: Probabilities for molecules being active

Fig. 8 shows the probability of 6 out of the 17 molecules namely PDGFR, SHP2, Ras, PIP3, MEK12, and Akt. The 6 molecules are chosen according to their positions in the signaling pathway. We can see from Fig. 8(a) that cCbl affects PDGFR more than the other two molecules (SHP2 and PI3K). This is due to strong negative feedback (red arrow) from cCbl to PDGFR in Fig. 1. There is also a negative feedback from SHP2 to PDGFR. However, the reaction rate of the negative feedback from cCbl is 7 times as large as the one from SHP2, hence cCbl can affect PDGFR more than SHP2. In contrast, there would be no effect of PI3KMutant to PDGFR as there is neither a direct regulatory signal from PI3K on PDGFR nor indirect regulation through the molecule which can regulate PDGFR such as SHP2 or cCbl. Apart from this, the red curve in Fig. 8(d) shows that not only PIP3 in PI3KMutant is less active than in other conditions, but also the time of it becoming active is delayed. In biological experiments, this delay is *not* observed. This is due to the fact that there is another molecule PLCg giving a fast positive crosstalk to PI3K. This interaction is outside of the scope of the PDGF model we consider here.

If we focus on the light blue curves in Fig. 8(a) – Fig. 8(f), we can see that cCbl has much impact on PDGFR but little impact on the other molecules. This is because after PDGFR activates SHP2 and PI3K, the states of the other molecules are determined by SHP2 and PI3K. Furthermore, from the green curve of Fig. 8(b), Fig. 8(c) and Fig. 8(e), we see that SHP2 has great effect on the status of the molecules on the MAPK pathway. However, the more downstream the molecule is, the less prominent the effect becomes (the effect to MEK12 is less than that to Ras). Due to the influence of the positive feedback from PI3K and PIP3, the molecules in the MAPK pathway can also become active without SHP2, which is the result from positive crosstalk interactions. Besides, the red curves of Fig. 8(d) and Fig. 8(f) show that the influence of PI3K to the molecules in the PI3K/Akt pathway is still present due to the positive feedback from Ras.

We analyze the long-run probability of the molecules being active after PDGFL stimulation (shown in Tab. 1). The result demonstrates the activities of each molecule in this model in WildType condition. In the table, we observe that the two downstream signaling molecules, MEK12 and Akt, have high long-run probabilities, i.e., 0.77 and 0.82, respectively. This is sensible in the real biology as these two signaling molecules should be at sufficiently high concentrations after PDGF ligand activation in order to push the cell toward proliferation and to keep the cell survive from apoptotic process. In addition, we also observe high steady state probabilities on PDK and MTOR which might be a result from the accumulative effect from the basal activity and the positive feedback loop, in addition to the activation from PDGF ligand. In the mutants situations, we expect to observe different steady state probability of each molecule which would then affect the cell fate decision differently in each condition.

PRISM supports reward properties (see Sect. 4). Fig. 9 shows the expected time of six molecules being active by time instant t. These six molecules are the same as in Fig. 8. All the six curves tend to be linear after time instant 12, which shows that the 6 molecules start to be in a steady state after time instant 12.

Molecule	Probability	Molecule	Probability
PDGFR	0.22	Grb2SOS	0.55
SHP2	0.45	Ras	0.72
Gab2SOS	0.53	MEK12	0.77
c-Raf	0.63	PIP3	0.72
PI3K	0.62	Akt	0.82
PDK	0.83	MTOR	0.84
cCbl	0.47	PPX	0

Table 1. PRISM verification results: Steady state probabilities of molecules

5.2 Analyzing the influence of the crosstalk reactions

For the second goal, to analyze the influence of the crosstalk reactions, the experiment is also performed as in silico genetics. After developing the base model, we get the model variants by removing one crosstalk arrow at a time. In total, there are four positive crosstalk reactions (green arrows) and two negative crosstalk reactions (red arrows pointing from Akt). In this analysis, we focus on the positive crosstalk from Ras to PI3K and the negative crosstalk from Akt to c-Raf. More precisely, we develop three models in this experiment. The first one is the base WildType model with all reactions included. The second one is the RasPI3KMutant model, in which the positive feedback from Ras to PI3K is removed. The last one is the Aktc-RafMutant model, in which the negative feedback from Akt to c-Raf is removed. After building the three models, we analyzed the influence of the two crosstalks by comparing the WildType model to the two mutant models, respectively.

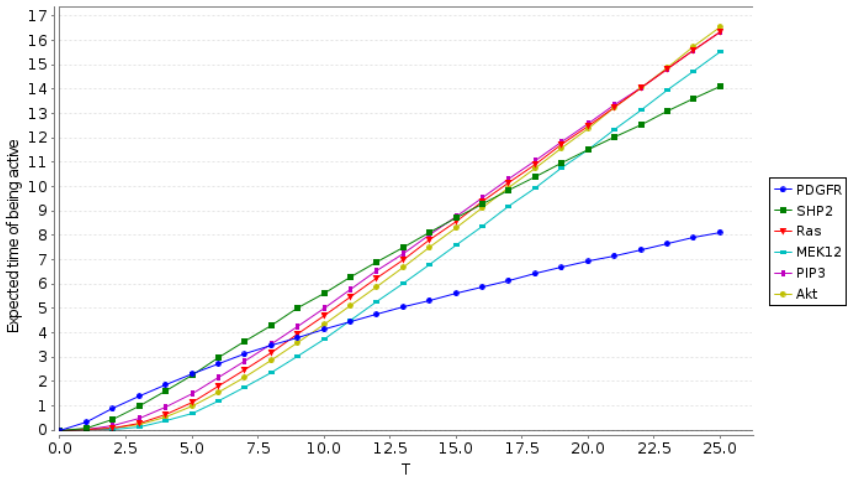
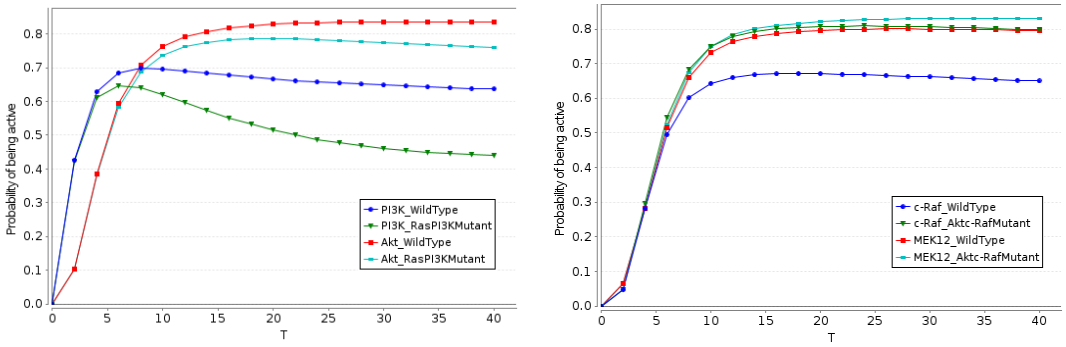


Fig. 9. PRISM verification results: Expected time of being active by time t



(a) Positive crosstalk from `Ras` to `PI3K` (b) Negative crosstalk from `Akt` to `c-Raf`

Fig. 10. PRISM verification results: Influence of crosstalks

Fig. 10 shows the results of the comparison. In both sub-figures, we compare the molecules which are directly related to the crosstalk arrows and the molecules at the end of the related main signaling pathways. In Fig. 10(a), we compare in each model the probabilities for molecules `PI3K` and `Akt` being active, which is at the end of the PI3K/Akt pathway; while in Fig. 10(b) we have chosen the molecules `c-Raf` and `MEK12`, which is at the end of the MAPK pathway. The dark blue and green curves in Fig. 10(a) show how the green arrow influences the status of `PI3K`. If there is no positive crosstalk, the probability of `PI3K` being active (green curve) becomes smaller. The two curves are coincident for a time period because the molecule `Ras` needs some time to become active before it can activate `PI3K`. The red and light blue curves show that the influence of the positive feedback on `Akt` is smaller than that on `PI3K`. Just like the situation in Fig. 8(b), Fig. 8(c) and Fig. 8(e), the more downstream the molecules are, the smaller the influence becomes. The results in Fig. 10(b) are similar: the dark blue and green curves show that the influence the negative crosstalk gives to `Ras` is significant; while the positive crosstalk has little influence on `MEK12`.

5.3 Analyzing the importance of individual reactions

For the last goal to analyze the importance of individual reactions on downstream signaling molecules, we compute the steady state probabilities of `Akt` and `MEK12` in 31 different models (called an *edge knockout study*), each of which is obtained by removing one reaction from the `wildtype` model. For example, the '`PIP3-Gab2SOS`' model is obtained by removing reaction No. 20 in Fig. 2). The results are shown in Fig. 11. In parallel to the edge knockout study, we have also computed the steady state probability of `Akt` and `MEK12` in 17 different models (called a *node knockout study*), each of which is obtained by removing all related interactions around each node. For instance, the interactions between `PDGFR`, `H-Ras`, `PIP3` and `MEK12` to `PI3K` are all omitted in the '`PI3K-KO`' model. The results are shown in Fig. 12. By adding a horizontal and a vertical line through the dot of the wildtype model, we divide these figures into four areas. Apparently, the dots in different areas show that the removed reactions/nodes have different influences on the steady state probabilities of `Akt` and `MEK12`. For instance,

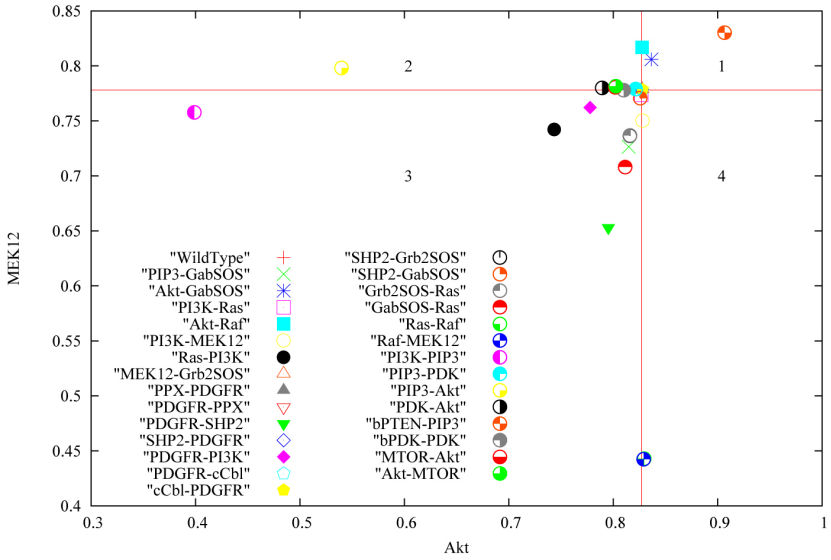


Fig. 11. PRISM verification results: Steady state probabilities of `Akt` and `MEK12` in different edge knockout models

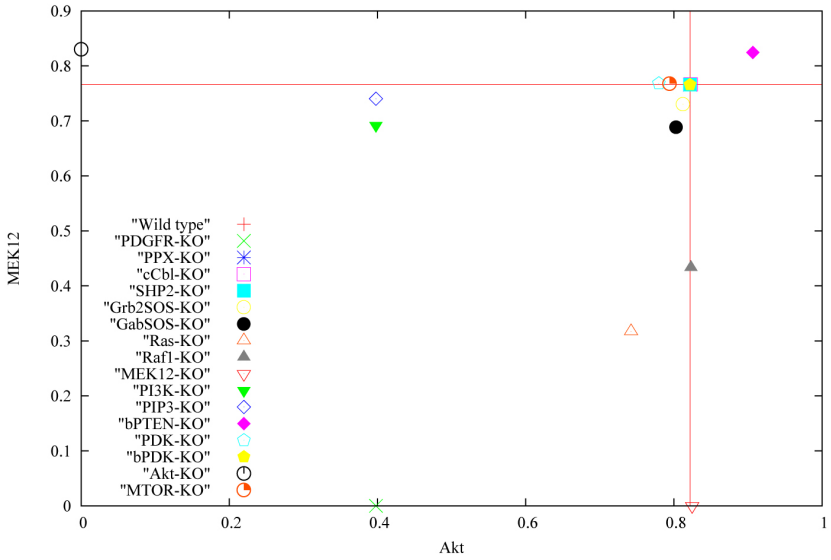


Fig. 12. PRISM verification results: Steady state probabilities of `Akt` and `MEK12` in different node knockout models

the reactions (corresponding to the dots) in area 3 can decrease the steady state probabilities of both `Akt` and `MEK12`, while those in area 1 can increase both probabilities. The reactions in area 2 can decrease the steady state probability of `Akt` and increase the one of `MEK12`; while reactions grouped in area 4 lead to the opposite effect. For those dots lying on the horizontal line, the corresponding removed reactions have little

impact on the steady state probability of MEK12. Similarly, the reactions on the vertical line have little impact on Akt.

Comparing the two studies, the node knockout study is more relevant to biological observations, which could be observed under the kinase inhibitor treatment scheme. For instance, the 'PI3K-KO' model represents the cell system which has been treated with PI3K inhibitors such as Wortmannin or LY294002 drugs [46]. Moreover, these observations have biological implications as well. Akt and MEK12 are downstream molecules in the signal transduction process that regulate different cellular functions. The signal from Akt keeps the cells to survive from apoptosis and the signal from MEK12 regulates the cells' growth and proliferation. In cancer, both of these two main pathways (see Fig. 1) are more active so they drive the cells to keep growing and dividing in an uncontrolled manner. Therefore, if we could find the targets to control these two signaling molecules to be at a desirable level, it would be beneficial for cancer therapeutic development.

6 Comparing Stochastic Verification and ODE Simulation

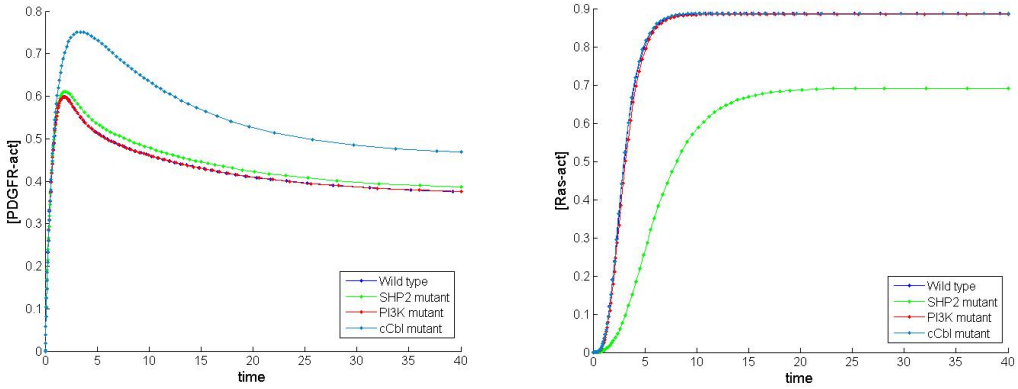
In general, stochastic verification and ODE simulation produce comparable results, e.g., see [18]. To compare these approaches for the PDGF signaling pathway, we have conducted some ODE simulations as well, where the normalized initial concentrations of the molecules are set either to 0 or 1. Moreover, we have only one instance of each molecule in the PRISM model. Thus, we would expect that the steady state probabilities obtained from PRISM stochastic verification are comparable to the steady state values (concentrations) obtained from ODE simulation.

Two particular results are shown in Fig. 13(a) and Fig. 13(b), which correspond to Fig. 8(a) and Fig. 8(c). We can see that the dynamic behavior of the two molecules, PDGFR and Ras, are similar between the two methods. However, considerable quantitative differences have been observed between the two methods in the node knockout study as shown in Tab. 2. For instance, the difference of the steady state probabilities/values for Akt is roughly 0.18 in the PDGFR-KO and PI3K-KO scenarios.

Model	Akt (PRISM)	Akt (ODE)	MEK12 (PRISM)	MEK12 (ODE)
WildType	0.8217	0.8993	0.7660	0.8849
PDGFR-KO	0.3979	0.5748	0	0
PPX-KO	0.8217	0.8993	0.7660	0.8849
MEK12-KO	0.8245	0.8993	0	0
PI3K-KO	0.3979	0.5748	0.6918	0.8724

Table 2. Steady state probabilities (PRISM verification) and steady state values (ODE simulations) of Akt and MEK12

In order to better understand the source of these differences, we investigated the possible role of different model structures, parameter values, and pathway length. The



(a) PDGFR (b) Ras

Fig. 13. ODE simulation results: Concentrations of molecules over time

prototypical model structures that we analyzed comprise a negative feedback loop as shown in Fig. 14(b) and Fig. 15(b), a positive feedback loop as shown in Fig. 14(c), and a linear chain (cascade) of molecules as shown in Fig. 15(a) and Fig. 15(b). Each molecule, except from the input node 'A', has two states active and inactive, to resemble the PDGF model. The impact of different parameter values was investigated by varying parameter 'k' in Fig. 14(c).

According to Fig. 14(d) and Fig. 15(c), we find that model structures with positive or negative feedbacks only have a minor contribution to the differences between the steady state probabilities from PRISM verification and steady state values from ODE simulation. Moreover, long-chain negative feedback loops which are generally present in biological systems, e.g., as shown in Fig. 15(b), do not cause such high differences either. On the other hand, when we consider the model with a positive feedback loop in Fig. 14(c), the differences between the two methods highly increase once the parameter values for the input node vary (see Fig. 14(e)). We observe that the decrease of parameter k in the model in Fig. 14(c) is correlated to the increase of the differences between the two methods. The steady state probabilities (PRISM verification) and steady state values (ODE simulation) for nodes B1 and B2 of the model in Fig. 14(c) with different k values are presented in Fig. 14(f) and Fig. 14(g). We can observe that the steady state probabilities from PRISM verification change drastically with different k values for both nodes, while the steady state values from ODE simulation do not change that much. This is essentially due to the different ways to compute steady state probabilities for CTMCs [47, 29] and to compute steady state values for ODE [48].

Thus, we see that in our setting the difference between the two methods mainly depends on parameter values and to some extent also on the model structure which correlates to the findings in the PDGF model. Here, we observed large differences e.g. for Akt in the PDGFR−KO conditions in the node knockout study as shown in Tab. 2. Under this condition Akt is activated only by PDK (from bPDK) and positive feedback from MTOR with a low parameter value of 0.1. This leads to a larger difference between the two methods, in contrast to the wild-type situation with larger parameter values.

This was in addition verified by analyzing model structures and parameters equal to the positive feedback loop in the PDGF model (which is not detailed in this paper).

In conclusion, we have demonstrated by examples the impact of specific parameter values and model structure on the differences between ODE simulation and stochastic verification. It has been stated that the reason for these differences might be that the approach based on verification leads to more accurate results when the number of molecules is small [49]. Nevertheless, the influence of different parameter values needs to be addressed in future research.

7 Conclusion and Future Work

In this paper, we have given a detailed description of using the probabilistic model checker PRISM to study the PDGF signaling pathway. Based on intensive literature reviews, we have built a PDGF signal transduction model in ODE format consisting of 17 molecule species and we converted it for the analysis in the probabilistic model checker PRISM. We have analyzed the dynamics of PDGF induced signaling, the influence of crosstalk reactions in the signaling pathway and the importance of individual reactions on downstream signaling molecules. Our experimental results show that stochastic verification can provide us with a good understanding of the PDGF signaling pathway, especially the result discussed in the end of Sect. 5.3 potentially can give rise to better behavior prediction of the pathway.[1]

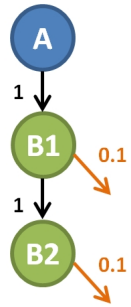
Furthermore, we have also investigated the differences of the results of the two analysis methods, ODE simulation and stochastic verification (PRISM) as discussed in Sect. 6. We have demonstrated that the two methods can predict the results differently, especially when parameter values are small. Such results will bring more insights to determine which method is more suitable for the analysis of biological systems.

Nevertheless, as indicated at the end of Sect. 3, our model is not intended to be fully accurate. The reaction rates, especially the crosstalk reaction rates, are based on literature reviewing and are only relative values. Currently, experiments in a biological laboratory are performed to get more precise values for these reaction rates. We hope techniques like parametric verification for Markov chains (e.g., [50–52]) can help to synthesize values which are consistent with the lab experimental results.

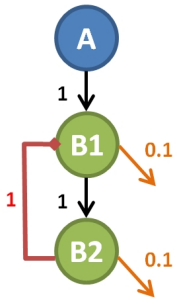
References

1. Yuan, Q., Pang, J., Mauw, S., Trairatphisan, P., Wiesinger, M., Sauter, T.: A study of the PDGF signaling pathway with PRISM. In: Proc. 3rd Workshop on Computational Models for Cell Processes. Volume 67 of EPTCS. (2011) 65–81
2. Regev, A., Shapiro, E.: Cellular abstractions: Cells as computation. Nature **419** (2002) 343
3. Bonzanni, N., Anton Feenstra, K., Fokkink, W.J., Krepska, E.: What can formal methods bring to systems biology? In: Proc. 2nd World Congress on Formal Methods. Volume 5850 of LNCS., Springer (2009) 16–22

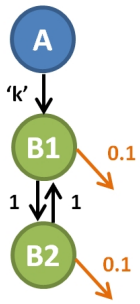
[1] The PRISM model and property specifications can be found at `http://satoss.uni.lu/jun/models/PDGF.zip`.



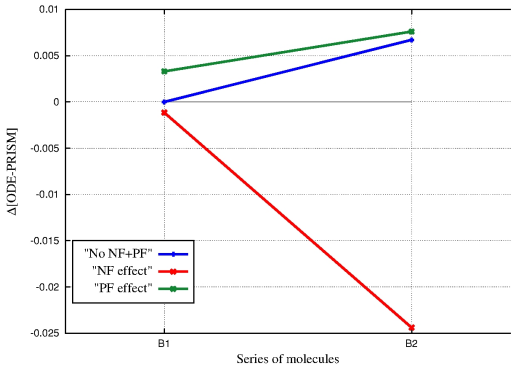
(a) Linear chain



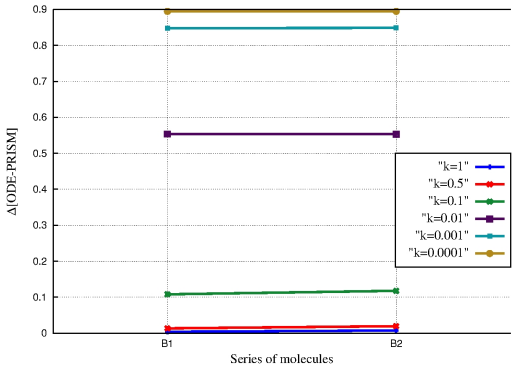
(b) Negative feedback loop



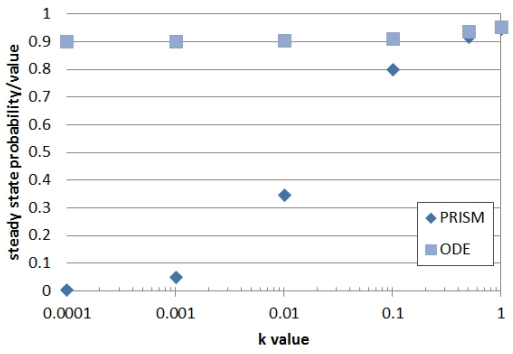
(c) Positive feedback loop



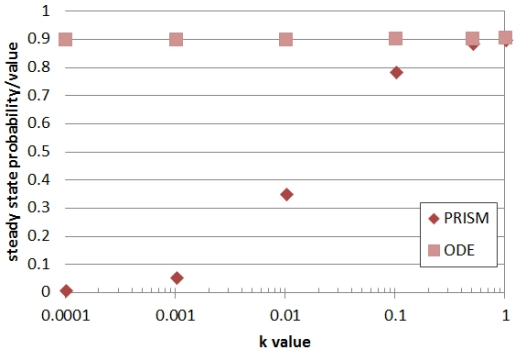
(d) Differences between the steady state probabilities (PRISM verification) and steady state values (ODE simulation) for the models in Fig. 14(a)-14(c) with k=1



(e) Differences between the steady state probabilities (PRISM verification) and steady state values (ODE simulation) for the model in Fig. 14(c) with different k values

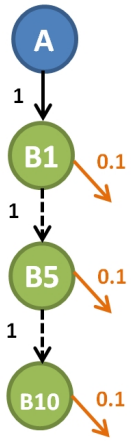


(f) The steady state probabilities (PRISM verification) and steady state values (ODE simulation) for node B1 in the model of Fig. 14(c) with different k values

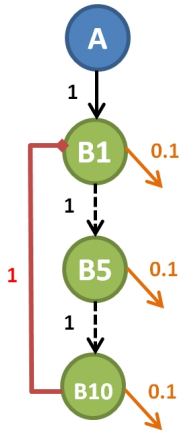


(g) The steady state probabilities (PRISM verification) and steady state values (ODE simulation) for node B2 in the model of Fig. 14(c) with different k values

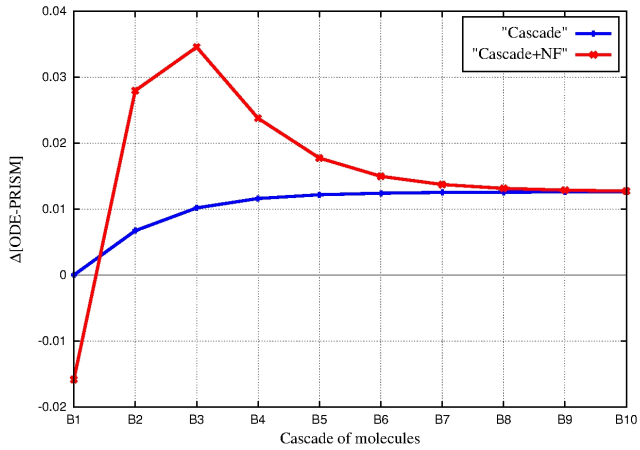
Fig. 14. Influence of model structures and parameter values on differences between PRISM verification and ODE simulation (PF: positive feedback; NF: negative feedback)



(a) Linear chain of 10 molecules



(b) Linear chain with negative feedback loop



(c) Differences between the steady state probabilities (PRISM verification) and steady state values (ODE simulation) for the models in Fig. 15(a)-15(b)

Fig. 15. Influence of pathway length on differences between PRISM verification and ODE simulation (NF: negative feedback)

4. Fisher, J., Henzinger, T.A.: Executable cell biology. Nature Biotechnology **25**(11) (2007) 1239–1249
5. Sadot, A., Fisher, J., Barak, D., Admanit, Y., Stern, M.J., Jane Albert Hubbard, E., Harel, D.: Toward verified biological models. IEEE/ACM Transactions on Computational Biology Bioinformatics **5**(2) (2008) 223–234
6. Fisher, J., Piterman, N.: The executable pathway to biological networks. Briefings in Functional Genomics and Proteomics **9**(1) (2010) 79–92
7. Kwiatkowska, M.Z., Norman, G., Parker, D.: Stochastic model checking. In: Formal Methods for the Design of Computer, Communication and Software Systems: Performance Evaluation. Volume 4486 of LNCS., Springer (2007) 220–270
8. Baier, C., Katoen, J.P.: Principles in Model Checking. MIT Press (2008)
9. Hart, S., Sharir, M., Pnueli, A.: Termination of probabilistic concurrent programs. ACM Transactions on Programming Languages and Systems **5**(3) (1983) 356–380
10. Kwiatkowska, M.Z., Norman, G., Parker, D.: Quantitative Verification Techniques for Biological Processes. In: Algorithmic Bioprocesses. Springer (2009) 391–409

11. Kwiatkowska, M.Z., Norman, G., Parker, D.: Probabilistic Model Checking for Systems Biology. In: Symbolic Systems Biology. Jones and Bartlett (2010) 31–59
12. Aldridge, B.B., Haller, G., Sorger, P.K., Lauffenburger, D.A.: Direct lyapunov exponent analysis enables parametric study of transient signalling governing cell behaviour. Systems Biology **153**(6) (2006) 425–432
13. Yu, J., Ustach, C., Kim, H.R.: Platelet-derived growth factor signaling and human cancer. Journal of Biochemistry and Molecular Biology **36**(1) (2003) 49–59
14. Golub, T.R., Barker, G.F., Lovett, M., Gilliland, D.G.: Fusion of PDGF receptor beta to a novel ets-like gene, tel, in chronic myelomonocytic leukemia with t(5;12) chromosomal translocation. Cell **77**(2) (1994) 307–316
15. Anan, K., Morisaki, T., Katano, M., Ikubo, A., Kitsuki, H., Uchiyama, A., Kuroki, S., Tanaka, M., Torisu, M.: Vascular endothelial growth factor and platelet-derived growth factor are potential angiogenic and metastatic factors in human breast cancer. Surgery **119**(3) (1996) 333–339
16. Cools, J., DeAngelo, D.J., Gotlib, J., Stover, E.H., Legare, R.D., Cortes, J., Kutok, J., Clark, J., Galinsky, I., Griffin, J.D., Cross, N.C., Tefferi, A., Malone, J., Alam, R., Schrier, S.L., Schmid, J., Rose, M., Vandenberghe, P., Verhoef, G., Boogaerts, M., Wlodarska, I., Kantarjian, H., Marynen, P., Coutre, S.E., Stone, R., Gilliland, D.G.: A tyrosine kinase created by fusion of the PDGFRA and FIP1L1 genes as a therapeutic target of imatinib in idiopathic hypereosinophilic syndrome. The New England journal of medicine **348**(13) (2003) 1201–1214
17. Miettinen, M., Lasota, J.: Gastrointestinal stromal tumors: Pathology and prognosis at different sites. Seminars in Diagnostic Pathology **23**(2) (2006) 70–83
18. Calder, M., Vyshemirsky, V., Gilbert, D., Orton, R.: Analysis of signalling pathways using continuous time Markov chains. Transactions on Computational Systems Biology VI **4220** (2006) 44–67
19. Kwiatkowska, M.Z., Norman, G., Parker, D.: Using probabilistic model checking in systems biology. SIGMETRICS Performance Evaluation Review **35** (2008) 14–21
20. Pronk, T., de Vink, E., Bosnacki, D., Breit, T.: Stochastic modeling of codon bias with PRISM. In: Proc. 3rd Workshop on Methods and Tools for Coordinating Concurrent, Distributed and Mobile Systems, Computer Science Department, University of Cyprus, Nicosia (2007)
21. Bosnacki, D., Pronk, T., de Vink, E.: In silico modelling and analysis of ribosome kinetics and aa-tRNA competition. Transactions on Computational Systems Biology XI **5750** (2009) 69–89
22. Heath, J., Kwiatkowska, M., Norman, G., Parker, D., Tymchyshyn, O.: Probabilistic model checking of complex biological pathways. Theoretical Computer Science **319**(3) (2008) 239–257
23. Owens, N., Timmis, J., Greensted, A., Tyrrell, A.: Modelling the Tunability of early T cell signalling events. In: Proc. 7th Conference on Artificial Immune Systems. Volume 5132 of LNCS. (2008) 12–23
24. Jha, S.K., Clarke, E.M., Langmead, C.J., Legay, A., Platzer, A., Zuliani, P.: A bayesian approach to model checking biological systems. In: Proc. 7th Conference on Computational Methods in Systems Biology. Volume 5688 of LNCS., Springer (2009) 218–234
25. Liò, P., Merelli, E., Paoletti, N.: Multiple verification in computational modeling of bone pathologies. In: Proc. 3rd International Workshop on Computational Models for Cell Processes. Volume 68 of EPTCS. (2011) 82–96
26. Zhang, R., Shah, M.V., Yang, J., Nyland, S.B., Liu, X., Yun, J.K., Albert, R., Loughran, T.P.: Network model of survival signaling in large granular lymphocyte leukemia. Proceedings of the National Academy of Sciences of the United States of America **105**(42) (2008) 16308–16313

27. Wang, C.C., Cirit, M., Haugh, J.M.: PI3K-dependent cross-talk interactions converge with Ras as quantifiable inputs integrated by Erk. Molecular Systems Biology **5** (2009) 246
28. Aziz, A., Sanwal, K., Singhal, V., Brayton, R.: Model checking continuous time markov chains. ACM Transactions on Computational Logic **1**(1) (2000) 162–170
29. Baier, C., Hermanns, H., Haverkort, B.R., Katoen, J.P.: Model-checking algorithms for continuous-time markov chains. IEEE Transactions on Software Engineering **29**(6) (2003) 524–541
30. Krauss, G.: Biochemistry of Signal Transduction and Regulation. Wiley-VCH, Weinheim (2008)
31. Bhalla, U.: Understanding complex signaling networks through models and metaphor. Progress in Biophysics & Molecular Biology **81**(1) (2003) 45–65
32. Heldin, C.H., Westermarkt, B., Wasteson, A.: Platelet-derived growth factor. Molecular and Celluar Endocrinology **39**(3) (1985) 169–187
33. Tallquist, M., Kazlauskas, A.: PDGF signaling in cells and mice. Cytokine & Growth Factor Review **15**(4) (2004) 205–213
34. Yokote, K., Hellman, U., Ekman, S., Saito, Y., Roennstrand, L., Saito, Y., Heldin, C.H., Mori, S.: Identification of Tyr-762 in the platelet-derived growth factor alpha-receptor as the binding site for Crk proteins. Oncogene **16**(10) (1998) 1229–1239
35. Gutiérrez-Uzquiza, A., Arechederra, M., Molina, I., Banos, R., Maia, V., Benito, M., Guerrero, C., Porras, A.: C3G down-regulates p38 MAPK activity in response to stress by Rap-1 independent mechanisms: involvement in cell death. Cellualar Signalling **22**(3) (2010) 533–542
36. Valgeirsdóttir, S., Paukku, K., Silvennoinen, O., Heldin, C.H., Claesson-Welsh, L.: Activation of Stat5 by platelet-derived growth factor (PDGF) is dependent on phosphorylation sites in PDGF beta-receptor juxtamembrane and kinase insert domains. Oncogene **16**(4) (1998) 505–515
37. Joazeiro, C.A., Wing, S.S., Huang, H., Leverson, J.D., Hunter, T., Liu, Y.C.: The tyrosine kinase negative regulator c-Cbl as a RING-type, E2-dependent ubiquitin-protein ligase. Science **286**(5438) (1999) 309–312
38. Hanahan, D., Weinberg, R.A.: Hallmarks of cancer: The next generation. Cell **144**(5) (2011) 646–674
39. Yu, J., Gutkind, J.S., Mahadevan, D., Li, W., Meyers, K.A., Pierce, J.H., Heidaran, M.A.: Biological function of PDGF-induced PI-3 kinase activity: its role in alpha PDGF receptor-mediated mitogenic signaling. Journal of Cell Biology **127**(2) (1994) 479–487
40. Bazenet, C.E., Gelderloos, J.A., Kazlauskas, A.: Phosphorylation of tyrosine 720 in the platelet-derived growth factor alpha receptor is required for binding of Grb2 and SHP-2 but not for activation of Ras or cell proliferation. Molecular and Cellular Biology **16**(12) (1996) 6926–6936
41. Reddi, A.L., Ying, G., Duan, L., Chen, G., Dimri, M., Douillard, P., Druker, B.J., Naramura, M., Band, V., Band, H.: Binding of Cbl to a phospholipase Cgamma1-docking site on platelet-derived growth factor receptor beta provides a dual mechanism of negative regulation. Journal of Biological Chemistry **282**(40) (2007) 29336–2947
42. Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. Journal of Physical Chemistry **81**(25) (1977) 2340–2361
43. Witt, J., Barisic, S., Schumann, E., Allgöwer, F., Sawodny, O., Sauter, T., Kulms, D.: Mechanism of PP2A-mediated IKKβ dephosphorylation: a systems biological approach. BMC Systems Biology **3** (2009) 71
44. Schmidt, H., Jirstrand, M.: Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology. Bioinformatics **22**(4) (2006) 514–515
45. Saltelli, A., Tarantola, S., Chan, K.P.S.: A quantitative model-independent method for global sensitivity analysis of model output. Technometrics **41**(1) (1999) 39–56

46. Davies, S., Reddy, H., Caivano, M., Cohen, P.: Specificity and mechanism of action of some commonly used protein kinase inhibitors. Biochemical Journal **351**(Pt 1) (2000) 95–105
47. Stewart, W.J.: Introduction to the Numerical Solution of Markov Chains. Princeton University Press (1994)
48. Tenenbaum, M., Pollard, H.: Ordinary Differential Equations: An Elementary Textbook for Students of Mathematics, Engineering, and the Science. Harper & Row (1985)
49. Kwiatkowska, M.Z., Norman, G., Parker, D., Tymchyshyn, O., Heath, J., Gaffney, E.: Simulation and verification for computational modelling of signalling pathways. In: Proc. 38th Winter Simulation Conference. (2006) 1666–1674
50. Han, T., Katoen, J.P., Mereacre, A.: Approximate parameter synthesis for probabilistic time-bounded reachability. In: Proc. 29 IEEE Real-Time Systems Symposium, IEEE Computer Society (2008) 173–182
51. Hahn, E.M., Hermanns, H., Zhang, L.: Probabilistic reachability for parametric Markov models. In: Proc. 16th Spin Workshop on Model Checking Software. Volume 5578 of LNCS., Springer (2009) 88–106
52. Andreychenko, A., Mikeev, L., Spieler, D., Wolf, V.: Parameter identification for Markov models of biochemical reactions. In: Proc. 23rd Conference on Computer Aided Verification. Volume 6806 of LNCS., Springer (2011) 83–98