

Regina Samaga

**Qualitative and
Semi-Quantitative Analysis of
Signal Transduction Networks**



Qualitative and Semi-Quantitative Analysis of Signal Transduction Networks

Dissertation

zur Erlangung des akademischen Grades

**Doktoringenieurin
(Dr.-Ing.)**

von Dipl. Biomathematikerin Regina Samaga, geb. Reiner,
geboren am 18.02.1982 in Dachau.

Genehmigt durch die Fakultät für Elektrotechnik und Informations-
technik der Otto-von-Guericke-Universität Magdeburg.

Gutachter:

Prof. Dr.-Ing. Rolf Findeisen

Dr.-Ing. Steffen Klamt

Prof. Dr.-Ing. Thomas Sauter

Promotionskolloquium am 4. Juli 2018

Forschungsberichte aus dem Max-Planck-Institut
für Dynamik komplexer technischer Systeme

Band 51

Regina Samaga

**Qualitative and Semi-Quantitative Analysis
of Signal Transduction Networks**

Shaker Verlag
Aachen 2018

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.d-nb.de>.

Zugl.: Magdeburg, Univ., Diss., 2018

Copyright Shaker Verlag 2018

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers.

Printed in Germany.

ISBN 978-3-8440-6203-8

ISSN 1439-4804

Shaker Verlag GmbH • P.O. BOX 101818 • D-52018 Aachen

Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9

Internet: www.shaker.de • e-mail: info@shaker.de

Preface

This work evolved during my time as a research assistant at the Max Planck Institute for Dynamics of Complex Technical Systems in Magdeburg. I have been lucky to encounter as a part of the Systems Biology Group headed by Prof. Ernst Dieter Gilles an inspiring and cooperative working atmosphere, which was shaped by Prof. Gilles' scientific curiosity in all the projects and by his interest in the people behind the work. Later, I was part of the research group Analysis and Redesign of Biological Networks headed by Steffen Klamt. This work would not be what it is without Steffen's continuous support. Throughout the entire time, I could count on his encouragement and in-depth discussions and on an open door and ear whenever I needed an advice.

During my early time at the Max Planck Institute, I had with Julio Saez Rodriguez a great tutor who introduced me to the logical modeling of signaling networks and to the modeling tool PROMoT. I gratefully acknowledge the support of the PROMoT Development Team, in particular Sebastian Mirschel and Katrin Kolczyk, and the support of Axel von Kamp with regards to CellNetAnalyzer issues. My special thanks go to J r my Huard, with whom I shared the office over a long time, for his help in any Linux, LaTeX, Matlab, or PROMoT topic, just to name a few. I am also very grateful to Renate Wagner, Janine Holzmann and Anke Goettert for all their effort to assist the whole group and to provide a very comfortable working atmosphere.

The last chapter of this work would not exist without the substantial support of Dietrich Flockerzi. He always found the time to recalculate and discuss all my results, to help to sharpen the mathematical formulations and to explain everything to me with a great patience. I also want to thank Carsten Conradi for all discussions and advice about Chemical Reaction Network Theory.

I would like to express my gratitude to my collaboration partners Leonidas Alexopoulos, Iannis Melas, Seong-Hwan Rho, and Tim Maiwald, for our fruitful cooperations. Special thanks to Lorenza D'Alessandro for our intensive collaboration and friendship throughout the years. Thanks for all the open discussions, your great commitment and scientific enthusiasm.

I am grateful to Prof. Rolf Findeisen, who accepted to co-advise and review this thesis.

All in all, I want to thank all my colleagues from the Max Planck Institute in Magdeburg for the enjoyable and inspiring time we had together.

My warmest thanks go to my family, for all their support, sympathy, and love.

The work presented in this thesis was part of the Systems Biology funding initiatives *HepatoSys* and *The Virtual Liver Network*, which were both financially supported by the German Federal Ministry of Education and Research.

Dachau, 2018

Regina Samaga

Contents

List of Figures	vii
List of Tables	xi
Zusammenfassung	xiii
1 Introduction	1
2 Systems Biology of Signal Transduction Networks	5
2.1 Cell Signaling	5
2.2 Liver Regeneration	6
2.3 Signaling through the EGF Ligand Family and through HGF	6
2.3.1 The Signaling System of the EGF Ligand Family	6
2.3.2 Hepatocyte Growth Factor Signaling	10
2.4 Modeling Approaches for Signaling Networks	11
2.4.1 Interaction Graphs	11
2.4.1.1 Cycles in Interaction Graphs	13
2.4.1.2 Dependency Matrix	14
2.4.1.3 Minimal Cut Sets	15
2.4.2 Logical Models of Signal Transduction Networks	16
2.4.2.1 Hypergraph Representation of Logical Models	18
2.4.2.2 Dynamical Analysis of Logical Models	20
2.4.2.3 Logical Steady State and Signal Flow Analysis	21
2.4.2.4 Minimal Intervention Sets	24
2.4.3 Ordinary Differential Equation Modeling	26
2.4.3.1 Differential Equation Models Derived from Boolean Models	26
2.4.3.2 Qualitative Analysis of Dynamical Models	27

3	Data-Driven Interrogation and Training of Signaling Network Structures Based on Interaction Graphs	33
3.1	Data Analysis Based on the Dependency Matrix	34
3.2	Detection and Removal of Inconsistencies between Data and Signaling Network Topologies Using the Concept of Sign Consistency	36
3.2.1	General Framework	38
3.2.2	Formulation as Optimization Problems	43
3.2.2.1	Basic Definitions and Formulation of Sign Consistency	43
3.2.2.2	SCEN_FIT Optimization Problem	44
3.2.2.3	MCoS Optimization Problem	45
3.2.2.4	OPT_SUBGRAPH Optimization Problem	47
3.2.2.5	OPT_GRAPH Optimization Problem	48
3.3	Application to ERBB Signaling	48
3.3.1	The Interaction Graph Model of ERBB Signaling	49
3.3.2	Analysis of High-Throughput Experimental Data	51
3.3.2.1	Data Description and Discretization	51
3.3.2.2	Analysis with the Dependency Matrix	53
3.3.2.3	Analysis Based on Sign Consistency	61
3.4	Discussion	74
4	Logical Models of Signal Transduction Networks	79
4.1	Species Equivalence Classes	79
4.2	Computation of Intervention Strategies and Failure Modes	81
4.2.1	Definition of Minimal Intervention Sets and General Algorithm	81
4.2.2	Methods for Search Space Reduction	83
4.2.2.1	Exploiting Dependencies	83
4.2.2.2	Exploiting Fault Equivalence Classes	85
4.2.2.3	Exploiting Dependencies in Intervention Goals with Multiple Targets	88
4.3	Analysis of Experimental Data	90
4.4	The Logic of ERBB Signaling	91
4.4.1	From a Stoichiometric Model to a Logical Model	92
4.4.2	Theoretical Analysis of the Logical ERBB Model	96
4.4.2.1	Qualitative Input–Output Behavior	96
4.4.2.2	Species Equivalence Classes in the ERBB Model	98

4.4.2.3	Computation of Minimal Intervention Sets in the Logical ERBB Model	98
4.4.3	Data Analysis	104
4.5	Discussion	110
5	From Structure to Dynamics: Combined Interaction Graph and ODE Modeling Unravels Network Structure of HGF Signaling	115
5.1	An Interaction Graph Model of HGF Signaling	117
5.2	Experimental Data	118
5.3	Selection of Minimal Model Structures	121
5.4	Translation into Ordinary Differential Equation Models	128
5.5	Ordinary Differential Equation Model Selection	130
5.6	Experimental Validation and Model Predictions	134
5.6.1	Experimental Validation of a Negative Crosstalk	134
5.6.2	Inhibitor Combination: Model Predictions and Experimental Validation	136
5.7	Discussion	139
6	Prediction of Qualitative Dynamics from Network Structure	141
6.1	Perturbation Analysis for General ODE Systems	142
6.1.1	General Framework	142
6.1.2	Determinants and Graphs	146
6.1.3	A Graph-Theoretic Criterion for Initial Response	149
6.1.4	Determinant Criteria for Ultimate System Response	149
6.1.5	Some Graph-Theoretic Implications	153
6.2	Perturbation Analysis for Chemical Reaction Networks	155
6.2.1	Dynamic Chemical Reaction Networks	156
6.2.2	Determinant Criteria for Ultimate Response of Chemical Reaction Networks	160
6.2.3	Graphical Representation of Dynamic Chemical Reaction Networks	166
6.2.3.1	Labeled Interaction Graphs and R-Subgraphs	167
6.2.3.2	Species–Reaction Graphs and Directed Species– Reaction Graphs	170
6.2.3.3	Interrelation between Labeled Interaction Graphs and Directed Species–Reaction Graphs	175

6.2.4	Necessary Motifs for Inverse Response	176
6.2.5	Relation to Multistationarity	186
6.3	Discussion	191
7	Conclusions	195
A	Integer Linear Programming Formulations for Interaction Graph Training	197
A.1	Implementation.	197
A.2	ILP Formulation of Sign Consistency.	197
A.3	ILP Formulation of SCEN_FIT	199
A.4	ILP Formulation of MCoS	200
A.5	ILP Formulation of OPT_SUBGRAPH	201
B	Model Compression for Interaction Graphs	203
C	Documentation of ERBB Models	205
C.1	List of Species	205
C.2	List of Interactions of the Logical ERBB Model (M1)	211
C.3	Interactions that are only included in ERBB interaction graph model	227
C.4	Description of Model M2	228
D	Sensitivities of Binarization of HepG2 Data to Chosen Parameters	229
E	Documentation of the HGF Interaction Graph Master Model	233
E.1	List of Species	233
E.2	List of Interactions of the HGF Core Model	235
E.3	List of Candidate Mechanisms	237
F	Documentation of HGF ODE Models	239
F.1	Reactions in the ODE Models	239
F.1.1	Reactions in the Core Model	239
F.1.2	Candidate mechanisms	241
F.2	Parameter Names and Values of the Final Model Structure	242
G	Proofs from Chapter 6	245
	Bibliography	249

List of Figures

1.1	Different modeling formalisms for cellular signaling	4
2.1	Interaction graph of the EGF/NRG1 network example	9
2.2	Dependency matrix of the EGF/NRG1 example model	16
2.3	Logical model of the EGF/NRG1 network example	20
3.1	A simple example network used for illustration purposes.	37
3.2	Interaction Graph Model of ERBB Signaling	50
3.3	Data set of TGF α -stimulated primary human hepatocytes and HepG2 cells	52
3.4	Data set from HepG2 cells, stimulated with five different EGF-like ligands	54
3.5	Dependency matrix of ERBB interaction graph model, main activation routes	56
3.6	Comparison between experimental data from primary hepatocytes and ERBB model predictions derived from the dependency matrix	57
3.7	Comparison between experimental data from HepG2 cells and ERBB model predictions derived from the dependency matrix	58
3.8	Interaction graph model of the ERBB signaling network, processed for ILP approach.	64
3.9	Discretized measurements of the 16 considered experimental scenarios and the resulting SCEN_FIT solutions computed from the ERBB interaction graph model.	66
3.10	Cumulative fitting error of optimal SCEN_FIT solutions over all 16 scenarios in the ERBB network as a function of the two discretization thresholds	67
3.11	Optimal model structures derived from the compressed ERBB model by applying OPT_SUBGRAPH and OPT_GRAPH procedures	69
3.12	Discretized data and the two SCEN_FIT solutions that result from the optimal subgraphs given in Figure 3.11(a).	72

3.13	Comparison of the fitting errors of the initial model structure and of the optimal interaction graph	73
4.1	Pseudocode for the computation of minimal intervention sets	89
4.2	Examples illustrating the translation of the stoichiometric EGFR model into a logical description.	93
4.3	Logical model of the ERBB receptor signaling pathway represented in PROMoT.	95
4.4	Species equivalence classes in the logical ERBB model.	99
4.5	Sensitivities of the binarization to the chosen parameters (primary human hepatocytes)	105
4.6	Comparison of the discretized data with predictions from the logical model	107
4.7	Comparison of the discretized data with predictions from model M2 .	108
5.1	Workflow of model selection strategy	116
5.2	Interaction graph master model	119
5.3	Experimental results	120
5.4	Discretized experimental data	121
5.5	Predictions by interaction graph models	123
5.6	Pseudocode for the selection of minimal model structures from the interaction graph master model	125
5.7	Pseudocode for the comparison of model predictions with discretized data	126
5.8	Selected minimal model structures, core and complete model	127
5.9	Underlying interaction graph of the complete ODE model	129
5.10	ODE model selection	131
5.11	ODE model fit	133
5.12	Negative crosstalk: experimental validation	136
5.13	Model predictions of inhibitor combinations	137
5.14	Inhibitor combinations: experimental validation	138
6.1	Scheme illustrating initial, inverse, and compensatory response	144
6.2	Scheme of dual phosphorylation and dephosphorylation mechanism .	163
6.3	Chemical reaction network with bimolecular reaction	167
6.4	Labeled interaction graph and R-subgraphs of order 3 for the example system (6.36)	171

6.5	SR graphs for the CRN example (6.36)	174
6.6	DSR graph for the CRN example (6.37)	175
6.7	R-subgraphs corresponding to amplifying motifs for the CRN example (6.36) taken with NAC kinetics	190
6.8	R-subgraphs for the CRN example (6.36) taken with weakly monotonic kinetics	192
B.1	Basic network compression rules for interaction graphs	204
D.1	Sensitivity to Parameter for Relative Significance	229
D.2	Sensitivity to Parameter for Absolute Significance	230
D.3	Sensitivity to Parameter for Experimental Noise	231

List of Tables

2.1	Logical steady states in the EGF/NRG1 example model	23
2.2	Minimal intervention sets in the EGF/NRG1 logical model	25
3.1	Example scenarios and optimizations for the example network in Figure 3.1	40
3.2	Proposed ERBB model changes to improve agreement between interaction graph model structure and experimental data.	61
3.3	MCoS for scenario 11 in Figure 3.9.	68
3.4	Optimal subgraphs that fit the discretized data shown in Figure 3.9. . .	70
3.5	Suggestions for new single edges as computed by OPT_GRAPH.	72
4.1	Species equivalence classes in the Logical ERBB models M1 and M2. . .	100
4.2	Minimal intervention sets of size 2 to activate ERK1/2 and AKT in the ERBB model	101
4.3	Benchmark tests showing the power of reduction techniques for the computation of minimal intervention sets	103
4.4	Proposed ERBB model changes to improve agreement between model structure and experimental data.	111
5.1	Candidate edges of selected models, model combinations, and random models.	135
6.1	Example 6.51, analysis of initial and ultimate response to positive perturbation in E_1	166

Zusammenfassung

An den Signaltransduktionsvorgängen in der Zelle ist eine Vielzahl von Molekülen beteiligt, die sich gegenseitig beeinflussen. Diese Moleküle bilden komplexe, hoch regulierte Signalnetzwerke. Um ein ganzheitliches Verständnis von diesen Netzwerken zu erhalten, werden Methoden der Systembiologie eingesetzt, welche mathematische und computergestützte Methoden mit experimentellen Daten kombinieren. Die experimentellen Daten, die zur Verfügung stehen, ermöglichen es zum Teil, detaillierte quantitative Informationen zu erhalten; manchmal liefern sie jedoch auch nur ein qualitatives Bild. Bei manchen Signalwegen ist die Netzwerktopologie bereits im Detail bekannt, bei anderen ist diese gänzlich unbekannt. Um diesen Unterschieden in der Qualität der vorliegenden Information zu begegnen, wurden in den letzten Jahren Modellierungsmethoden von unterschiedlicher Komplexität entwickelt.

Physiko-chemische Modellierungsmethoden, beispielsweise gewöhnliche Differentialgleichungssysteme mit Massenwirkungskinetiken, erlauben eine detaillierte Beschreibung des zeitlichen Systemverhaltens. Diese Art der Modellierung erfordert eine gute Kenntnis der biologischen Mechanismen und kinetischen Parameter, so dass die Anwendung auf kleinere und bereits relativ gut beschriebene Netzwerke beschränkt ist.

Im Gegensatz dazu stehen qualitative Modellierungsmethoden, die hauptsächlich die Netzwerkstruktur beschreiben und keine Information über die kinetischen Parameter benötigen. Somit sind diese Methoden im Allgemeinen für großskalige Netzwerke geeignet. Einfache Graphenmodelle, bei denen die biologischen Species als Knoten und die Interaktionen zwischen den Species als Kanten dargestellt werden, wurden bisher hauptsächlich zur Beschreibung topologischer Eigenschaften von Netzwerken mit bis zu mehreren Tausend Proteinen eingesetzt. Erweiterte Graphenmodelle, wie zum Beispiel Constraint-based Modellierung, Petrinetze oder Logische Netzwerke leiten sich auch rein von der Netzwerkstruktur ab, ermöglichen jedoch die Analyse wichtiger funktionaler Eigenschaften und ermöglichen darüberhinaus bestimmte Vorhersagen zum qualitativen Systemverhalten.

Parameterfreie Methoden im Kontext der gewöhnlichen Differentialgleichungssy-

steme haben zum Ziel, Aussagen über das qualitative dynamische Verhalten eines Systems zu treffen. Eine typische Frage ist beispielsweise, ob die Struktur eines gegebenen Differentialgleichungssystems ein bestimmtes dynamisches Verhalten ermöglicht, das heißt ob es Parameterwerte gibt, mit denen das System beispielsweise mehrere stationäre Lösungen besitzt oder Oszillationen zeigt. Auch wenn diese Methoden parameterfrei sind, so ist dennoch ein detailliertes mechanistisches Verständnis der einzelnen Reaktionen nötig.

In der vorliegenden Arbeit wurden mathematische Modellierungsmethoden entwickelt, um die Netzwerkstruktur von zellulären Signaltransduktionswegen anhand experimenteller Daten und qualitativer Beobachtungen der Dynamik zu analysieren. Dabei wurden drei verschiedene Modellierungsformalismen eingesetzt: Interaktionsgraphen, logische Modelle und gewöhnliche Differentialgleichungssysteme. Diese drei Formalismen hängen eng miteinander zusammen. Ein Interaktionsgraph beschreibt paarweise Zusammenhänge zwischen biologischen Species. Ein logisches Modell kann von einem Interaktionsgraph abgeleitet werden, indem logische Regeln ergänzt werden, die beschreiben, wie verschiedene Eingänge an einem Knoten des Graphen kombiniert werden. Bestimmte Systemeigenschaften eines gewöhnlichen Differentialgleichungssystems können anhand des zugrundeliegenden Interaktionsgraphen abgeleitet werden, welcher die Vorzeichenstruktur der Jacobi-Matrix des Systems repräsentiert. In dieser Arbeit wird gezeigt, dass insbesondere die Kombination der verschiedenen Modellierungsmethoden von unterschiedlicher Komplexität es ermöglicht, wichtige Erkenntnisse über die Struktur und Funktion von Signaltransduktionswegen zu erhalten. Als Anwendungsbeispiele wurden Modelle der Signalwege des Epidermalen Wachstumsfaktors (EGF) und des Hepatozyten-Wachstumsfaktors (HGF) erstellt und analysiert, was neue Einblicke in diese Signalwege ermöglichte. Beide Wege spielen eine zentrale Rolle bei der Leberregeneration.

Basierend auf Informationen aus der Literatur wurde ein großskaliges Modell der Signalwege erstellt, die von der Familie der EGF Liganden aktiviert werden. Dieses Interaktionsgraph-Modell wird in dieser Arbeit präsentiert. Neue Methoden und Algorithmen zur Analyse der Netzwerkstruktur – basierend auf Interaktionsgraphen – werden beschrieben. Die Anwendung dieser Methoden auf das EGF Modell zeigt, dass Abweichungen zwischen den experimentellen Daten und der kanonischen Netzwerkstruktur bestehen, welche zum Teil Zelltyp-abhängig sind. Dies erlaubte die Formulierung von neuen Hypothesen. Außerdem werden Erweiterungen eines bestehenden Formalismus zur statischen Analyse von logischen Modellen präsentiert. Diese beinhalten unter anderem neue Algorithmen zur Berechnung

von Interaktionsstrategien. Zudem wird gezeigt, wie das qualitative Input–Output Verhalten automatisiert mit diskretisierten experimentellen Daten verglichen werden kann. Die Anwendung dieser neuen Methoden auf ein ebenfalls innerhalb dieser Arbeit erstellten logischen Modells ergänzt die Ergebnisse der Analyse des Interaktionsgraph-Modells. In einem neuen Modellierungsansatz, der in dieser Arbeit präsentiert wird, werden qualitative und quantitative Modellierungsmethoden miteinander kombiniert. Gezeigt wird die Analyse der HGF-induzierten Aktivierung zweier zentraler Signaltransduktionswege. Eine datengetriebene Analyse ermöglicht es, aus einer Vielzahl von möglichen Modellstrukturen, die sich aus der Kombination verschiedener Crosstalk- und Feedback-Mechanismen ergeben, eine Vorselektion verschiedener Interaktionsgraph-Modelle zu treffen. Eine darauffolgende Analyse mit gewöhnlichen Differentialgleichungssystemen erlaubt, die Netzwerkstruktur zu identifizieren, die das transiente dynamische Verhalten, welches in den experimentellen Daten widerspiegelt wird, am besten beschreibt. Die Differentialgleichungssysteme wurden dabei von den vorselektierten Interaktionsgraphen abgeleitet. Im letzten Teil der Arbeit werden Interaktionsgraphen analysiert, die einem gewöhnlichen Differentialgleichungssystem zu Grunde liegen. Es werden neue graphentheoretische und algebraische Methoden präsentiert, die es ermöglichen, bestimmte steady-state Antworten auf Perturbationen für chemische Reaktionsnetzwerke auszuschließen. Schließlich wird die Verbindung zu bereits bekannten strukturellen Bedingungen für mehrfach stationäre Zustände gezeigt.

1 | Introduction

Cellular signaling is made up of complex networks of interacting molecules that are tightly interconnected and regulated. In order to gain an integrated understanding of these networks, systems biology approaches combining mathematical and computational methods with experimental data are becoming increasingly important. Available experimental data might provide a detailed quantitative knowledge or just a qualitative view, and detailed prior knowledge on the network topology might or might not exist. To account for the different quality of information that is available for a network under study, modeling formalisms of different levels of complexity have been developed over the last years (de Jong 2002; Kestler et al. 2008).

Physicochemical modeling approaches, typically networks of differential equations, provide a detailed description of the biochemical processes that is based on physical and chemical theory (Aldridge et al. 2006). Most widely used are sets of coupled ordinary differential equations (ODEs) that describe the system's development over time using mass action kinetics for the rates of production and consumption of the biomolecular species (e.g., Swameye et al. (2003)). This type of modeling requires sufficient knowledge of biological mechanisms and kinetic parameters, what limits its applicability to small and well-characterized networks.

In contrast, qualitative modeling approaches are primarily based on the network structure and do not require information on the kinetic parameters. This makes them generally applicable to large-scale networks. The class of qualitative modeling approaches comprises various formalisms of different complexity. Graph models representing biological species as nodes and interactions between the species as edges are arguably the simplest mathematical description of signaling networks. They have mainly been applied to study global topological properties of networks containing up to several thousand proteins (Barabási and Oltvai 2004; Jeong et al. 2001). More refined qualitative modeling approaches include constraint-based modeling (Papin and Palsson 2004), Petri nets (Chaouiya 2007; Hardy and Robillard 2004), and logical modeling (Klamt et al. 2006; Morris et al. 2010; Wang et al. 2012). As graph models, these frameworks solely rely on the network structure, yet they enable the analysis of important functional properties of large-scale signal transduction net-

works such as input–output relationships, feedback loops, or signal transfer routes, and they also allow certain predictions, for instance, regarding the expected qualitative response to perturbations. Besides static investigations, Petri nets and logical models enable to derive qualitative properties of the system’s dynamics by means of discrete dynamic modeling (Chaouiya 2007; Wang et al. 2012). Other parameter-free approaches also aim at gaining insights into the qualitative dynamic properties of the system, however, in the context of ODE systems (Angeli and Sontag 2003; Conradi and Flockerzi 2012; Craciun and Feinberg 2005; Feinberg 1987; Radde et al. 2010). A typical question that is tackled by these approaches is whether a given ODE network structure is able to exhibit, for some parameter values, a certain qualitative behavior such as multistationarity, oscillatory behavior, or non-monotonicity. Although these approaches are parameter-free, a detailed knowledge of the involved reactions, particularly mechanistic details, is usually required, whereas Petri nets and logical models are based on a more abstract understanding.

Within this thesis, methods to interrogate network structures of cellular signaling pathways based on experimental data and qualitative dynamic observations were developed. To this end, three modeling formalisms that are of different complexity were used (Figure 1.1). As representatives of graph models, *interaction graphs* capture pairwise relationships between biological compounds. Applications of interaction graphs to cellular signaling networks include the identification of signaling pathways and feedback loops, and the analysis of global interdependencies, which is useful to check the consistency of experimental data with a given network structure. In *logical models*, the information that is contained in an interaction graph is extended by rules defining how the discrete state of a node is governed by the states of other nodes. This enables one to compute the qualitative input–output behavior of a signaling pathway under study as well as the identification of intervention strategies. Based on *ODE models*, it is possible to explain and predict quantitative and dynamic system behavior. Interaction graphs, logical models, and ODE models are tightly interlinked since every logical model and every ODE model has an underlying interaction graph (Figure 1.1): Logical models can be derived from interaction graphs by introducing logical rules specifying how to combine different inputs into one node. For an ODE model, several system properties can be derived from its underlying interaction graph, which represents the sign structure of the Jacobian. Within this thesis, it was shown that by combining modeling approaches of different complexity, important insights into the structure and functioning of signaling pathways can be gained. As application examples, different models of the Epidermal

Growth Factor (EGF) and of Hepatocyte Growth Factor (HGF) signaling pathways, which both play a central role in liver regeneration, were built and analyzed. This led to new biological insights about these pathways.

The outline of this thesis is as follows: In Chapter 2, a general introduction on cellular signaling is given and the signaling systems that will be studied in this work are introduced. Furthermore, the different modeling approaches that will be used are described. In Chapter 3, a large-scale interaction graph model of signaling pathways activated by ligands of the EGF ligand family is presented, which was built from literature information. New methods for the interrogation of network structures based on interaction graphs are described. By applying these methods to the EGF model, cell-type specific discrepancies between the canonical model structure and experimental data were identified, and new hypotheses were derived. In Chapter 4, some extensions of the methods introduced by Klamt et al. (2006) for the static analysis of logical models are presented. In addition, it is shown how the qualitative input–output behavior and discretized experimental data can be compared in an automated way. The application of the newly derived methods to the EGF logical model complements the results received from the interaction graph analysis. In Chapter 5, the benefit from combining different modeling techniques gets even more concrete. There, a novel hybrid modeling approach is presented, which combines qualitative and quantitative modeling approaches. Considering HGF-induced activation of two central signaling pathways, data-driven interaction graph analysis enabled to pre-select interaction graph model structures out of a large number of possible structures resulting from different combinations of possible feedback and crosstalk regulations. A subsequent analysis of ODE models revealed which network structure reflected the transient behavior observed in the experimental data best. While in Chapter 5 ODE models are derived from interaction graphs, Chapter 6 is dedicated to analyze the properties of interaction graphs underlying an ODE model. In this way, conclusions on possible dynamic behaviors of a model can be drawn that depend solely on the system structure and not on parameter values. This kind of analysis also enables the identification of network structures that can cause an observed dynamic behavior. In particular, methods to exclude certain responses to perturbations in steady state for chemical reaction networks are presented. Both graph-theoretic and algebraic results were derived. Finally, the connection to previously derived structural requirements for multistationarity are shown.

Several publications by the author of this thesis are included in this work. These are D’Alessandro, Samaga, Maiwald et al. (2015), Melas, Samaga et al. (2013), Sam-

aga and Klamt (2013), and Samaga et al. (2009, 2010). Text passages taken from these publications are not explicitly marked as citation in this thesis, but the publications are referred to in the respective chapter or section introductions.

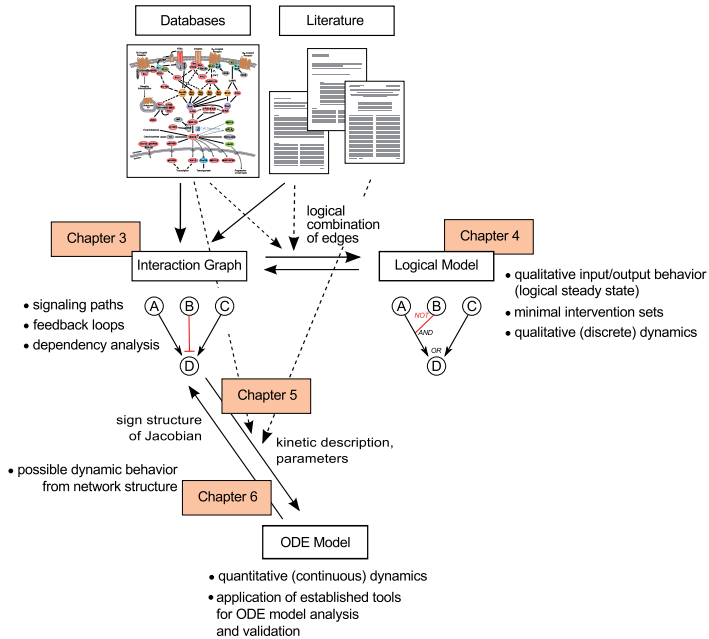


Figure 1.1: Different modeling formalisms for cellular signaling. Interaction graphs can often be directly derived from biological databases. By choosing logical combinations for the edges, logical models can be built. A particular hypergraph representation of logical models enables one to derive the underlying interaction graph from the logical model by splitting the AND gates (Klamt et al. 2006). Based on interaction graph structures, ODE models can be derived by adding kinetic descriptions and parameters. Each ODE model has an underlying interaction graph which represents the sign structure of the Jacobian matrix and enables predictions on the possible dynamic behavior from the network structure (Thieffry 2007). For further explanations see text. Illustration of the pathway scheme reproduced courtesy of Cell Signaling Technology [2]. Whole figure adapted from Samaga and Klamt (2013).

2 | Systems Biology of Signal Transduction Networks

The aim of this chapter is to give a general introduction to Systems Biology of signaling networks and to introduce the signaling systems that will be studied in this work. After a short introduction to Cell Signaling (Section 2.1), we will give an overview of liver regeneration (Section 2.2). Liver regeneration is a highly complex phenomenon, and it has become an example of use of systems biology approaches. We built and analyzed mathematical models of two signaling pathways playing a major role in liver regeneration, and we will introduce these pathways in Section 2.3. Finally, we will give a detailed description of the different modeling methods that are central to this work (Section 2.4).

2.1 Cell Signaling

Multicellular organisms use complex mechanisms to enable cell-to-cell communication and to respond to their environment. Various signal transduction pathways link mechanical or chemical stimuli to a cellular response. In general, extracellular signaling molecules bind to receptors at the cell surface, thus activating intracellular signaling pathways. The components of these pathways are signaling proteins such as kinases, phosphatases, small GTPases, or adapter proteins, and they are connected by activating and deactivating mechanisms, each of which passes the signal from one species to another. Examples of these include chemical modifications such as phosphorylations, triggering of conformational changes, and colocalizations. In this way, the initial stimulus is transduced to target proteins such as transcription factors, metabolic enzymes, or cytoskeletal proteins, and changing their activity affects the cellular response by changing, for example, protein syntheses, the cellular metabolism, or cell movement (see, e.g., Alberts et al. 2004, Krauss 2003).

2.2 Liver Regeneration

A phenomenon that already fascinated ancient Greeks and that is regulated by a well-concerted interplay of various signaling processes is liver regeneration (reviewed in Fausto et al. 2006; Michalopoulos 2007; Taub 2004). The liver as important organ—it plays a central role in metabolic homeostasis and is the main detoxifying organ of the body—has the capacity to return to normal size after loss of liver mass. The main functional cells of the liver are hepatocytes, which make up about 80% of hepatic cells (Taub 2004). Although being highly differentiated and long-lived cells, hepatocytes are still able to proliferate: During liver regeneration, activation of cytokines provokes that quiescent hepatocytes re-enter the cell cycle. Growth factors then drive cell cycle progression by overriding a restriction point in the late G1 phase (Fausto et al. 2006). Two growth factor signaling systems play a major role in liver regeneration. These are the hepatocyte growth factor (HGF) and its receptor Met, and the epidermal growth factor (EGF)-like ligands with their receptor family ERBB (Michalopoulos and Khan 2005). In order to understand liver regeneration, it is important to unravel the complexity of these systems. Therefore, we built within this work mathematical models of signaling through ERBB receptors and of HGF/Met signaling and applied newly developed methods for their analysis in combination with experimental data from hepatocytes. This was part of the funding initiatives HepatoSys [6] and Virtual Liver Network [8] of the German Federal Ministry for Education and Research. In the next section, we give an overview of these two signaling systems and briefly describe the developed models.

2.3 Signaling through the EGF Ligand Family and through HGF

2.3.1 The Signaling System of the EGF Ligand Family

The signaling cascade that is stimulated by epidermal growth factor (EGF)-like ligands is among the best studied signaling pathways in mammalian cells. Aside from EGF, members of the EGF ligand family include transforming growth factor alpha (TGF- α), the neuregulins (NRGs), amphiregulin (AR), betacellulin (BTC), heparin-binding EGF-like growth factor (HB-EGF), and epiregulin (EPR). EGF-like growth factors bind to receptors of the ERBB protein family¹, which consists of four

¹The name ERBB was originally chosen because of the homology to the erythroblastoma viral gene product, *v-erbB* (Citri and Yarden 2006).

receptor tyrosine kinases: epidermal growth factor receptor (EGFR/ERBB1), ERBB2 (HER2/Neu), ERBB3, and ERBB4. Ligand binding induces the formation of ERBB receptor homo- and heterodimers and subsequent autophosphorylation of specific tyrosine residues, which provide docking sites for proteins containing Src homology 2 (SH2) or phosphotyrosine binding (PTB) domains (Olayioye et al. 2000). Due to distinct biochemical properties, the individual ERBB ligands activate different ERBB dimers and also show differential binding affinities. Furthermore, each receptor dimer displays an individual pattern of autophosphorylation sites and, thus, recruits distinct effector proteins (Olayioye et al. 2000). Two of the four ERBB receptor types do not form functional homodimers, but only heterodimers: ERBB3 receptors have impaired kinase activity, whereas ERBB2 does not bind any ligand of the EGF family (Citri and Yarden 2006). The latter is regarded as a non-autonomous amplifier of ERBB signaling (Citri and Yarden 2006); it is the preferred heterodimerization partner of the other ERBB receptors and as such impairs the formation of ERBB1/ERBB3, ERBB1/ERBB4, and ERBB3/ERBB4 heterodimers (Graus-Porta et al. 1997; Olayioye et al. 1998). All in all, this complex system made up of multiple ligands, which signal through eight potential receptor homo- and heterodimers, gives rise to a great signaling diversity: signaling through ERBB receptors regulates various cellular processes such as survival, proliferation, development, and growth (Citri and Yarden 2006; Oda et al. 2005; Olayioye et al. 2000). Furthermore, ERBB receptors play an important role in the development and malignancy of human cancers (Olayioye et al. 2000) and are targets for new and existing anti-cancer drugs (Schoeberl et al. 2009).

In order to untangle the complexity of ERBB signaling, various mathematical models have been developed within the last years (reviewed in Citri and Yarden (2006) and Wiley et al. (2003)). The first models focused on EGFR and described its internalization, ligand binding, and homodimerization (Wiley and Cunningham 1981; Wofsy et al. 1992). Later models included downstream signaling events (e.g., Kholodenko et al. (1999), Resat et al. (2003), and Schoeberl et al. (2002)) and also addressed homo- and heterodimerization among members of the ERBB receptor family and the effects on downstream signaling in response to different ligands (e.g., Birtwistle et al. (2007), Chen et al. (2009), Hatakeyama et al. (2003), and Hendriks et al. (2006)). All these models describe the dynamics of the involved species, but cover only limited parts of the ERBB pathway.

Regarding liver regeneration, it has been shown that the different ERBB ligands have differing effects; thus, it is crucial to disentangle the complexity of signaling of

the ERBB family members in hepatocytes (Michalopoulos and Khan 2005). Within the last years, high-throughput experimental techniques were developed that enable the generation of data where various readouts are measured in response to several stimuli and to perturbations all over considered networks. Thus, system-wide analyses have come into reach. However, there is now also the need for large-scale mathematical models and for techniques that enable the assessment of these data with the models. In this work, we present a large-scale model that comprises the main and also the side routes of ERBB signaling. This model is based on the EGFR pathway map developed by Kitano and colleagues (Oda et al. 2005) and additional information from the literature. We use a logical modeling formalism (Klamt et al. 2006), which is of a qualitative nature, but still enables to derive important functional properties and predictions. The model describes signaling induced by thirteen members of the EGF ligand family through the four ERBB receptors, leading to the activation of various kinases and transcription factors that affect proliferation, growth, and survival (Samaga et al. 2009). This model is studied in Section 4.4 of this work. The list of included species and mechanisms is given in Appendix C. Furthermore, we analyzed the underlying interaction graph of this model; this is subject of Section 3.3. Both modeling formalisms are described in Section 2.4, and graphical representations of the logical ERBB model and the interaction graph are given in Figures 4.3 and 3.2, respectively. Our ERBB model describes ligand binding and subsequent autophosphorylation, what is followed by binding of adapter proteins such as GAB1, GRB2, and SHC. These proteins transmit signals to the small GTPases Ras and Rac, leading to the activation of mitogen-activated protein kinase (MAPK) cascades. Among these, ERK1/ERK2 is the best studied, but our model also contains the p38 and JNK cascades. Highly interconnected with the MAPKs and also downstream of the ERBB receptors is phosphatidylinositol 3-kinase (PI3K)/AKT signaling, another major branch of the model. Furthermore, activation of different STATs and the PLC γ /PKC pathway are included. Both the logical ERBB model and the interaction graph were used to analyze phosphoproteomic data from primary human hepatocytes and the human liver carcinoma cell line HepG2 in order to reveal hepatocyte-specific characteristics of the network (see Sections 4.4 and 3.3).

Besides the large-scale model of ERBB signaling, for illustration purposes, we will use a smaller version, which was manually derived from the large-scale model, and which we introduced in Samaga and Klamt (2013). From the thirteen EGF-like ligands, we chose EGF and NRG1. EGF binds specifically to ERBB1, whereas NRG1 binds to ERBB3 and ERBB4 (Olayioye et al. 2000). The main purpose of the

EGF/NRG1 example network is to illustrate the presented methods; thus, we tried to keep the network simple while still being biologically plausible.

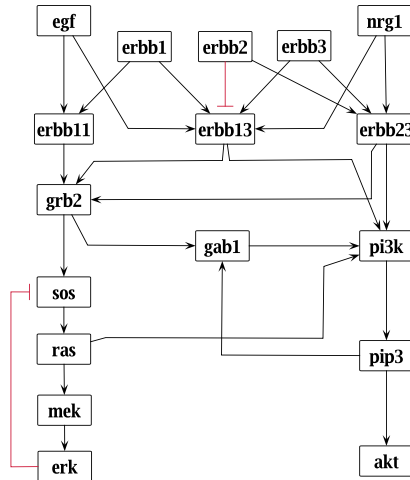


Figure 2.1: Interaction graph of the EGF/NRG1 network example. Activation of MAPK signaling through ERBB receptors requires the recruitment of SOS to the receptors via the adapter protein GRB2. SOS activates the small GTPase Ras by provoking guanine nucleotide exchange, and GTP-bound Ras activates Raf, a MAP-Kinase-Kinase-Kinase, what initiates the MAPK signaling cascade. Activated Raf phosphorylates the MAP kinase kinase (MEK), which in turn phosphorylates ERK. PI3K signaling in response to EGF-like ligands is initiated by PI3K binding to the receptors, either directly, or indirectly via GAB1. Activated PI3K generates phosphatidylinositol (3,4,5)-trisphosphate (PIP₃), which serves as a docking site for the serine/threonine protein kinase AKT at the plasma membrane. AKT is then activated by phosphorylation of two specific phosphosites. Black arrows indicate positive (activating) edges, red blunt-ended lines negative (inhibiting) edges. Figure adapted from Samaga and Klamt (2013).

Of the different downstream signaling pathways, we focused on two major ones, the MAPK signaling cascade activating ERK, and the PI3K signaling pathway activating AKT. In the example network, both pathways are described in a compressed way, neglecting some of the intermediate species. Furthermore, we did not consider all the various feedback and crosstalk mechanisms that have been reported for both pathways (see, e.g., Mendoza et al. (2011)), but focused on some exemplary ones. In order to keep the activation mechanisms at the receptor level simple, ERBB4 was not included, and only three out of the four functional dimers formed by ERBB1, ERBB2, and ERBB3 are part of the network. Including ERBB3 rather than ERBB4

was motivated by the fact that ERBB3 can directly activate PI3K, whereas ERBB4 can only indirectly activate PI3K via GAB1 and thus shows a similar signaling response as ERBB1 (see, e.g., Citri and Yarden (2006)). Again, we study a logical model (Figure 2.3) and the interaction graph (Figure 2.1) of the EGF/NRG1 example network.

2.3.2 Hepatocyte Growth Factor Signaling

Hepatocyte growth factor (HGF) signals through the receptor tyrosine kinase Met (reviewed in Birchmeier et al. (2003), Furge et al. (2000), and Gentile et al. (2008)). Ligand binding induces autophosphorylation of two tyrosines within the tyrosine kinase domain, what activates the intrinsic kinase activity of the receptor (Naldini et al. 1991). Phosphorylation of two additional tyrosines in the C-terminal tail of Met activates a multisubstrate docking site, which binds several adapter proteins via their SH2 or PTB domains (Ponzetto et al. 1994). A key player in HGF signaling is the adapter protein GAB1, which binds to the multisubstrate docking site via its specific Met binding domain. Receptor-bound GAB1 then recruits other signaling molecules (Birchmeier et al. 2003; Gentile et al. 2008). HGF/Met signaling activates different signaling branches through which it regulates various biological responses such as proliferation, survival, migration, and morphogenesis (Birchmeier et al. 2003; Furge et al. 2000). It plays a key role in cancer progression and seems to be a promising target for cancer treatment (Furge et al. 2000).

Intact HGF/Met signaling is crucial to enable hepatocyte proliferation (Borowiak et al. 2004). Two signaling pathways that are activated by HGF are the MAPK and the PI3K pathway. Both pathways were shown to be regulated during liver regeneration: in conditional Met mutant mice, no MAPK activation and a reduced activation of the PI3K pathway was observed in response to partial hepatectomy; as a consequence, cell cycle progression of hepatocytes was inhibited (Borowiak et al. 2004). In response to HGF, PI3K binds to the Met receptor both directly and indirectly via GAB1 (Furge et al. 2000; Ponzetto et al. 1993). As in ERBB signaling, activation of PI3K is followed by generation of PIP₃ and AKT recruitment and phosphorylation. MAPK signaling is initiated by SOS binding to the activated Met receptor via GRB2, what, again as in ERBB signaling, results in ERK phosphorylation via activation of Ras, Raf, and MEK. PI3K and MAPK signaling are highly interlinked, and both pathways are regulated by feedforward and feedback mechanisms. In order to understand the mechanisms controlling hepatocyte proliferation during liver regeneration, it is necessary to unravel which crosstalk, feedback, and feedforward

mechanisms are relevant in hepatocytes.

To this aim, mathematical models provide unique tools to disentangle the complexity and to predict the impact of perturbations. Mathematical models of the MAPK signaling pathway have been developed that only consider negative feedback (Kholodenko 2000), negative and positive feedback loops (Shin et al. 2009), or that analyze the signal-to-response relation (Schoeberl et al. 2002). Mathematical models describing both PI3K and MAPK signaling pathways upon single or combinatorial stimuli revealed the presence of crosstalk mechanisms between MAPK and PI3K pathways (Borisov et al. 2009; Klinger et al. 2013; Won et al. 2012) or differences in the stimulus-specific network topology (Cirit and Haugh 2012; Santos et al. 2007). As indicated, most of the studies considered only single feedbacks or a limited number of crosstalk mechanisms. Therefore, to unravel a more complex network structure, a systematic unbiased approach is required.

We built an interaction graph model describing MAPK and PI3K signaling in response to HGF that contains reported crosstalk, feedback, and feedforward mechanisms. This model serves as starting point for a novel hybrid modeling approach, which combines qualitative and quantitative modeling techniques to unravel the HGF-induced activation of MAPK and PI3K signaling in primary mouse hepatocytes based on time-resolved experimental data (subject of Chapter 5). The list of species and mechanisms that are included in the interaction graph model is given in Appendix E, and a graphical representation is shown in Figure 5.2.

2.4 Modeling Approaches for Signaling Networks

In the following, we introduce the modeling concepts that are subject of this work. Parts of this section are derived from a review article we published (Samaga and Klamt 2013).

2.4.1 Interaction Graphs

Cellular signaling networks can intuitively be described as signed directed graphs, known as *interaction graphs* (sometimes also called *influence graphs* or *regulatory graphs*). The nodes in these graphs represent the components of signaling such as hormones, receptors, protein kinases and phosphatases, adapter proteins, transcription factors, second messengers, or genes. The activating and deactivating mechanisms connecting these components in a signaling network (e.g., chemical modifications such as

phosphorylations) are represented as edges in the interaction graph: each edge connects a pair of nodes and is directed from the species passing the signal to the species receiving it. Furthermore, an associated sign indicates whether the edge represents an activating (positive sign) or deactivating mechanism (negative sign). Formally, an interaction graph G consists of a set $V = V(G)$ of *nodes* (or vertices), a set $A = A(G)$ of *edges* (or arcs) that are defined as ordered pairs of nodes, and a sign mapping $\sigma : A(G) \rightarrow \{-1, 1\}$. Given an edge (u, v) pointing from node $u \in V$ to node $v \in V$, u is called *tail* and v *head* of the edge. A node's *in-degree* is the number of edges pointing to the node. Accordingly, a node's *out-degree* is the number of edges pointing from the node.

Interaction graphs are often represented as “pathway cartoons” and can thus be seen as the prevalent formalism describing signaling networks in the biological literature. They are also commonly used to represent signaling pathways in pathway databases such as Reactome (Joshi-Tope et al. 2005), KEGG (Ogata et al. 1999), WikiPathways (Kelder et al. 2012), or in public repositories provided, for example, by BioCarta [1], or Cell Signaling Technology [2].

Graph models such as interaction graphs can be used to study global topological network properties (such as degree distributions) and thus to unravel common design principles of biological networks (reviewed in (Aittokallio and Schwikowski 2006; Albert 2005; Barabási and Oltvai 2004)). For instance, many biological networks were found to have a scale-free topology, where the majority of nodes has a low degree, while still a relatively large number of nodes (compared to random networks) is connected to many compounds. A well-known example for such a highly connected “hub” is the tumor suppressor protein p53 (Vogelstein et al. 2000). In addition to these statistical features characterizing the overall architecture of a given biological network, an interaction graph encodes other important properties highly relevant for understanding basic network functions.

Before discussing those properties, it is important to realize that interaction graphs are often implicitly contained as underlying network structure in models of more complex formalisms. In particular, this holds true for Boolean and ODE models. For example, given an ODE system, the entries of its Jacobian matrix reflect pairwise influences between species. Therefore, we can associate with the system an interaction graph that is defined on the basis of the signs of these entries (Thieffry 2007). Accordingly, functional properties derived from interaction graphs are directly relevant for all models having this graph as underlying structure (Radde et al. 2010). We will further discuss interaction graphs in the context of ODE modeling in Section 2.4.3.2.

The relation between Boolean models and interaction graphs will be explained in Section 2.4.2.1.

2.4.1.1 Cycles in Interaction Graphs

Feedback loops are sequences of edges by which components can influence their own activation level (Thomas and D’Ari 1990). They are found in almost all known signaling pathways and have been shown to have major impacts on network dynamics and to mediate important biological functions (Brandman and Meyer 2008; Tyson et al. 2003).

Formally, a feedback loop is a directed cycle and is defined as alternating sequence of nodes and edges starting and ending at the same node, while visiting no node (except the start/end node) twice. Thus, a feedback loop is a sequence $v_1 a_1 v_2 a_2 \dots v_{k-1} a_{k-1} v_k$ such that (i) node v_1 is equal to node v_k , (ii) the tail of edge a_i is node v_i , and the head of edge a_i is node v_{i+1} , and (iii) all nodes v_1, \dots, v_{k-1} are distinct. Depending on the parity of the number of negative edges the sequence contains, a feedback loop is said to be negative (odd number of negative edges) or positive (zero or even number of negative edges). Equivalently, the sign of a feedback can also be determined by multiplying the sign of all edges making up the loop.

The interaction graph representation of the EGF/NRG1 network (Figure 2.1) contains two feedback loops: (i) the sequence $PI3K \rightarrow PIP_3 \rightarrow GAB1 \rightarrow PI3K$ forms a positive feedback loop as all edges are positive, and (ii) the sequence $SOS \rightarrow Ras \rightarrow ERK \rightarrow SOS$ forms a negative feedback loop as it contains one negative edge ($ERK \rightarrow SOS$).

Positive feedback loops may cause a discontinuous switch in the cellular response (Tyson et al. 2003) as has been, for example, shown in frog oocytes, where a positive feedback loop (in combination with ultrasensitivity) triggers the conversion of a continuous stimulus (progesterone) into an all-or-none biological response (oocyte maturation; Ferrell and Machleder (1998)). A bistable behavior like this is in general associated with positive feedback loops, and, indeed, it was shown that a system that displays more than one steady state—both in a Boolean or ODE model representation—must contain a positive feedback loop in its interaction graph (Cinquin and Demongeot 2002; Gouzé 1998; Plahte et al. 1995; Snoussi 1998; Soulé 2003; Thomas and Kaufman 2001a,b). We will come back to this in Section 2.4.3.2.

Negative feedback loops stabilize the system’s response and are a common de-

sign principle of biochemical systems to achieve homeostasis, that is, to keep the (activation) level of certain components at an optimal value (Brandman and Meyer 2008; Novák and Tyson 2008; Thomas and D'Ari 1990; Tyson et al. 2003). They have also been shown to create oscillations, and, given an ODE model, just as a positive feedback loop in the associated interaction graph is necessary for multistationarity, a negative feedback loop is a prerequisite for an oscillatory behavior for constant signed Jacobian (Angeli et al. 2009; Gouzé 1998; Plahte et al. 1995; Richard and Comet 2011; Snoussi 1998). Although sustained biochemical oscillations can be generated by a single negative feedback loop, as, for example, in NF- κ B signaling (Hoffmann et al. 2002), they often arise from motifs containing both positive and negative feedbacks (Brandman and Meyer 2008; Novák and Tyson 2008). An example is that of periodic calcium spikes as they have been observed after growth factor or hormone stimulation (Meyer and Stryer 1988).

Thus, the identification and investigation of feedback structures might help to understand core design principles of non-trivial dynamic behavior (see Section 2.4.3.2).

2.4.1.2 Dependency Matrix

Perhaps the most direct questions that can be answered with an interaction graph at hand are related to *signaling paths* between pairs of nodes. A signaling path from node v_1 to node v_k is a sequence $v_1 a_1 v_2 a_2 \dots v_{k-1} a_{k-1} v_k$, where all nodes v_1, \dots, v_k are distinct, and edge a_i points from node v_i to node v_{i+1} . Just as for feedback loops, a path is negative if it contains an odd number of negative edges, else positive. We will refer to v_1 as the source node and v_k as the target node of the signaling path.

First of all, one might be interested in identifying all different signaling routes that exist between a given pair of nodes, for example, the different paths through which a ligand influences the activity of a transcription factor. Signaling paths reveal how the often well-known local interactions are combined to network-wide influences. If applied in a systematic manner, this enables one to classify a source species with respect to a target species, depending on the sign(s) of the signaling path(s) connecting them (Klamt et al. 2006): (1) if all paths from the source to the target node are positive, the source is an *activator* of the target; (2) if all paths from the source to the target node are negative, the source is an *inhibitor* of the target; (3) if there exist positive and negative paths from the source to the target node, the source is said to be an *ambivalent factor* of the target; and (4) if there exists no path from the source to the target node, the source has no influence on the target and is therefore called

neutral factor. For certain predictions it is advantageous to refine the classification of activators and inhibitors by considering also information about negative feedback loops: if A is an activator of B and none of the species lying on a path from A to B is part of a negative feedback, B behaves monotone with respect to changes in A (see Kunze and Siegel (1994) and Section 2.4.3.2 how this translates to ODE systems), that is, increasing A results (after some time) in an increase of B. In this case, we call A a *strong activator* for B in contrast to *weak activators*, where at least one of the activating paths touches a negative feedback loop (Klamt et al. 2006, 2007). Accordingly, if A is an inhibitor of B and none of the species lying on a path from A to B is part of a negative feedback loop, increasing A results in a decrease of B, and A is called a *strong inhibitor* of B, otherwise *weak inhibitor*. For weak activators and inhibitors, we can only predict that the initial response (starting from a steady state) of the target nodes will be positive/negative, but nothing can be said on the asymptotic behavior.

The information how the species influence each other can be stored in a compact manner in a dependency matrix (Klamt et al. 2006). The dependency matrix for the EGF/NRG1 example model is shown in Figure 2.2. The diagonal entries of this matrix represent how a species acts on itself: just as the influence of a species on another is characterized by the sign of the connecting paths, the influence of a species on itself is characterized by the sign(s) of the feedback loop(s) it is involved in. Based on the dependency matrix, the effect of stimulation or perturbation experiments can be predicted and then be compared with the measured behavior. Within this work, we will present new methods for data analysis and network inference that make use of this concept (Chapters 3 and 5).

Although the computation of the dependency matrix is in general an NP-complete problem, the algorithms we are using compute the matrix reasonably fast (less than one second) in realistic networks we considered (Klamt and von Kamp 2009). Alternatively, algorithms delivering approximations with high precision are available, and in networks without negative feedbacks, an exact polynomial algorithm can be used (Klamt and von Kamp 2009).

2.4.1.3 Minimal Cut Sets

A task that is of particular importance for medical applications is the problem of identifying possible strategies to prevent signal propagation through certain signaling paths. If, for example, one is interested in blocking the activation of AKT in response to NRG1 or EGF in the EGF/NRG1 example (Figure 2.1), one possibility

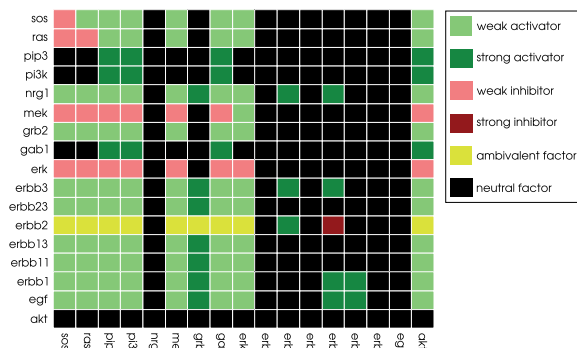


Figure 2.2: Dependency matrix of the EGF/NRG1 example model. Shown is the dependency matrix of the EGF/NRG1 example model given in Figure 2.1. The color of row i , column j indicates the influence of species i on species j (see legend). Figure adapted from Samaga and Klamt (2013).

is to prevent signaling through the essential PI3K by inhibiting its kinase activity or by removing it from the system. To block AKT activation by intervening at the receptor level, one has to make sure that signaling through all three receptor dimers ERBB11, ERBB13, and ERBB23 is prevented as they give rise to redundant routes. To tackle those problems, *Minimal Cut Sets* (MCSs) can be computed in interaction graphs which are minimal sets of compounds or/and edges that have to be removed to interrupt a given set of paths and/or feedback loops (Klamt et al. 2006). MCSs correspond to *feedback arc sets* or *feedback vertex sets* in the special case where feedback loops are to be disrupted (Festa et al. 2009). MCSs in interaction graphs are also very similar to MCSs introduced for metabolic networks, which disrupt a given set of metabolic pathways, for example, those synthesizing an undesired product (Klamt 2006). Both types of MCSs can be computed by the minimal hitting set algorithm, and it is also possible to consider side constraints, for example, to keep certain (desired) paths/pathways intact (Hädicke et al. 2011).

2.4.2 Logical Models of Signal Transduction Networks

Given a signaling pathway, a question that immediately arises is whether pathway stimulation leads eventually to full activation of a certain downstream protein, for example, a transcription factor. This is an example for a question of qualitative

nature that can often not directly be answered based on an interaction graph, but requires more complex and deterministic modeling formalisms. First of all, the state of a node—in the easiest case *active* or *inactive*—is not defined in an interaction graph; rather, state changes can be considered (*up* or *down*). Furthermore, interaction graphs reflect pairwise interactions, whereas the biochemical processes in the cell often involve more than two players. Thus, whether a signal can be transmitted from a source node to a target node often depends on a third node. An example from Figure 2.1 is the activation of ERBB1: both the receptor monomer ERBB1 and the ligand EGF are needed to get the phosphorylated receptor dimer that is able to trigger downstream signaling events. This shows that information on how the different interactions influencing a species are combined is necessary to make functional predictions on a node's state. One possibility is to decode this information into a logical function. Together with logical variables associated with each species and representing the activation level as discrete states, these functions define a logical network.

Logical modeling of biological systems was pioneered by Kauffman (1969) and has since then emerged as valuable formalism in systems biology (for recent reviews see, e.g., Morris et al. (2010), Wang et al. (2012), and Wynn et al. (2012)). Various applications to modeling gene regulatory and signaling networks can be found in the literature (Albert and Othmer 2003; Christensen et al. 2009; Giacomantonio and Goodhill 2010; Huard et al. 2012; Li et al. 2004; Naldi et al. 2010; Ryll et al. 2011; Schlatter et al. 2009; Zhang et al. 2008). Most frequently, Boolean networks are studied where the logical variables are only allowed to take the values 0 (inactive or absent) or 1 (active or present). In more general approaches, the variables can take an arbitrary number of discrete (multivalued logical models; see Thomas and D'Ari (1990)) or even continuous values (fuzzy logic models; see, e.g., Aldridge et al. (2009) and Morris et al. (2011)). In the following, we focus on Boolean models.

As mentioned above, every node in a logical network possesses a logical function defining how the state of the node (that is, the value of the associated logical variable) can be derived from the state of other nodes. Generally, a logical function can be composed by using arbitrary logical operations (such as AND, OR, NOT, XOR, NAND etc.), and different representations of one and the same logical function may exist (Thomas and D'Ari 1990). It is often useful and intuitive to restrict the logical operators to AND (also called logical product), OR (also called logical sum), and NOT, and then to express the logical functions as sum of products (SOP): AND terms consisting of several logical variables or their negated form are ORed together

(Thomas and D’Ari 1990). The SOP representation is also known as disjunctive normal form (DNF). Any Boolean function can be expressed in this way.

We exemplify the SOP representation by means of activation of ERBB13 in the EGF/NRG1 model (Figure 2.1). Dimerization of ERBB1 and ERBB3 and subsequent autophosphorylation of the receptor dimer arises both after EGF and NRG1 stimulation and is impaired if ERBB2 is present (see introduction on ERBB signaling in Section 2.3.1). Thus, assuming each species can be either active/present (1) or not (0), and using the symbols \cdot for AND, $+$ for OR, and $!$ for NOT, the logical function describing ERBB13 activation reads in SOP representation

$$\text{ERBB13} = \text{EGF} \cdot \text{ERBB1} \cdot \text{ERBB3} \cdot !\text{ERBB2} + \text{NRG1} \cdot \text{ERBB1} \cdot \text{ERBB3} \cdot !\text{ERBB2}. \quad (2.1)$$

Usually, logical variables in a Boolean network are binary, that is, they take only the values 1 or 0. Here, we consider a formalism that uses a third value $*$ to represent an undefined or unknown state. The Boolean operators can be generalized to the three-valued logic in a straightforward way (Abramovici et al. 1990): the AND operation returns 1 if and only if all of its inputs are 1; it returns 0 iff at least one input is 0, and it returns $*$ in all other cases. The OR operation returns 0 if and only if all of its inputs are 0; it returns 1 iff at least one input is 1, and it returns $*$ in all other cases. Finally, the negated value of $*$ is $*$ and, as in the binary case, NOT(0) is 1 and NOT(1) is 0.

In the following, we consider logical networks as follows: Given a set of species $V = \{v_1, \dots, v_n\}$, each species v_i has an associated logical variable x_i that can be determined by its corresponding logical function $f_i : x \rightarrow \{0, 1, *\}$, $x = (x_1, \dots, x_n)^T$. Furthermore, we assume that all logical functions are given in SOP representation.

2.4.2.1 Hypergraph Representation of Logical Models

A representation of logical networks that is well suited to study signal transduction pathways is based on *directed hypergraphs*, which in turn relies on SOP-represented logical functions (Klamt et al. 2006). Hypergraphs are generalizations of graphs, as an edge in a hypergraph (also called hyperedge) is not restricted to connect a pair, but can connect an arbitrary number of nodes. Accordingly, a hyperedge in a directed hypergraph connects a set of start nodes with a set of end nodes (Klamt et al. 2009). In our particular case, the set of end nodes consists of only one element.

Just as in interaction graphs, the nodes of the hypergraph represent the biological species. Now, each summand (which is an AND term or a single, possibly negated, logical variable) within the SOP-represented logical function of a node A becomes a hyperedge pointing into this node A .

Thus, each hyperedge in the hypergraph can be interpreted as a signaling event, that is, one mode of activation of the downstream node. If a hyperedge has several start nodes, the associated logical variables are inputs of an AND operation. In case of a single start node, the hyperedge becomes a simple edge, which indicates that the activation level of a single species determines the state of the downstream node. Furthermore, each edge branch has an associated sign indicating whether the value of the node it arises from is negated by a NOT operation ($-$) or not ($+$). Different activation modes of one species, that is, edges that point into the same node, are connected by an OR operation.

Considering again the logical function of ERBB13 given in Equation (2.1), each summand (AND term) is represented as a hyperedge pointing into ERBB13 in the hypergraph: the first hyperedge connects the start nodes EGF, ERBB1, ERBB3, and ERBB2 with the end node ERBB13; the second hyperedge connects the start nodes NRG1, ERBB1, ERBB3, and ERBB2 with ERBB13 (Figure 2.3). In both hyperedges, the branch coming from ERBB2 is marked in red, indicating that it enters the logical function in its negated form. Figure 2.3 shows the complete hypergraph representation of the logical model of EGF/NRG1 signaling from which the logical functions of each node can easily be derived.

The typical workflow when building a logical model is to first determine and analyze the interaction graph before defining the logical functions for each node. Choosing an appropriate logical function for a signaling process is not an easy task and requires a competent knowledge of the molecular mechanisms behind; therefore, this step often involves an intense literature study. Obviously, several logical models can be derived from the same interaction graph. Even in the Boolean case—as long as a node has more than one ingoing edge—one has to decide whether to use an AND or an OR operation, or, in the case of three or more inputs, a combination thereof.

In cases where one cannot gather from the available knowledge whether an AND or an OR operation is the more apposite description of a biological process, an alternative is to use logical operators with an incomplete truth table (Klamt et al. 2006). In general, this limits the determinacy of the model. However, as signaling pathways often feature redundant network structures, a model containing logical operations

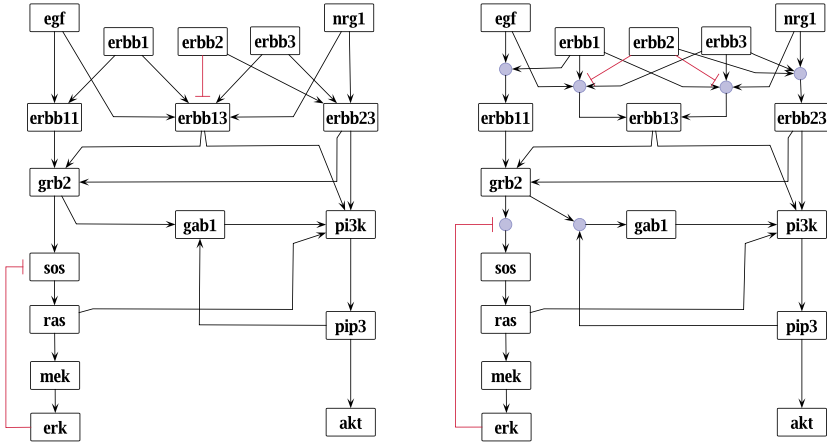


Figure 2.3: Logical model of the EGF/NRG1 network example. Shown is the hypergraph representation of a Boolean model (right) with underlying interaction graph shown on the left (this interaction graph was already given in Figure 2.1). Blue circles denote AND operations, that is, a hyperedge with n inputs is represented as n arrows pointing into a blue circle and one arrow pointing out of it. Red blunt-ended lines indicate that the respective input value is negated. Several arrows pointing into one node are OR-connected. Figure adapted from Samaga and Klamt (2013).

with incomplete truth tables can still have a high predictive power (see Chapter 4).

A further advantage of the hypergraph representation is that it enables one, if desired, to return to the interaction graph which underlies the logical model and from which the logical model was built: one only needs to split the hyperedges representing AND operations into simple edges (with a minus sign if the edge stems from a negated branch of the hyperedge) followed by a removal of possibly arising duplicate edges (Klamt et al. 2006). In this way, the characteristics of the interaction graph are preserved in the logical description and can easily be derived from it, for example, if the user wants to compute the feedback loops (implicitly) contained in the logical network.

2.4.2.2 Dynamical Analysis of Logical Models

One reason that logical models have emerged as valuable modeling approach for biological systems is the fact that the logical description—despite its simplicity—is

able to capture essential qualitative features of the system's dynamics (Glass and Kauffman 1973; Thomas and D'Ari 1990). In the classical approach, the dynamics of a logical model are defined by a synchronous updating scheme (Kauffman 1969): the value of node v_i at time $t + 1$ is determined by the logical values of its input nodes v_{i_1}, \dots, v_{i_k} at time t as given by its logical function f_i :

$$x_i^{t+1} = f(x_{i_1}^t, x_{i_2}^t, \dots, x_{i_k}^t). \quad (2.2)$$

All states are updated simultaneously, assuming that the modeled biological processes all have the same duration. The synchronous scheme is deterministic as each state is followed by one subsequent state. In contrast, with the asynchronous logical description a more realistic updating scheme was introduced by which different time delays for the individual biological events can be accounted for (Thomas 1973).

Of particular importance when studying logical models of biological systems is the identification of attractors. These attractors represent the long-term behavior of the system and can often be associated with cellular phenotypes or steady cellular states (Wang et al. 2012). The simplest attractors are made up of a single state, referred to as fixed point or (logical) steady state. Complex (cyclic) attractors are made up of several states among which the system oscillates.

2.4.2.3 Logical Steady State and Signal Flow Analysis

Dynamical modeling of logical networks has been successfully applied to a variety of biological regulatory networks (e.g., Albert and Othmer 2003; Giacomantonio and Goodhill 2010; Li et al. 2004; Naldi et al. 2010; Zhang et al. 2008). However, as the knowledge of initial conditions and timescales is often incomplete in biological systems, an application to large-scale networks is difficult (Wang et al. 2012). Thus, in addition to the described dynamic simulations of a logical network, there are static methods particularly suited for the analysis of large-scale networks.

As already stated above, a fundamental question that arises when studying signaling pathways is how the system reacts to different stimulations, for example, different combinations of ligands and inhibitors. Given a logical network, the qualitative input-output response can be computed by propagating the logical values of a set of fixed input nodes according to the logical functions. Apart from the inputs, all other logical values are assumed to be unknown. The goal of this procedure is to infer the logical steady state that results from the given inputs (Klamt et al. 2006).

We consider a given input state $x^0 = (x_1^0, \dots, x_n^0)^T$, where $x_i^0 \in \{0, 1, *\}$, and denote with $I_0 \subseteq \{1, \dots, n\}$ the species indices whose associated value is fixed to 0 or 1, that is, $I_0 = \{i : x_i^0 \in \{0, 1\}\}$. The species v_i with $i \in I_0$ represent, for example, external stimuli or knock-ins/knock-outs. For all $i \in I_0$, the Boolean functions f_i are overwritten by the simple binary values $x_i^0 \in \{0, 1\}$. The logical steady state resulting from the initial state x^0 is then a state vector $x^S(x^0) = (x_1^S(x^0), \dots, x_n^S(x^0))^T$, $x_i^S \in \{0, 1, *\}$, for which holds that $x_i^S = f_i(x^S)$ for all $i \in \{1, \dots, n\} \setminus I_0$, whereas $x_i^S = x_i^0$ for all $i \in I_0$. While in the case of two-valued Boolean networks several logical steady states might result from a given set of fixed values, or a logical steady state may not exist at all, a unique logical steady state follows for any given set of clamped values in the three-valued logic. For example, if no value is fixed in the network, all nodes will have state $*$ in the associated logical steady state. Logical states with value $*$ in logical steady state can be interpreted as states which cannot be uniquely resolved to a 0/1 value, either because the initial values of some other nodes need to be known to infer a unique 0/1 value, or because these nodes will never reach a fixed value because they are part of a negative feedback loop inducing oscillations in the Boolean (and possibly also in the continuous) dynamics.

The logical steady state in three-valued logic can be determined in an interactive and intuitive way by signal propagation using the logical functions for three-valued logic. The algorithm actually simulates the signal propagation taking place in the real situation: First, we initialize the state vector x with x^0 and introduce an index set I_K , which contains all those node indices for which we have found a 0 or 1 value that cannot change in the future. Clearly, at the beginning we have $I_K = I_0$. In each iteration, we then look for states x_i , $i \notin I_K$, for which we can deduce a value 0 or 1 given the current state vector x . If we can change a state x_m from $*$ to 0 or 1, we include the index m in I_K and start a new iteration. The algorithm stops if no further indices can be added to I_K , and x then represents the logical steady state $x^S(x^0)$. The algorithm is polynomial in input size and will terminate at latest after $n - \#I_0$ iterations, where $\#I_0$ denotes the cardinality of I_0 . This follows from the conservative nature of three-valued logic: a node will not change its state once it has been assigned a 0/1 value.

As an example, suppose the system in Figure 2.3 is stimulated with EGF in presence of all receptors, that is, $x_{\text{egf}}^0 = 1$, $x_{\text{nrg1a}}^0 = 0$, $x_{\text{erbb1}}^0 = x_{\text{erbb2}}^0 = x_{\text{erbb3}}^0 = 1$. By clamping these values and initializing the states of all other nodes with $*$, the states of the receptor dimers ERBB11, ERBB13, and ERBB23 will be updated in the first iteration of the algorithm (see Table 2.1). In the next step, we can only conclude

$x_{\text{grb2}}^S = 1$. Then the algorithm stops as no further node with state * can be updated. The other states cannot be uniquely determined due to the positive and negative feedback loops: as GRB2 is 1, the state of SOS depends on the activation level of ERK that is in turn dependent on SOS. The state of GAB1 depends on PIP₃ activity and thus on the state of PI3K. However, as two of the OR-connected inputs of PI3K are 0 (ERBB13 and ERBB23) and the state of another one cannot be determined (Ras), the state of PI3K depends on the state of its fourth input, which is again GAB1. Stimulation with NRG1 results in a different situation (see Table 2.1): in this case, PI3K can be activated directly by ERBB23, thus, independently of the states of the other nodes in the positive feedback loop, PIP₃ and GAB1. As the state of one node of the positive feedback is now determined, the other states in the loop can also be computed. The values of the nodes forming the negative feedback loop are still undefined. The subset of 0/1 values in the logical steady state corresponds to the term *partial logical steady state* in Klamt et al. (2006).

Table 2.1: Logical steady states in the EGF/NRG1 example model.

		With negative feedback		Negative feedback removed	
<i>Fixed input values</i>	egf	1	0	1	0
	nrg1a	0	1	0	1
	erbb1	1	1	1	1
	erbb2	1	1	1	1
	erbb3	1	1	1	1
<i>Computed logical steady state values</i>	erbb11	1	1	1	1
	erbb13	0	0	0	0
	erbb23	0	1	0	1
	grb2	1	1	1	1
	sos	*	*	1	1
	ras	*	*	1	1
	mek	*	*	1	1
	erk	*	*	1	1
	gab1	*	1	1	1
	pi3k	*	1	1	1
	pip ₃	*	1	1	1
	akt	*	1	1	1

The logical variables of the input nodes were set to the specified value and, according to the logical functions (see Figure 2.3), propagated through the network. The entry * indicates that the respective steady state value could not be uniquely resolved to a 0/1 value. Table adapted from Samaga and Klamt (2013).

The example showed that, in presence of functional feedback loops, it might happen that a 0/1 value in logical steady state can only be determined for a subset of

nodes. Thus, negative feedback loops are sometimes removed before computing the logical steady state. This is often justifiable as, from a qualitative perspective, one might only be interested in which signals can be activated at all and does not want to consider the downregulating effect of a negative feedback loop coming into play once the initial response occurred (Klamt et al. 2006). In our example model, we can break the negative feedback loop by removing the negative effect of ERK on SOS. As a consequence, all states can be computed in response to both EGF and NRG1 (see Table 2.1).

Positive feedback loops amplify the signaling response. Thus, their effect can often not be described in a satisfactory way using Boolean states. The biology behind the positive feedback loop in our example model is as follows (Rodrigues et al. 2000): in response to growth factor stimulation, GAB1 is recruited to the plasma membrane through binding to the ERBB1-GRB2 complex. This leads to activation of PI3K and, in turn, generation of PIP₃. The latter recruits additional GAB1 molecules to the receptor complex at the plasma membrane what enhances downstream signaling. In this case, a multi-level logical description would be most appropriate: active GRB2 (i.e., GRB2 bound to ERBB1 homo- or heterodimers) activates GAB1 to level 1, whereas GRB2 AND PIP₃ activate GAB1 to level 2.

In Chapter 4, we will show that computing the qualitative network response as described above enables to compare predictions derived with a given network structure with discretized data from stimulus-response experiments. Of course, one has to ensure that the measured time points and possible assumptions that are made for the logical steady state analysis, for example, regarding the activity of feedback loops, are valid (MacNamara et al. 2012).

2.4.2.4 Minimal Intervention Sets

Another problem that uses the concept of logical steady states is the identification of sets of interventions (an intervention representing logical values fixed to a certain value thus corresponding to knock-outs or constitutive activations) to achieve a predefined intervention goal, for example, a certain phenotypic response of the cell (Klamt et al. 2006). Similar as minimal cut sets in interaction graphs, a concept of *Minimal Intervention Sets* (MISs) can be introduced for logical models, and the resulting sets in interaction graphs and logical models tackling the same target nodes are naturally correlated. However, whereas minimal cut sets in interaction graphs are restricted to questions regarding signaling paths and feedback loops (e.g., “How can

all negative feedback loops be interrupted?”), in logical models a certain functional behavior (state) should be achieved. Typical problems that can be addressed by the computation of MISs in logical models are the identification of drug targets, the identification of failure modes that might cause an observed pathological behavior (diagnosis problem), and the identification of nodes that are of central importance for a certain biological function. Furthermore, MISs can be used to identify necessary changes in a proposed network structure to remove inconsistencies between model predictions and data (Saez-Rodriguez et al. 2007).

In the logical model of the example network (Figure 2.3), we looked for interventions to activate ERK ($ERK = 1$) and deactivate PI3K ($PI3K = 0$) in presence of all receptors. The identified MISs are given in Table 2.2. At least three interventions are required. MEK has to be set to 1 in all MISs: a more upstream intervention with the goal to activate ERK would at the same time lead to activation of PI3K through Ras, although deactivating PI3K is desired. In order to achieve $PI3K = 0$, two further interventions are required: (i) set NRG1 or ERBB23 to 0, and (ii) set EGF, ERBB1, GRB2, GAB1, or PIP₃ to 0.

Table 2.2: Minimal intervention sets in the EGF/NRG1 logical model.

<i>Intervention goal: erk = 1, pi3k = 0</i>											
<i>Side constraints: erbb1 = 1, erbb2 = 1, erbb3 = 1</i>											
	egf	nrg1	erbb11	erbb13	erbb23	grb2	sos	ras	mek	gab1	pip3
1	0	0							1		
2		0	0						1		
3		0				0			1		
4		0							1	0	
5		0							1		0
6	0				0				1		
7			0		0				1		
8					0	0			1		
9					0				1	0	
10					0				1		0

Shown are the computed minimal intervention sets to activate ERK and deactivate PI3K in presence of ERBB1, ERBB2, and ERBB3 (see Figure 2B). An entry 1 means constitutive activation, whereas 0 indicates a required deactivation. Interventions for species with fixed values (i.e., ERBB1, ERBB2, ERBB3) and the target species ERK and PI3K have not been considered.

2.4.3 Ordinary Differential Equation Modeling

Cells behave as nonlinear dynamical systems (Sontag 2005). Thus, in order to derive a comprehensive quantitative understanding, dynamical modeling approaches are indispensable. The most accurate descriptions of signaling pathways are mechanistic models (also termed physicochemical models). Characteristic for these models is that equations refer to identifiable processes such as phosphorylations, and parameters can be interpreted physically, representing, for example, binding affinities or concentrations (Aldridge et al. 2006). The mathematical formulations of physicochemical models are in terms of ordinary or partial differential equations, which can be deterministic or stochastic. To date, the most widely used representation of biological pathways are coupled systems of ordinary differential equations (ODEs) using mass-action kinetics for the rates of production and consumption of the biomolecular species (Becker et al. 2010; Chen et al. 2009; Huang and Ferrell 1996; Kholodenko et al. 1999). However, this type of modeling requires sufficient knowledge of biological mechanisms and parameters, what limits its applicability to well-characterized networks of small to medium size. In the following, we will discuss comparatively simpler differential equation models and parameter-free approaches in the context of ODE modeling.

2.4.3.1 Differential Equation Models Derived from Boolean Models

An alternative to physicochemical ODE models are ODE models where mass-action kinetics are approximated by other kinetics, for example, Michaelis-Menten or Hill-type kinetics, what reduces model complexity, however, at the expense of loss of mechanistic details.

Among the simplest ODE models are ODE models that are continuous counterparts of Boolean models and as such closing the gap between qualitative and mechanistic dynamical modeling (Mendoza and Xenarios 2006; Wittmann et al. 2009). One possibility to translate Boolean models into ODE models is to use multivariate polynomial interpolation (Wittmann et al. 2009). Continuous counterparts of the discrete Boolean functions can be derived, for example, by linear interpolation of the functions. Alternatively, linear interpolation can be performed after applying Hill functions to the function arguments (Wittmann et al. 2009). The resulting ODE models must be considered as phenomenological models, whose reactions and parameters lack a physical interpretation. Nevertheless, these models can be fitted against time-resolved experimental data and can, in principle, be used to explain

and predict the quantitative and dynamic behavior of the system. As an example, an ODE model derived from a Boolean model of T cell activation has been shown to be able to reproduce time courses in response to different ligand concentrations and to predict binding affinities of different ligands (Wittmann et al. 2009).

Already in 1973, systems of piecewise-linear (PL) differential equations were introduced as continuous counterparts of Boolean models (Glass and Kauffman 1973). In a PL model, the Boolean functions are replaced by sums and products of step functions. The use of step functions implies discontinuities hampering the simulation of a PL model with standard numerical integrators. Instead, PL models can be analyzed by means of state transition graphs, that is, in a similar manner as the qualitative dynamics in logical networks are studied (De Jong et al. 2004).

2.4.3.2 Qualitative Analysis of Dynamical Models

A major challenge in ODE modeling is that a sufficient amount of quantitative data is needed, while many of the available data are qualitative or semi-quantitative. Within the last years, there has been substantial effort to improve the experimental setup and subsequent data analysis in order to generate quantitative data (e.g., Schilling et al. (2005)). Furthermore, analytic methods such as mass spectrometry, which enable the collection of large amounts of quantitative data, are advancing forward. Nevertheless, mathematical methods that allow incorporating qualitative information in ODE modeling are still desirable. An example for such an effort is the approach presented in Rumschinski et al. (2012), which enables the invalidation of ODE models based on qualitative and semi-quantitative data.

Qualitative information about a biological system can be available at different levels. Besides information about the interactions between different players, for example, “addition of species A increases the activity of species B”, qualitative knowledge about more global system properties might exist. For example, one might know from experimental observations that a system exhibits bistability. Obviously, a dynamical model reflecting the system should also be able to reflect the observed properties. A common strategy is to explore parameter spaces and initial conditions, which have been estimated before based on experimental data, by means of numerical simulations. This is often combined with methods such as bifurcation analysis that enable to determine the classes of dynamical behavior a parameterized model can produce (Aldridge et al. 2006; Sontag 2005). However, the determination of parameter values, especially in large systems, is very difficult. High-dimensional

parameter and state spaces give rise to high-dimensional and nonlinear optimization problems, which are difficult to solve. In addition, incomplete, noisy, and often only semi-quantitative data contribute to parameter uncertainty. Furthermore, conclusions from numerical simulations are local and it is difficult to derive a fundamental understanding (Radde et al. 2010; Sontag 2005). This again shows the desirability of parameter-free, qualitative approaches, here in the context of ODE models. The typical question that is tackled by these approaches is whether a given ODE network structure is able to exhibit, for some parameter values, a certain observed dynamical behavior.

Qualitative ODE models. The structure of an ODE system $\dot{x} = f(x)$ is often analyzed by considering the sign pattern of the Jacobian matrix J with entries $J_{ij}(x) = \frac{\partial f_i(x)}{\partial x_j}$. In a biological system, the state variables x_i usually denote the species concentrations, and functions f_i describe the rate of production and consumption of the species. A monotonicity assumption on the rate functions f_i is often required to guarantee that the sign pattern of the Jacobian is constant over all time points. A set of models whose sign patterns of the Jacobian are identical can be considered as a *qualitative model* (Radde et al. 2010). The qualitative information determining the sign pattern of the Jacobian matrix is often available in biological systems. A special class of ODE models are dynamic models that derive from chemical reaction networks. A dynamic chemical reaction network is a system of the form $\dot{x} = Nv(x)$, where N is known as the *stoichiometric matrix* of the system, and v is a vector of reaction rates. Assuming that the reaction rates fall into certain classes, for example, mass-action kinetics, the network structure is solely represented by the stoichiometric matrix N , as conclusions on the sign pattern of the Jacobian matrix can in this case be derived from the entries in N .

Graphical representation of ODE models. As already stated in Section 2.4.1, the structure of an ODE system can be represented by an interaction graph. Considering a system $\dot{x} = f(x)$ with x_i denoting the concentration of species S_i , the nodes in the interaction graph underlying the ODE model represent the species S_i . A positive edge from S_j to S_i is drawn in the interaction graph if $\frac{\partial f_i}{\partial x_j}(x) > 0$ for some x , and a negative edge is drawn if $\frac{\partial f_i}{\partial x_j}(x) < 0$ for some x . Note that, in general, the interaction graph constructed in this way can contain both a positive and a negative edge between a pair of species nodes. Under certain monotonicity assumptions on the rate functions f_i , the interaction graph only contains a single edge of unique sign between each pair of nodes. Many functions describing regulatory functions and

chemical kinetics show the requested monotonicity property; hence, their structure is represented by a constant interaction graph structure.

For dynamic chemical reaction networks, bipartite graphs containing species and reaction nodes have been shown to be a valuable representation. We will discuss these Species–Reaction graphs (SR graphs) and their relation to interaction graphs in more detail in Chapter 6.

In the following, we will briefly discuss multistationarity and static response to perturbations as examples of dynamical system properties that have been tackled with parameter-free approaches.

Multistationarity. Given an ODE system $\dot{x} = f(x)$, one can ask whether this system has several steady states, that is, whether the function f has several zeros. Considering systems arising in biology, one is usually interested in the existence of more than two *positive* steady states. Feinberg, Horn, and Jackson already introduced in the 1970s with their chemical reaction network theory an approach aiming at receiving results on the system behavior that are valid independently of chosen rate constants (Feinberg 1987; Horn and Jackson 1972). Feinberg’s deficiency zero and deficiency one theorems enable to exclude multistationarity based on the system’s structure (Feinberg 1987). In subsequent works, Feinberg and coworkers also derived structural conditions guaranteeing multistationarity (Ellison and Feinberg 2000). All these methods apply to chemical reaction networks with mass-action kinetics and were implemented in the Chemical Reaction Network Toolbox [4]. With subnetwork analysis, a complementary approach was provided by Conradi et al. (2007). This type of analysis also requires mass-action kinetics and enables the assertion of multistationarity in some cases where the Chemical Reaction Network Toolbox cannot decide about it.

Another class of methods is based on the attempt to find sufficient conditions for injectivity: only non-injective systems have the capacity for multiple equilibria. Craciun and Feinberg (2005) derived criteria for injectivity of a system which concern the determinant expansion of the system’s Jacobian matrix. The original results were obtained for chemical reaction networks with mass-action kinetics; additionally, it was assumed that each species can feed in and flow out of the system (“fully open” network; Craciun and Feinberg (2005)). In subsequent works, the results were generalized to chemical reaction networks with less-restrictive kinetics (Banaji et al. 2007; Wiuf and Feliu 2013) and to networks where some species were considered not to enter or leave the system (Craciun and Feinberg 2006a, 2010). In parallel, graphical methods were derived from the injectivity conditions on the Jaco-

bian. First, it was shown that a positive cycle in the interaction graph is necessary for multistationarity (Cinquin and Demongeot 2002; Gouzé 1998; Plahte et al. 1995; Snoussi 1998; Soulé 2003; Thomas and Kaufman 2001a,b). For chemical reaction networks, stronger graphical requirements were stated in terms of SR graphs (Banaji and Craciun 2010; Craciun and Feinberg 2006b), and this work was extended to general interaction networks (Banaji and Craciun 2009). Recently, Shinar and Feinberg (2012) introduced *concordant chemical reaction networks*. This class of networks is equivalent to the class of networks that give—under certain assumptions on the kinetics—rise to injective dynamical systems. Notably, the decision whether a chemical reaction network is concordant or not is not derived from the Jacobian matrix, but from algebraic properties based on the stoichiometric matrix. Algorithms that enable to decide about network concordance have been integrated in the previously mentioned Chemical Reaction Network Toolbox [4]. Additionally, graph-theoretic criteria for concordance based on SR graphs have been derived (Knight et al. 2015; Shinar and Feinberg 2013), thus connecting the theory of concordant networks to previous work (Banaji and Craciun 2010; Banaji and Craciun 2009; Craciun and Feinberg 2006b).

Static response to perturbations. Qualitative system properties are often studied by steady-state-shift experiments: given a system at steady state, one changes system parameters or inputs and observes the resulting change of the variables once the system reaches a new steady state. Steady-state-shift experiments have been shown to allow the recovery of chemical reaction mechanisms (Chevalier et al. 1993) and to unravel interactions between modules in gene regulatory and signaling networks (Kholodenko et al. 2002; Klinger et al. 2013). To predict the qualitative effect a certain perturbation has on the system variables, one often uses linear approximations. The deviation of a state variable is expressed by a local functional relation between this state variable and the directly perturbed variables, which follows from the implicit function theorem. However, results obtained from the linear system may be valid only if the perturbations are sufficiently small. In order to obtain global results, additional assumptions are usually required. One possibility is to restrict oneself to the study of systems whose structure excludes the existence of multiple steady states. In this case, the global existence of the implicit function is guaranteed (Radulescu et al. 2006; Siegel et al. 2006). Alternatively, one can consider a particular path, known as *homotopy path*, which connects the two steady states and whose analysis allows to derive global conclusions from local observations (Shiomura 1995). For this analysis it is required that the perturbation induces a continuous steady-

state shift. We will consider the application of homotopy methods to perturbation analysis in Chapter 6.

Siegel et al. (2006) developed a framework to connect the network topology of a given gene regulatory network to experimental data resulting from a steady-state-shift experiment. The dynamics of the system is supposed to be described by a set of ordinary differential equations, and the network topology is given by an interaction graph representing the Jacobian matrix. In many applications, the interaction graph is directly reconstructed from literature information or databases (see Section 2.4.1) without explicitly defining the ODE system. The data retrieved from the steady-state-shift experiment are assumed to provide the qualitative information whether the perturbation leads to an increased, decreased, or not significantly changed state of a measured species. In order to compare experimental results and network topology, Siegel et al. (2006) derived a system of linear qualitative equations from the interaction graph and the differential dynamics defined by it: the quantitative equations resulting, for example, from the implicit function theorem as indicated above were transformed into a linear system in the sign algebra (Kuipers 1994). Solutions of the system describe, under certain assumptions, the possible qualitative responses of the state variables to small and large perturbations (Siegel et al. 2006). Several approaches and application examples for network inference resulted from this work, all based on the basic result that the direction of change of the state of a variable must be explainable by the direction of change of at least one of its predecessor nodes in the interaction graph (Gebser et al. 2011; Guziolowski et al. 2009; Veber et al. 2008). In Chapter 3 of this work, we will present another framework for interrogating and training of interaction graphs, which is based on a similar *sign consistency* rule as the one derived by Siegel et al. (2006).

Predicting the possible effects of perturbations in a qualitative system also plays an essential role in fault diagnosis in chemical process systems. Again, the system structure is assumed to be given by an interaction graph, and consistency rules in terms of sign conditions are derived (Iri et al. 1979; Oyeleye and Kramer 1988). A quite intuitive rule is that an observed effect results from propagating the perturbation along the paths in the interaction graph (Iri et al. 1979). We already discuss this in terms of the dependency matrix (Section 2.4.1.2). However, when a variable exhibits certain types of complex dynamics resulting from negative feedback, a simple propagation along the paths is not possible any more (Oyeleye and Kramer 1988). This is accounted for in the dependency matrix by the notion of *weak* activators and inhibitors. Oyeleye and Kramer (1988) derived graph-theoretic conditions to iden-

tify variables that have the potential to show such a complex dynamic response. In Chapter 6, we revisit this work. In particular, we show how conditions for the class of chemical reaction networks can be derived.

3 | Data-Driven Interrogation and Training of Signaling Network Structures Based on Interaction Graphs

Although interaction graphs merely capture the positive and negative influences between pairs of species, their structure already constrains the possible qualitative behavior of the nodes in response to stimulations or perturbations. Thus, one can compare predictions derived from an interaction graph model with qualitative changes in the activation levels of certain components that are caused by introducing, for example, a ligand or inhibitor. In this chapter, we present two approaches for the analysis of experimental data based on interaction graphs. In both cases, one considers changes of the system caused by a perturbation added in steady state. The first method we will present is based on the dependency matrix (Section 3.1). This method is suited to analyze the qualitative transient system response, in particular the initial response to a perturbation. In Section 3.2, we present a method based on the concept on sign consistency, and which considers steady state shifts. By application to ERBB signaling in hepatocytes, we demonstrate how both analyses enable the identification of cell-type-specific discrepancies between model structure and experimental data. Furthermore, new hypotheses about the ERBB signaling topology and conclusions on missing or probably inactive interactions could be derived (Section 3.3).

In the first approach, model predictions are derived from the dependency matrix (Klamt et al. 2006). We already introduced the presented technique for analyzing phosphoproteomic data and its application to ERBB signaling in Samaga et al. (2009), and the descriptions in Section 3.3.2.2 were adapted from this publication.

The second approach is based on the concept of *sign consistency* (Gebser et al. 2011; Guziolowski et al. 2009; Siegel et al. 2006), for which we developed an integer linear programming (ILP) formulation. Within this framework, we defined optimization problems (i) to detect discrepancies between experimental data and model structure, (ii) to identify possible places in the network structure causing these dis-

crepancies and, (iii) to propose changes in the network structure to minimize them. These results have been derived in close collaboration with Ioannis Melas from the National Technical University of Athens, who made substantial contributions to the ILP formulations. We published the approach and its application to ERBB signaling in Melas, Samaga et al. (2013), and Sections 3.2 and 3.3.2.3 were derived from this work.

3.1 Data Analysis Based on the Dependency Matrix

The dependency matrix contains the information how the species of the interaction graph model influence each other (see Section 2.4.1.2). Based on this information, one can predict the qualitative effect a stimulation or perturbation of certain species has on the activation levels of other species of the model. We briefly describe this idea with the help of the EGF/NRG1 example model, whose dependency matrix is given in Figure 2.2. Stimulating the cell with EGF should result in increased phosphorylation levels of ERBB11 and ERBB13 compared to the unstimulated case as EGF is a strong activator of these two receptor dimers. In contrast, phosphorylation of ERBB23 should, according to the model, not be influenced by EGF stimulation. As another example, the phosphorylation level of ERK should be increased in response to EGF when considering the initial response after stimulation. However, as EGF is only a weak activator of ERK, at later time points the negative feedback loop might cause a decrease even beneath the phosphorylation level of ERK in the unstimulated cell; thus, no predictions can be derived from the dependency matrix for the later time points (see Section 2.4.1.2). Introducing an inhibitor blocking the MEK kinase activity should lead to an increase in Ras activity. As MEK is a weak inhibitor of Ras, this prediction is limited to the early time points. However, if we assume that MEK kinase activity is completely blocked by the inhibitor, Ras activity must be increased also at later time points as the negative feedback loop is disconnected. Finally, a change in the expression level of ERBB2 might lead to an increased, decreased, or unchanged state of a number of downstream nodes of which ERBB2 is an ambivalent factor: these nodes can be reached by positive paths running over ERBB23, while at the same time being targets of negative paths from ERBB2 via ERBB13. The qualitative response is in this case dependent on the strength of the respective paths, which cannot be revealed solely from the structural information represented by the interaction graph.

We will use this concept for the analysis of stimulus–response data. As starting

point for the data analysis, we assume we are given an interaction graph model, for example, a “master topology” of a signaling pathway that contains all reported interactions, and a set of experimental scenarios showing the up- or downregulation of measured readout nodes in response to perturbations of some nodes. Typically, such an experimental scenario is derived by comparing the measured responses of the readouts under two different treatments. As an example, one might compare the phosphorylation state of a particular readout in response to adding a ligand L with the phosphorylation state in response to adding the ligand L in combination with an inhibitor I. The perturbation characterizing the scenario is given by the difference of both treatments; in the example, this is the addition of the inhibitor I. A perturbation can be positive (e.g., the addition of a ligand) or negative (e.g., the addition of an inhibitor), and a scenario may contain several positive and/or negative perturbations.

Given a set of experimental scenarios, one can derive predictions for the qualitative response of the readouts to the perturbations of each scenario from the dependency matrix: (i) if all positively perturbed nodes are strong activators and all negatively perturbed nodes are strong inhibitors of the readout, the model predicts “increase” in response to the perturbation; (ii) if all positively perturbed nodes are strong inhibitors and all negatively perturbed nodes are strong activators of the readout, the model predicts “decrease”; (iii) if all perturbed nodes are neutral factors of the readout, the model prediction is “no change”; (iv) if all positively perturbed nodes are activators and all negatively perturbed nodes are inhibitors, and if at least one of the perturbed nodes is a weak activator/inhibitor, the initial response to the perturbation is “increase”, but, due to negative feedback loop(s), the response at later time points is not restricted by the model; (v) if all positively perturbed nodes are inhibitors and all negatively perturbed nodes are activators, and if at least one of the perturbed nodes is a weak activator/inhibitor, the initial response is “decrease”, whereas, for later time points, the response is not restricted by the model; (vi) in all other cases, the model does not restrict the response to the perturbation; this applies if one of the perturbed nodes is an ambivalent factor or if two perturbations have opposing effects on the readout.

We can compare the measured effect of the perturbations with the model predictions from the interaction graph to see whether the experimental data are in accordance with the model structure. In this way, it is possible to uncover inconsistencies between experimental results and the qualitative knowledge captured in the model in a systematic way, and, hence, to get insights into cell-type and context specific

network structures. We show this in two applications: in Section 3.3.2, the analysis of experimental data of ERBB signaling is studied, and in Chapter 5, the described concept is used in combination with ODE modeling to untangle MAPK and PI3K signaling in response to HGF.

In order to compare a measured effect with the model predictions derived from the dependency matrix, the effect has to be classified as increase, decrease, or no change. This can be achieved in different ways. We describe the data discretization in the context of the two applications (Section 3.3.2 and Chapter 5). Furthermore, the time point the measurements are taken has to be chosen carefully: the positive effect of a weak activator can, for example, only be predicted for the initial response.

3.2 Detection and Removal of Inconsistencies between Data and Signaling Network Topologies Using the Concept of Sign Consistency

The dependency matrix is particularly useful to get an overview on how a node can potentially influence the other nodes in the network; however, it may become limiting if multiple node values are measured in one experimental scenario. Given an interaction graph topology, state changes measured for certain nodes are, in general, not independent; thus, one can derive stronger constraints. This is particularly true if we assume that the system moves from one steady state to another. Consider the small example model given in Figure 3.1(a) and assume there would be another node Z that is activated by F (edge $F \rightarrow Z$). As A is an ambivalent factor for F and Z , we would predict from the dependency matrix that the levels of F and Z can decrease or increase in response to a perturbation in A ; however, it is not possible that their new steady-state levels change in different directions.

This is considered by a class of methods that are relying on a *sign consistency rule* and that have also been developed to detect discrepancies between topology and experimental data (Gebser et al. 2011; Guziolowski et al. 2009; Siegel et al. 2006). The key idea is that, in a steady state shift experiment, the direction of change of the state of a node must be explainable by the direction of change of at least one of its predecessor nodes (except for the directly perturbed node(s)). For example, in Figure 3.1(a), after a perturbation in A , the steady-state level of F may have become larger only if E decreased its activation level (as E inhibits F) or/and if C increased its level (as C activates F). The sign consistency rule gives rise to constraints on

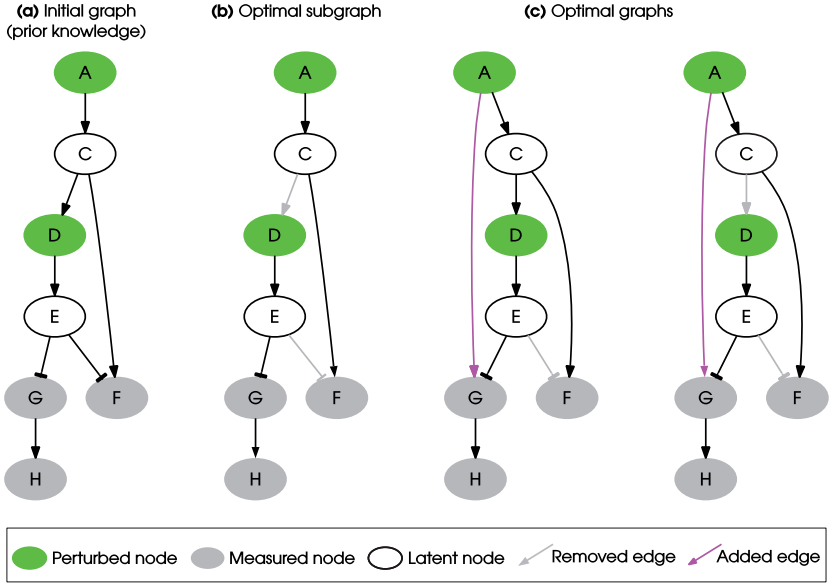


Figure 3.1: A simple example network used for illustration purposes. The interaction graph consists of seven nodes and seven edges. The green nodes A and D can be perturbed externally; the gray nodes F, G, and H are the readouts of the network, whose activation states are measured in the experiments; the white nodes C and E are latent nodes, which are neither perturbed nor measured (see scenarios in Table 3.1). **(a)** The initial topology of the interaction graph representing the prior knowledge. This graph produces a total fitting error of 5 over the three scenarios in Table 3.1. **(b)** The unique optimal subgraph of (a) minimizing the total fitting error of the experimental scenarios to 2 (see Table 3.1). **(c)** Two optimal graphs obtained from (a) by applying OPT_GRAPH: by adding edge $A \rightarrow G$ and either (left) removing $E \rightarrow F$ or (right) removing $E \rightarrow F$ and $C \rightarrow D$, the fitting error is eradicated completely and becomes 0 (cf. Table 3.1). Figure adapted from Melas, Samaga et al. (2013).

the possible patterns of ups and downs of the nodes' activation levels in a given interaction graph. These constraints can be encoded, for example, by Answer Set Programming (Gebser et al. 2011). Confronting these constraints with experimental data may then lead to the detection of topological inconsistencies, namely if no sign pattern complying with the given measurements and perturbations can be found (Gebser et al. 2011; Guziolowski et al. 2009; Siegel et al. 2006).

The methods we will present in this section are based on a similar sign consistency rule; however, they differ in a number of aspects. First, we will encode the sign constraints as an Integer Linear Programming (ILP) problem. This formulation gives us the opportunity to utilize the large corpus of effective algorithms developed for ILP problems. Second, we explicitly allow a "0" change to mark non-affected node states. Furthermore, for the situation that multiple stimulus–response experiments are available, we will address aspects that go beyond the detection of inconsistencies from single experiments, namely to correct a given network structure such that the number of mismatches is minimized. For the structure optimization process we will consider edge removals as well as edge additions.

3.2.1 General Framework

As in Section 3.1, we assume we are given an initial interaction graph topology and a set of experimental scenarios in each of which some nodes were perturbed and the resulting up- or downregulation of some readout nodes was measured in the new steady state. Figure 3.1(a) and the three experimental scenarios in Table 3.1 defined by the columns *Perturbations* and *Measurements* provide an illustrative example. Here, A and D are nodes that can be perturbed; F, G, and H are the readout nodes, which are measured, and C and E are latent nodes, which are neither perturbed nor measured.

Our goal is now to analyze and improve the consistency of an interaction graph topology with respect to a given set of experimental data. Central to all algorithms presented herein is the following definition of sign consistency.

Definition 3.1. (Sign Consistency). Let G denote an interaction graph with node set V , edge set A , and sign mapping σ . Furthermore, let s denote a node labeling (sign pattern), which stores for each node $v_i \in V$ a sign $s_i \in \{-1, 0, 1\}$. We say that s is *sign consistent with respect to* G if the following conditions hold for each node $v_i \in V$:

- a) If $s_i = -1$: either s_i was fixed to -1 (perturbed node), or there is a predecessor node $v_j \in V$ and an edge $a = (v_j, v_i)$, $a \in A$, with $\sigma(a) \cdot s_j = -1$.

- b) If $s_i = 1$: either s_i was fixed to 1 (perturbed node), or there is a predecessor node $v_j \in V$ and an edge $a = (v_j, v_i)$, $a \in A$, with $\sigma(a) \cdot s_j = 1$.
- c) If $s_i = 0$: either (i) s_i was fixed to 0, or (ii) v_i has no predecessor, or (iii) for all $v_j \in V$ with $a = (v_j, v_i)$, $a \in A$, we have $s_j = 0$, or (iv) there is a predecessor node $v_j \in V$ and an edge $a = (v_j, v_i)$, $a \in A$, with $\sigma(a) \cdot s_j = -1$, and another predecessor node $v_k \in V$ and an edge $b = (v_k, v_i)$, $b \in A$, with $\sigma(b) \cdot s_k = 1$.

In our setting, the signs of the external perturbations as well as the measured signs of the readout nodes can be described by a specific node labeling, which we call the *associated labeling* of the scenario. In realistic applications one usually has latent nodes, which are neither perturbed nor measured; hence, the associated node labeling of an experimental scenario may contain unknown values, which we denote by NaN. We call incomplete sign patterns *partial labelings*. A partial labeling \tilde{s} is sign consistent if there exists a complete sign consistent labeling s for which we have $\tilde{s}_i = s_i$ wherever $\tilde{s}_i \neq \text{NaN}$. In this sense, we say that an experimental scenario is sign consistent if its associated (partial) labeling is sign consistent. Finally, if we have a collection of scenarios, we say that this collection is sign consistent with the interaction graph if all the (partial) labelings associated with the scenarios are sign consistent.

We can now consider four fundamental problems on the consistency of experimental scenarios with respect to a given interaction graph:

(1) SCEN_FIT. Given a single experimental scenario, we fix the states of the perturbed nodes according to the experimental interventions and search for a sign consistent node labeling showing the minimum mismatch with the given measurements. We define the *fitting error* ε_{fit} as the sum of the absolute difference between the measurements m_i and the optimal sign pattern s : $\varepsilon_{\text{fit}} = \sum_{i: m_i \neq \text{NaN}} |m_i - s_i|$. In the ideal case, where the associated labeling of the experimental scenario is sign consistent, the fitting error will be 0.

From Figure 3.1(a)/Table 3.1, we see that scenario 1 is sign consistent: A was externally increased and D decreased, and with $s_A = s_C = s_G = s_F = s_H = 1$ and $s_D = s_E = -1$, we obtain a sign consistent labeling giving us a possible explanation for the measurements. In contrast, scenario 2 is not consistent with the interaction graph topology: if D is increased externally (no perturbation in A), then we expect to see a decrease in F, G, and H, which is not seen in F (unchanged). Thus, the minimum resulting fitting error for an optimal sign pattern is 1. Generally, an error

Table 3.1: Example scenarios and optimizations for the example network in Figure 3.1.

	Perturbations		Measurements			Initial fitting error (Fig. 1A)			MCoS	Remaining fitting error (Fig. 1B/Fig. 1C)		
	A	D	F	G	H	F	G	H		F	G	H
Sc1	1	-1	1	1	1	0	0	0		0/0	0/0	0/0
Sc2		1	0	-1	-1	1	0	0	{1 → F}, {1 → C}, {1 → A}	0/0	0/0	0/0
Sc3	1		1	1	1	0	2	2	{1 → G}, {-1 → E}, {-1 → D}, {-1 → C}	0/0	1/0	1/0

Rows *Sc1*, *Sc2*, *Sc3* correspond to scenarios 1 to 3. Columns *Perturbations* show the externally imposed states of the nodes A and D, which can be -1 (downregulation), 0 (state of the node did not change), or 1 (activation level is increased). No value is given if the node was not perturbed. Columns *Measurements* show the measured change of the activation levels of F, G, and H in the respective scenarios. Columns *Initial fitting error* show the total mismatch of predictions and measurements (derived with SCEN_FIT) with respect to the initial topology shown in Figure 3.1(a). Column *MCoS* (minimal correction sets) shows artificial positive (1) or negative (-1) external inputs to some nodes, which would lead to a perfect fit of the data (resulting fitting error for the scenario becomes 0). Columns *Remaining fitting error* show the remaining mismatches for the optimal subgraph depicted in Figure 3.1(b) and for the two optimal graphs displayed in Figure 3.1(c). The original network in Figure 3.1(a) has a total fitting error of 5; it is 2 for the optimal subgraph in Figure 3.1(b), and it becomes 0 in the optimal graphs in Figure 3.1(c). Table adapted from Melas, Samaga et al. (2013).

of 1 or -1 occurs if a change was expected/not expected, but was not seen/was seen in the experiments. For scenario 3, the predictions are even worse: increase in A (no perturbation in D, which thus depends on C) should lead to downregulation of G and H, but an increase is measured for both. Thus, we get an absolute error of 2 for each of the two predictions. It follows that the fitting error of a sign consistent node labeling closest to scenario 3 cannot be smaller than 4.

It may happen that several solutions exist, all explaining a given scenario equally well. For example, assume again that there was another node Z in Figure 3.1(a) that is activated by F through an edge $F \rightarrow Z$. If we now measured $G = H = F = -1$ and $Z = 1$ after positively perturbing A ($A = 1$), then the best scenario fit would result in an error value of 2 since F and Z must have the same value. However, there are three optimal solutions regarding F and Z, namely $F = Z = 0$, $F = Z = 1$, and $F = Z = -1$, all leading to the same minimum fitting error of 2. For some applications it will

be helpful to know all these optimal solutions; therefore, we will also address their enumeration.

(2) Minimal Correction Sets (MCoS). Another optimization problem for a single scenario directly follows if a given scenario is not sign consistent, that is, if no sign consistent labeling can be found that results in a fitting error of 0. We can then try to identify a minimal set of nodes whose states need to be corrected to obtain a consistent scenario. The correction of a node's state is simulated by adding an additional external input that is either 1 or -1 . We call these sets *minimal correction sets* (MCoS), the minimality property demanding that no subset of a MCoS would lead to a consistent labeling. For example, regarding scenario 3 in Table 3.1, there are four MCoS suggesting that there was either an external upregulation of G ($1 \rightarrow G$), or a downregulation in one of the nodes E, D, or C, each of unknown cause. Thus, MCoS show possible places in the network that have a high probability to cause the observed inconsistencies. With the MCoS problem we identify the enumeration of MCoS of minimum size for a given scenario. A simple extension not considered herein is to enumerate all MCoS irrespective of their size.

(3) OPT_SUBGRAPH. The first two problems focus on a single scenario; now we intend to optimize the network structure in such a way that the total fitting error over *all* scenarios is minimized. Initially, we allow only the removal of edges in the network, that is, we search for an optimal subgraph. In addition, we may also be interested in an enumeration of all subnetworks minimizing the number of inconsistencies between interaction graph topology and data. As an example, Figure 3.1(b) shows the unique optimal subgraph of the original interaction graph in Figure 3.1(a) minimizing the fitting error over all three scenarios in Table 3.1. This solution reduces the total fitting error from 5 to 2, and there is no solution that could reduce it further.

(4) OPT_GRAPH. The removal of certain edges may significantly improve the agreement between measurements and network topology, but some fitting errors can often only disappear if we have also the opportunity to add new interactions. Therefore, this fourth optimization problem intends to minimize the fitting error by allowing edge removals and insertions in parallel. Obviously, the fit cannot be worse than the one obtained by problem (3). For smaller networks, a full enumeration of all optimal solutions might be possible. However, as the insertion of new interactions

increases the solution space dramatically in large networks, we consider a greedy strategy which determines, in each iteration, the optimal edge whose inclusion—in combination with the pruning step (3)—decreases the fitting error the most. One may then add this edge permanently and repeat the algorithm described above until no further significant improvement can be obtained by inserting a new edge.

Figure 3.1(c) shows a result of this optimization step in our example: the edge $A \rightarrow G$ is identified as missing edge which, in combination with a pruning step, completely eradicates the original fitting errors in all scenarios. Thus, the resulting network is fully consistent with the entire set of experimental data. In this example, nine other edges can be identified whose addition, in combination with a pruning step by `OPT_SUBGRAPH`, lead to a fitting error of 0. Furthermore, for each added edge, the `OPT_SUBGRAPH` problem that is called after adding the edge might return several optimal solutions. Figure 3.1(c) shows the two existing optimal solutions (with a fitting error of 0) that are derived after adding edge $A \rightarrow G$.

Positive Cycles. Feedback loops often hamper the analysis of causality, and many network inference techniques therefore exclude cycles from the network or assume that no cycles exist (e.g., Sachs et al. (2005) and Saez-Rodriguez et al. (2009)). In contrast to many other approaches, our method can readily deal with negative cycles without any problems. However, positive cycles may become problematic as they can provide explanations for state changes without any external perturbation. A simple example for such “self-explaining” state changes is the following network: $A \rightarrow B \rightarrow C \rightarrow B$ (all edges are positive). Node A would normally serve as an input. However, assuming that A has not changed, a measured increase of B would be explainable by the sign-consistent labeling $(0, 1, 1)$, that is, B activates C , which then activates B again. Although such a shift without external perturbations could indeed happen in realistic systems (due to fluctuations in bistable systems), we recommend that the initial interaction graph should not contain a positive feedback—otherwise, many observations might become sign consistent just through the existence of positive cycles. This is also the reason why a new candidate edge can only be added to the network if it does not give rise to a new positive cycle. In many applications, this requirement is not a real limitation, in particular when describing early events in signaling networks.

3.2.2 Formulation as Optimization Problems

The four problems comparing experimental data and a given interaction graph model structure based on sign consistency, which were introduced above, can be encoded as integer linear programming (ILP) problems. In this section, we describe the objective functions and constraints for each problem. The ILP formulations, where the constraints are translated into linear equalities and inequalities, are given in Appendix A.

3.2.2.1 Basic Definitions and Formulation of Sign Consistency

As described above, we assume that we are given an interaction graph with node set $V = \{v_1, \dots, v_{n_V}\}$, edge set $A = \{a_1, \dots, a_{n_A}\}$, and sign function σ , which captures our prior knowledge on the signaling topology. Additionally, we are given a set of n_S experimental scenarios, each consisting of a specific set of perturbed nodes and a set of measurements. We will denote with $I_V := \{1, \dots, n_V\}$ the index set of the nodes, with $I_E := \{1, \dots, n_E\}$ the index set of the edges, and with $I_S := \{1, \dots, n_S\}$ the index set of the experimental scenarios. The experimental scenarios are specified by two matrices: (i) the $n_V \times n_S$ perturbation matrix \mathbf{p} with $p_{j,k} \in \{-1, 0, 1, \text{NaN}\}$ storing the enforced state of node v_j in scenario k through external perturbation, and (ii) the $n_V \times n_S$ measurement matrix \mathbf{m} with $m_{j,k} \in \{-1, 0, 1, \text{NaN}\}$ storing the measured change of the steady-state level of node v_j in scenario k . Thus, perturbation and measurement values indicate enforced/measured upregulation (1), downregulation (-1), or unchanged state (0). Usually, only a small subset of nodes is perturbed, and only a subset of nodes can be measured; therefore, unperturbed and non-measured states are marked by NaN.

In the ILP formulation of sign consistency, we will represent the predicted state of a node v_j in experiment k by an integer variable $x_{j,k} \in \{-1, 0, 1\}$. Again, $x_{j,k} = 1$ encodes increase and $x_{j,k} = -1$ decrease of node v_j in scenario k , whereas $x_{j,k} = 0$ indicates that the activation level of v_j remained unchanged.

We introduce the binary variables $u_{i,k}^+$ and $u_{i,k}^-$ to represent the potential of edge $a_i = (v_{i_t}, v_{i_h})$, $i_t, i_h \in I_V$, to up- or downregulate its head node v_{i_h} in experiment k . Furthermore, we let σ_i denote the sign of edge a_i , that is, $\sigma_i := \sigma(a_i)$. With that we can formulate the first constraint for a sign-consistent node labeling.

Constraint 1. Edge a_i has the potential to upregulate its head node v_{i_h} in experiment k (i.e., $u_{i,k}^+ = 1$) if and only if $\sigma_i \cdot x_{i_t,k} = 1$. In any other case we have $u_{i,k}^+ = 0$.

Accordingly, edge a_i has the potential to downregulate its head node v_{i_h} in experiment k (i.e., $u_{i,k}^- = 1$) if and only if $\sigma_i \cdot x_{i_t,k} = -1$. In any other case $u_{i,k}^- = 0$.

Next, we introduce the two binary variables $x_{j,k}^+$ and $x_{j,k}^-$ to represent the potential for node v_j to be increased or decreased depending on the activity of its upstream edges.

Constraint 2. Node v_j has the potential to be increased (i.e., $x_{j,k}^+ = 1$) if and only if an edge $a_i = (v_{i_t}, v_j)$, $i_t \in I_V$, exists with $u_{i,k}^+ = 1$. Node v_j has the potential to be decreased (i.e., $x_{j,k}^- = 1$) if and only if an edge $a_i = (v_{i_t}, v_j)$, $i_t \in I_V$, exists with $u_{i,k}^- = 1$.

The state $x_{j,k}$ of node v_j in scenario k is constrained by the values of $x_{j,k}^+$ and $x_{j,k}^-$ according to the definition of sign consistency (see Definition 1):

Constraint 3. (i) Node v_j may be increased ($x_{j,k} = 1$) if it has the potential to be increased ($x_{j,k}^+ = 1$). (ii) Node v_j may be decreased ($x_{j,k} = -1$) if it has the potential to be decreased ($x_{j,k}^- = 1$). (iii) Node v_j may stay unchanged ($x_{j,k} = 0$) if it has the potential to be both increased and decreased ($x_{j,k}^- = x_{j,k}^+ = 1$) or neither of the above ($x_{j,k}^- = x_{j,k}^+ = 0$).

A node labeling that is sign consistent according to Definition 3.1 must satisfy Constraints 1–3. The formulation with inequality constraints is given in Appendix A.2.

3.2.2.2 SCEN_FIT Optimization Problem

The goal of SCEN_FIT is to identify, for a given scenario k , a sign-consistent vertex labeling that is closest to the measurements of this scenario. We first have to constrain the values of the perturbed nodes in scenario k :

$$x_{j,k} = p_{j,k}, \quad \forall j \text{ with } p_{j,k} \neq \text{NaN}. \quad (3.1)$$

Realistic perturbations typically affect either input nodes (e.g., ligands) or internal nodes in the case where a specific inhibitor was added or where a constitutive activation or a knock-in/knock-out was introduced. The state of the perturbed nodes are thus fixed to the enforced value, and Constraint 3 is omitted for these nodes to preserve the consistency of the formulation.

We now search for a sign-consistent labeling $x_k = (x_{1,k}, \dots, x_{n_V,k})$ that minimizes the measurement-prediction-mismatch. This can be formulated as optimization

problem as follows:

$$\text{minimize}_{x_k} \sum_{\substack{j \in I_V \\ m_{j,k} \neq \text{NaN}}} |m_{j,k} - x_{j,k}| \quad (3.2)$$

subject to Constraints 1–3.

Thereby, the summation of mismatches is done over all nodes for which measurements exist.

As discussed in Section 3.2.1, we also consider the enumeration of *all* optimal SCEN_FIT solutions for a given scenario. To this end, we solve Problem (3.2) repeatedly, and, after each run, we exclude previously found solutions by adding the following constraints for each previous solution s :

$$\sum_{j \in I_V} |x_{j,k} - x_{j,k,s}| \geq 1, \quad (3.3)$$

where $x_{j,k,s}$ represents the value of $x_{j,k}$ in solution s . To ensure that only solutions with minimum fitting error are found, we replace, after the first iteration, the objective function in (3.2) by forcing instead the algorithm to find solutions with the same minimum fitting error objval as in the first run:

$$\sum_{\substack{j \in I_V \\ m_{j,k} \neq \text{NaN}}} |m_{j,k} - x_{j,k}| = \text{objval}. \quad (3.4)$$

The resulting problem becomes thus a simple search for a feasible solution and is repeated until no further solution can be found.

3.2.2.3 MCoS Optimization Problem

Computing a Single Minimal Correction Set. Next, we address the identification of a minimal correction set (MCoS) for a sign-inconsistent scenario k , that is, for a scenario where the fitting error after optimization is greater than zero. An MCoS indicates possible causes of discrepancies between measured data and assumed interaction graph topology. As described in Section 3.2.1, MCoS correspond to artificial perturbations of certain nodes, which render the measurements from a given scenario consistent with the given network.

Let a new set of binary variables $B_{j,k}^+$ and $B_{j,k}^-$ denote these artificial perturbations. Constraint 3 is then modified in the following way:

Constraint 4. (i) Node v_j may be increased ($x_{j,k} = 1$) if $x_{j,k}^+ = 1$ or if a positive input is added, that is, $B_{j,k}^+ = 1$. (ii) Node v_j may be decreased ($x_{j,k} = -1$) if $x_{j,k}^- = 1$ or if a negative input is added, that is, $B_{j,k}^- = 1$. (iii) Node v_j may stay unchanged ($x_{j,k} = 0$) if $x_{j,k}^- = x_{j,k}^+ = 1$, $x_{j,k}^- = B_{j,k}^+ = 1$, $B_{j,k}^- = x_{j,k}^+ = 1$, or if $x_{j,k}^- = x_{j,k}^+ = 0$.

Having introduced the correction terms $B_{j,k}^+$ and $B_{j,k}^-$, we set as an extra constraint the perfect fit of the data (which is now always feasible):

$$\sum_{\substack{j \in I_V \\ m_{j,k} \neq \text{NaN}}} |m_{j,k} - x_{j,k}| = 0. \quad (3.5)$$

As we are interested in MCoS with a *minimum* number of corrections, we consider the following optimization problem, where $B_k^+ = (B_{1,k}^+, \dots, B_{n_V,k}^+)$ and $B_k^- = (B_{1,k}^-, \dots, B_{n_V,k}^-)$:

$$\begin{aligned} & \underset{B_k^+, B_k^-}{\text{minimize}} \sum_{j \in I_V} (B_{j,k}^+ + B_{j,k}^-) \\ & \text{subject to (3.5) and Constraint 1, 2, and 4.} \end{aligned} \quad (3.6)$$

Enumeration of Minimal Correction Sets. In general, several MCoS with minimum size might exist; therefore, we address in this subsection the enumeration of *all* minimum MCoS. To this end, we solve the optimization problem (3.6) repeatedly, and, after each run, we exclude previously found solutions by adding the following constraints for each previous solution s :

$$\sum_{j \in I_V} (|B_{j,k}^+ - B_{j,k,s}^+| + |B_{j,k}^- - B_{j,k,s}^-|) \geq 1, \quad (3.7)$$

where $B_{j,k,s}^+$ and $B_{j,k,s}^-$ represent the value of $B_{j,k}^+$ and $B_{j,k}^-$ in solution s . To focus only on MCoS with the minimum number of corrections, we replace, after the first iteration, the objective function in (3.6) by forcing the algorithm to find a solution with the same minimum number of corrections objval as in the first run:

$$\sum_{j \in I_V} (B_{j,k}^+ + B_{j,k}^-) = \text{objval}. \quad (3.8)$$

Again, the resulting problem becomes a simple search for a feasible solution and is repeated until no further solution can be found.

3.2.2.4 OPT_SUBGRAPH Optimization Problem

As stated in Section 3.2.1, OPT_SUBGRAPH searches for an optimal subgraph of the original topology, that is, for a set of suitable edge removals minimizing the total fitting error *over all* scenarios. First, we describe how we can identify one particular solution to this problem before turning to the enumeration of optimal subgraphs.

Computing a single optimal subgraph. The removal of edges is implemented using binary variables y_i . The algorithm will set $y_i = 1$ if the edge a_i is removed by the optimization procedure to improve the fit of the data, otherwise $y_i = 0$. We impose again the Constraints 1–3 for sign consistency. The actual pruning is implemented by modifying Constraint 1 as follows:

Constraint 5. Edge a_i has the potential to upregulate its head node v_{i_h} in experiment k (i.e., $u_{i,k}^+ = 1$) if and only if $\sigma_i \cdot x_{i_v,k} - y_i = 1$. In any other case we have $u_{i,k}^+ = 0$. Accordingly, edge a_i has the potential to downregulate its head node v_{i_h} in experiment k (i.e., $u_{i,k}^- = 1$) if and only if $\sigma_i \cdot x_{i_v,k} + y_i = -1$. In any other case $u_{i,k}^- = 0$.

As for SCEN_FIT, we want to minimize the measurement-prediction-mismatch, but now minimize *over all* scenarios.

$$\underset{x_1, \dots, x_{n_s}}{\text{minimize}} \quad \sum_{\substack{(j,k) \in I_V \times I_S \\ m_{j,k} \neq \text{NaN}}} |m_{j,k} - x_{j,k}| \quad (3.9)$$

subject to Constraints 2, 3, and 5.

Usually, many optimal solutions may exist yielding the same residual fitting error. One might then be interested to focus on particular solutions, for example, on those containing the minimal/maximal number of edges in the remaining subgraph (Melas, Samaga et al. 2013).

Enumeration of Optimal Subgraphs. To identify all optimal subgraphs minimizing the inconsistencies between interaction graph topology and measurements for all scenarios, we solve the optimization problem (3.9) repeatedly, and, after each run, we exclude previous solutions s by adding the following constraint:

$$\sum_{i \in I_E} |y_i - y_{i,s}| \geq 1, \quad (3.10)$$

where $y_{i,s}$ represents the value of y_i in solution s . Moreover, after the first run, we replace the objective function in (3.9) by enforcing the algorithm to obtain the same, optimal, goodness of fit $objval$ as in the first run:

$$\sum_{\substack{(j,k) \in I_Y \times I_S \\ m_{j,k} \neq NaN}} |m_{j,k} - x_{j,k}| = objval. \quad (3.11)$$

3.2.2.5 OPT_GRAPH Optimization Problem

As motivated in Section 3.2.1, optimizing the interaction graph topology by edge removals may eliminate some, but often not all mismatches. One reason could be that some real effects cannot be transduced in the model due to missing edges. We therefore propose an algorithm suggesting new interactions, whose addition would minimize the fitting error. As the possibility to insert new interactions increases the solution space dramatically in large networks, we consider the following greedy strategy: for each interaction not contained yet in the interaction graph, we temporarily insert this edge and determine the resulting optimal solution for the fitting error by applying the OPT_SUBGRAPH algorithm introduced above. The single interaction that reduces the fitting error the most is picked by the greedy algorithm and permanently inserted in the interaction graph. This process is repeated until no further edge exists that could improve the goodness of fit to the data significantly (significance can be quantified by a certain threshold). Importantly, at the beginning of each iteration, a list of eligible edges is computed consisting only of those edges that do not form a positive cycle.

3.3 Application to ERBB Signaling

In the following, we describe the application of the above described methods for analyzing experimental data based on the interaction graph model of ERBB signaling introduced in Section 2.3.1. We presented the model as well as the data analysis based on the dependency matrix first in Samaga et al. (2009). The analysis based on sign consistency was a follow-up study, which we published in Melas, Samaga et al. (2013).

3.3.1 The Interaction Graph Model of ERBB Signaling

As already mentioned in Section 2.3.1, we present in this work a large-scale interaction graph model of ERBB signaling (Figure 3.2), which was based on the EGFR pathway map (Oda et al. 2005) and additional information from the literature. From this interaction graph model, we derived a logical model, which will be studied in Section 4.4. Our model contains most parts of the stoichiometric EGFR pathway map (Oda et al. 2005). However, endocytosis, the G1/S transition of the cell cycle, and the crosstalk with the G protein coupled receptor signaling cascade were not considered as our focus was on early signaling events induced by external stimuli. In addition to EGFR homodimers, we considered signaling through all different ERBB dimers, which is not part of the original stoichiometric model—though a simplified diagram is given in Oda et al. (2005). Finally, there are some reactions and species (e.g., the mammalian target of rapamycin (mTOR), p70S6 kinase) that are only included in our model so as to use the experimental data set (see Section 3.3.2). All in all, the interaction graph model contains 104 species and 266 interactions. A list of all species and interactions, including explanations on differences between the considered components and interactions in the stoichiometric and the interaction graph model, is given in Appendix C.

Before we come to the data analysis, we present some topological properties of the interaction graph model. The model has 236 feedback loops, thereof 139 negative. Strikingly, all positive feedback loops are composed of a negative feed-forward and a negative feedback, except one that describes the reciprocal activation of GAB1, an adapter protein, and PIP_3 , a lipid of the membrane layer (Rodrigues et al. 2000). All negative feedback loops arise from five mechanisms: (i) the kinases ERK1/2 and p90RSK downregulate their own activation by phosphorylation of SOS1, a guanine nucleotide exchange factor (GEF) for Ras; (ii) the phosphatase SHP1 binds to the autophosphorylated ERBB1 homodimers and dephosphorylates them; (iii) Ras positively influences its GTPase activating protein RasGAP via PI3K; (iv) the ubiquitin ligase c-Cbl binds to ERBB1, what induces degradation of the receptor in the lysosome; (v) Ras potentiates the Rab5A-GEF activity of RIN1, thus increasing the formation of endocytic vesicles. Thus, removing the two species Ras and ERBB1 homodimer breaks all negative feedback loops.

Remarkably, the small GTPase Ras is included in 98% of the loops, underlining its central role in the regulation of this network: Ras is a key regulator of cell fate (Downward 1998b) and a known oncogene in many human cancers (Bos 1989).

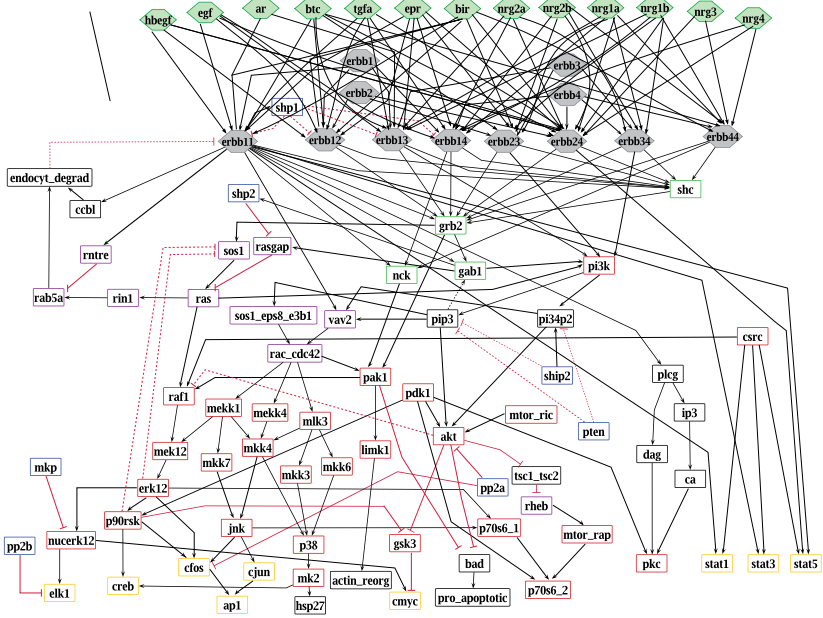


Figure 3.2: Interaction Graph Model of ERBB Signaling. Gray hexagons represent receptor monomers, homodimers, and heterodimers, and green hexagons stand for the 13 different ligands. The remaining species (symbolized with rectangles) are colored according to their function: red: kinases; blue: phosphatases; yellow: transcription factors; green: adapter molecules; violet: small G proteins and GAPs/GEFs; black: other. Black arrows indicate positive edges, red blunt-ended lines negative edges. Dotted lines indicate feedback mechanisms and downregulating feedforward loops that were removed prior to some analyses (for further explanations see text). Figure adapted from Samaga et al. (2009).

However, the high number of feedbacks containing Ras in our model can also reflect the fact that Ras is one of the best studied proteins, and, therefore, the feedback mechanisms of Ras are possibly better known than those of other proteins.

The large size of the network gives rise to a high number of possible signaling paths along which one node may affect another one. For example, there are 6786 paths, thereof 52% negative, leading from the input EGF to the transcription factor AP1 in the output layer.

3.3.2 Analysis of High-Throughput Experimental Data

One of the strengths of our model lies in the broad range of pathways it covers and in the easy simulation of the network wide response to different stimulations and interventions. It is therefore well-suited to analyze high-throughput data, where various readouts are measured in response to several stimuli and to perturbations all over the network. In Section 3.3.2.1, we describe two such data sets, one from primary human heptocytes and one from the hepatocarcinoma cell line HepG2. Both data sets were collected by Leonidas Alexopoulos and coworkers in the group of Peter Sorger, Harvard Medical School. In Section 3.3.2.2, we discuss the analysis of the two data sets based on the dependency matrix. For the primary human hepatocytes data, we also show the analysis based on sign consistency (Section 3.3.2.3).

3.3.2.1 Data Description and Discretization

The first set of measurements we consider is a subset of the phosphoproteomic data set published in Alexopoulos et al. (2010), which was created using a high-throughput method of bead-based fluorescent readings (Luminex, Austin, TX). Primary human hepatocytes and HepG2 cells were treated with/without $TGF\alpha$ in combination with specific molecular inhibitors, whereof six inhibit the activation of nodes considered in our model (MEK1/2, p38, PI3K, mTOR_raptor, GSK3, and JNK). The phosphorylation state of different signaling proteins was then measured after 0, 30, and 180 minutes. Eleven of the measured proteins are included in the ERBB model: AKT, CREB, ERK1/2, GSK3, HSP27, JNK, MEK1/2, p38, p70S6K_1, p90RSK, and STAT3. The data set is shown in Figure 3.3.

For the second data set, HepG2 cells were stimulated with different combinations of five ligands of the EGF family (EGF, NRG1, AR, EPR, and $TGF\alpha$) and treated with the PI3K inhibitor *ZSTK-474*. Again, the phosphorylation state of the eleven signaling proteins mentioned above was measured after 0, 30, and 180 minutes in

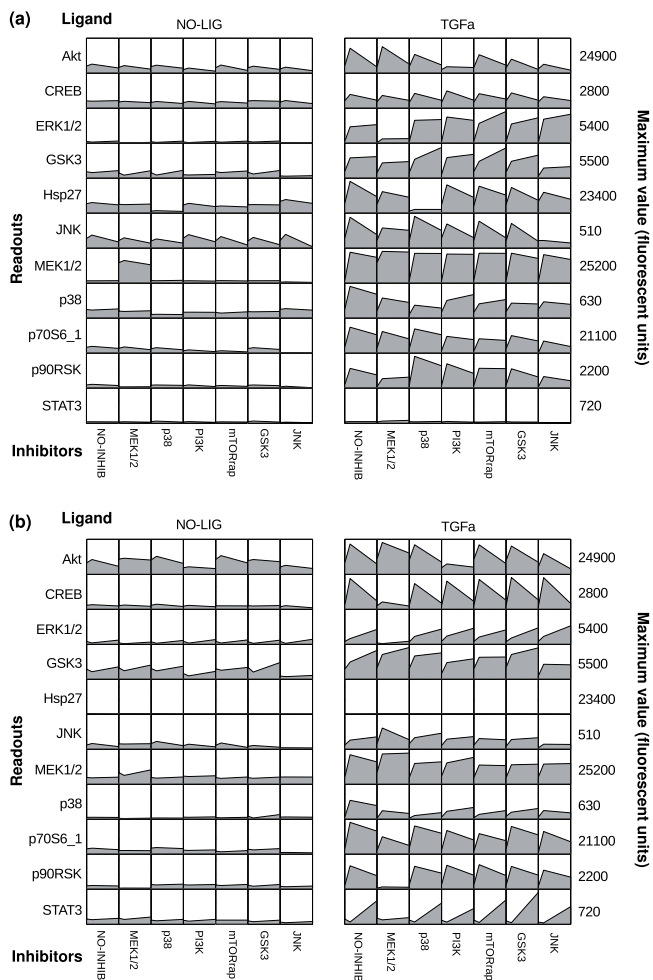


Figure 3.3: Data set of TGF α -stimulated primary human hepatocytes and HepG2 cells. In the rows, measurements of eleven proteins of the ERBB pathway are shown at 0, 30, and 180 minutes after potential ligand addition. The phosphorylation state of each protein has been measured after adding one of seven inhibitors, represented by the columns, in combination without (left panel) and with (right panel) TGF α stimulation. This data set is a subset of the phosphoproteomic data set published in Alexopoulos et al. (2010). It was processed using DataRail software (Saez-Rodriguez et al. 2008). **(a)** Primary human hepatocytes. **(b)** HepG2 cells. Figure adapted from Samaga et al. (2009).

fluorescence units using the same high-throughput method as for the first data set. The complete data set can be seen in Figure 3.4.

The dynamic range of the measured signals depends on the antibodies used for detection. In the first data set, for example, the signal for JNK ranges from 100 to 500 units, whereas MEK1/2 ranges up to 25000 units (Figure 3.3). Variations such as these do not necessarily reflect that JNK is less activated than MEK1/2, but may be attributed to protein abundance or assay calibration issues. Furthermore, the proposed interaction graph methods require a qualitative view of signal transduction, that is, the raw data need to be discretized. As already stated above, we consider three discrete states, which indicate the variation of the activation state of signaling nodes when changing external inputs or adding inhibitors (-1 for decrease, 0 for unchanged, and 1 for increase). We evaluated the ratios of all experiments that differ only by a single perturbation (ligand or inhibitor treatment), and the respective measurement was considered to be (i) increased, if the fold-increase of the signal (with versus without perturbation) was above 1.5, (ii) decreased, if the fold-decrease of the signal (with versus without perturbation) was below 0.66, and (iii) unchanged otherwise.

3.3.2.2 Analysis with the Dependency Matrix

The above described method that uses the dependency matrix for predicting the network response to perturbation experiments (Section 3.1) can, in general, be applied to any interaction graph model. However, in highly interlinked networks the number of ambivalent dependencies in the dependency matrix is usually quite high, what results in a poor predictive power of the method. To compare the ERBB network structure with the perturbation experiments from primary hepatocytes and HepG2 cells, we removed eight interactions from the network structure (see Figure 3.2). These interactions describe feedback mechanisms or downregulating feed-forward loops. Thus, we concentrate on the main activation routes rather than considering the fine-tuning and downregulation of the signaling pathway. This is appropriate for analyzing the 30 minutes time point of the given data sets, which represents a time point where the largest changes in protein modification after perturbation could be observed (Alexopoulos et al. 2010). If one also wants to analyze the late network response (e.g., the 180 minutes time point also included in the data set), the downregulating mechanisms can, of course, not be neglected. We will show in Chapter 5 an application, where predictions from the dependency matrix are com-

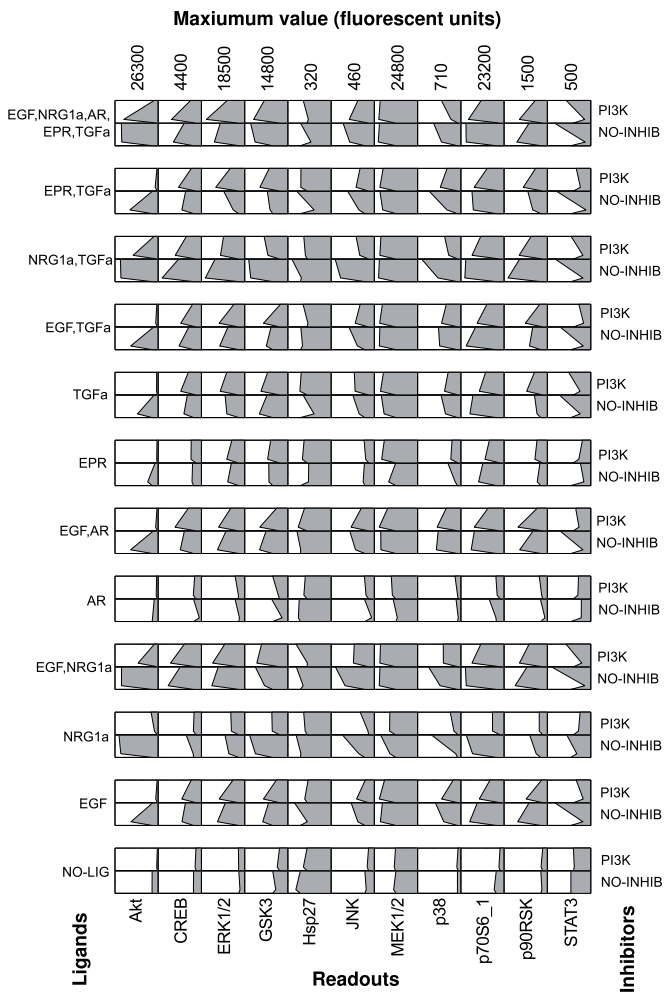


Figure 3.4: Data set from HepG2 cells, stimulated with five different EGF-like ligands. View rotated 90 degrees. In the rows, measurements of eleven proteins of the ERBB pathway are shown at 0, 30, and 180 minutes after potential ligand addition. The phosphorylation state of each protein has been measured after stimulation with different combinations of EGF-like ligands, represented by the paired columns, without (left of each pair of columns) and with (right of each pair of columns) addition of PI3K inhibitor. The data were processed using DataRail software (Saez-Rodriguez et al. 2008). Figure adapted from Samaga et al. (2009).

bined with transient measurements; there, the focus is in particular on the feedback regulation and side activation routes, showing that the method is not generally restricted to acyclic networks, but that additional assumptions are dependent on the data available and the question one wants to answer.

The dependency matrix that results for the ERBB network after removing the eight interactions is shown in Figure 3.5. It contains only a few ambivalent dependencies. Figures 3.6 and 3.7 show the comparison of the experimental data with the model predictions derived from this dependency matrix. All in all, the experimental network response to the different treatments agreed reasonably well with the structure of the model, in particular in primary cells. In HepG2 cells, 9.7% of the analyzed dependencies were contradictory to our model, whereas 44.6% agreed explicitly. In the remaining 45.7% of the cases (gray entries in Figure 3.7), the data showed no significant change, although the perturbed node can affect the readout in our model. Many of these gray entries will be discussed below. In primary cells, 13.2% of the predictions were false, 73.6% were fully correct, and for 13.2% we observed no significant changes, although the model contains paths between the perturbed node and the readout.

Most disagreements between model predictions and experimental results concentrated on certain experimental scenarios (rows) and readouts (columns). In the following, we discuss such systematic inconsistencies, and, using our model, we seek to provide explanations and conclusions:

- A significantly increased state of phosphorylation of STAT3 in response to any of the ligands could not be found both in HepG2 and primary hepatocytes. Whether this is due to the fact that the activation of STAT3 is very transient, as it has been reported, for example, for the human epithelial carcinoma cell line A431 (Olayioye et al. 1999), or if the activation of this transcription factor through ERBB receptors plays no role in hepatocytes, still has to be clarified.
- Stimulation of HepG2 cells with amphiregulin (not measured in primary cells) did not result in activation of the measured proteins (see Figure 3.7, lines 34–37). This is in agreement with findings of amphiregulin being a much weaker growth stimulator than EGF in some cell types (Shoyab et al. 1989).
- The systematic errors in the column of p38—for primary as well as HepG2 cells—might indicate missing edges in the model requiring further experimental studies to verify these findings. We cannot exclude that other, for example,

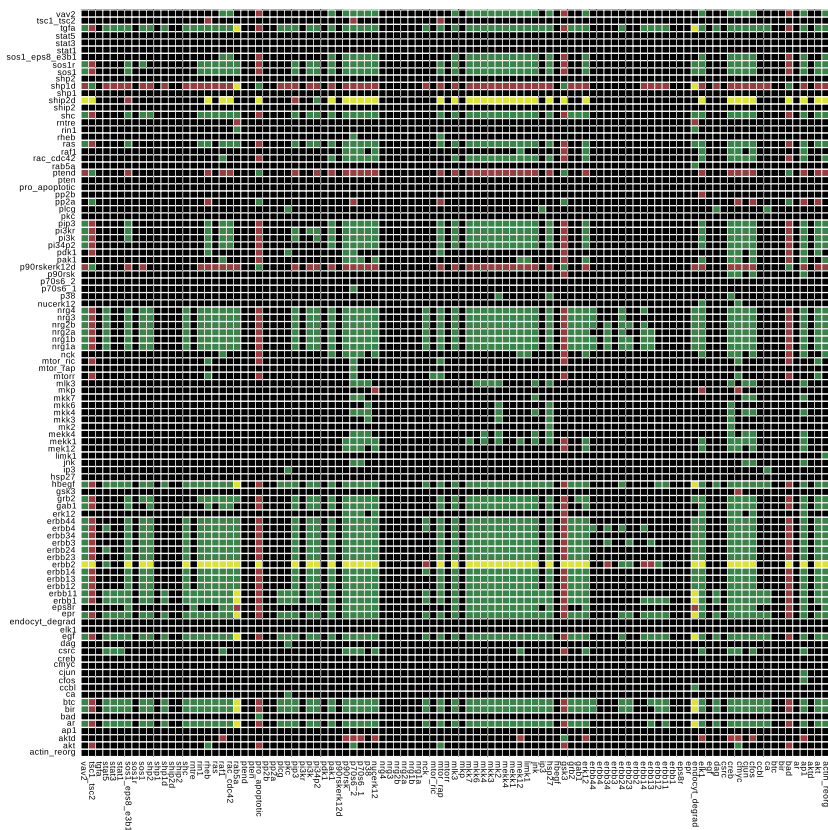


Figure 3.5: Dependency matrix of ERBB interaction graph model, main activation routes. Shown is the dependency matrix of the ERBB interaction graph model given in Figure 3.2. Only the main activation routes are considered, that is, eight interactions (indicated in Figure 3.2) were removed prior to computing the dependency matrix. The color of matrix element D_{ij} means the following: green: species i is a strong activator of species j ; red: i is a strong inhibitor of j ; yellow: i is an ambivalent factor for j ; black: i has no influence on j . Figure adapted from Samaga et al. (2009).

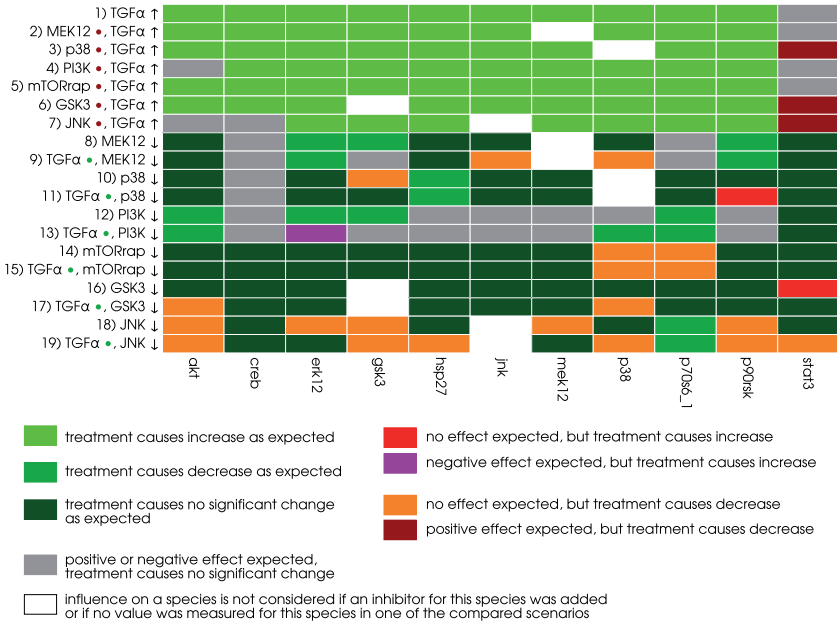


Figure 3.6: Comparison between experimental data from primary hepatocytes and ERBB model predictions derived from the dependency matrix. Shown is the comparison between the measured and predicted changes (ups and downs) in the activation levels of network elements in response to ligands and inhibitors in primary human hepatocytes (data obtained from Alexopoulos et al. (2010)). Each row compares two different scenarios A and B. A filled circle behind the species name in the row labels indicates that, in both scenario A and scenario B, this species was added as ligand (green filled circle), or an inhibitor for this species was added (red filled circle). Species whose input values differ in both scenarios are marked with an arrow pointing up or down, respectively. For example, the comparison of scenario A (EGF ligand, TGF α ligand, PI3K inhibitor) and scenario B (TGF α ligand, PI3K inhibitor) is labeled by TGF α \bullet , PI3K \bullet , EGF \uparrow , that is, the influence of an increased level of EGF on the readouts is analyzed under the side constraints that TGF α and a PI3K inhibitor were added as well. The readouts are shown in the columns. The color indicates whether the model predictions derived from the dependency matrix (Figure 3.5) and the measurements are consistent or not (see color legend). Figure reprinted from Samaga et al. (2009).

3 Interaction Graphs: Data-Driven Interrogation and Training of Network Structures

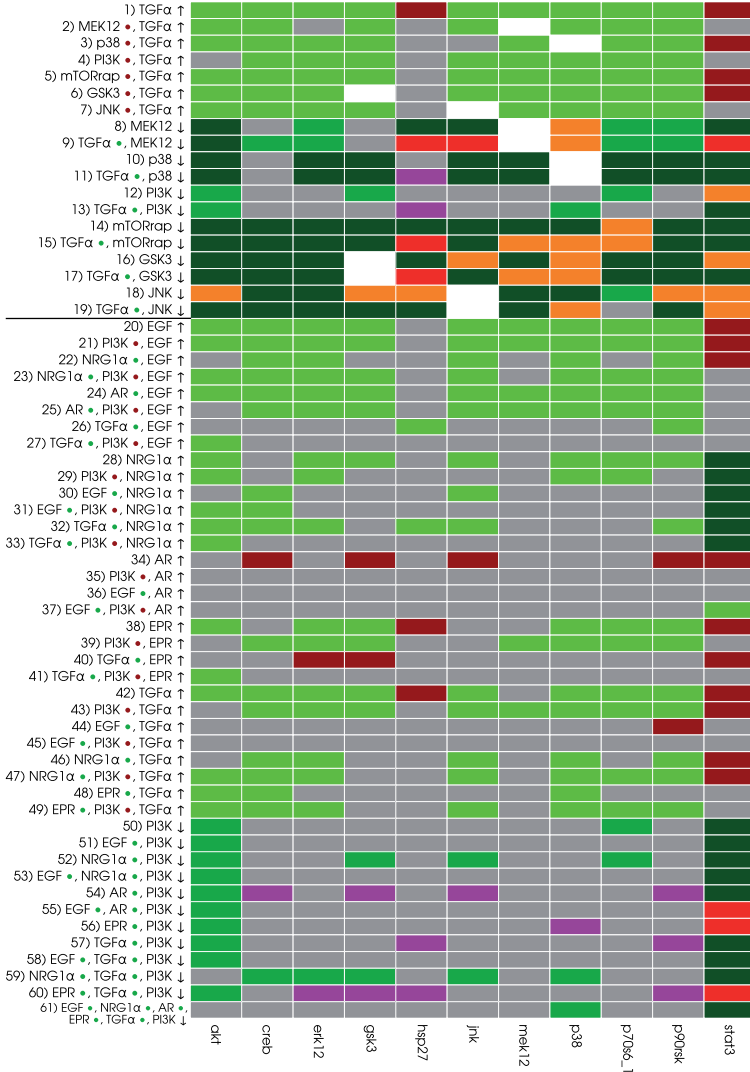


Figure 3.7: Comparison between experimental data from HepG2 cells and ERBB model predictions derived from the dependency matrix. Shown is the comparison between the measured and predicted changes (ups and downs) in the activation levels of network elements in response to ligands and inhibitors in HepG2 cells (data sets given in Figures 3.3 and 3.4). The horizontal line separates the first (top) from the second (bottom) data set for HepG2 cells. For further explanations and color legend see Figure 3.6. Figure reprinted from Samaga et al. (2009).

stress-induced, pathways not captured in our model may have caused these observations, also because some of the effects on p38 are also present without ligand stimulation.

- Stimulating HepG2 cells with both TGF α and EGF did not result in a significantly higher activation level of the readouts compared to adding only one of these ligands, as can be seen from the predominantly gray entries in lines 26/27 and 44/45 in Figure 3.7. This finding is in accordance with the fact that both ligands are very similar and bind to the same receptor dimers (see Appendix C.2).
- One of the major differences in the behavior of the two cell types was the activation of HSP27: whereas this heat shock protein was activated in response to cytokine stimulation in primary cells, no significant increase in the state of phosphorylation was observed in almost all studied scenarios in the cancer cell line.
- Another remarkable discrepancy between the experimental data and our model predictions is the influence of the mTOR inhibitor rapamycin on phosphorylation of the autoinhibitory domain of p70S6 kinase (termed p70s6k_1 in the model), which is not supported by our model (see lines 14/15 in Figures 3.6 and 3.7). Although mTOR mediates the phosphorylation of the catalytic site T389 (Hou et al. 2007), it has to the best of our knowledge not been implicated with the phosphorylation of T421 and S424, those sites, whose state of phosphorylation were measured in the analyzed data sets. However, an inhibitory effect of rapamycin on these sites has been reported earlier (Weng et al. 1998), even if the molecular mechanism that could explain this influence still has to be uncovered.
- According to our model, PI3K should influence all measured readouts except STAT3. However, the data showed a clear effect of the PI3K inhibitor only on the phosphorylation of AKT (see Figure 3.6, lines 12/13, and Figure 3.7, lines 50–61). Therefore, we searched for hypothetical changes in our model structure that could explain these experimental findings. We observed that node Rac/CDC42 lies on all paths connecting the ligands with the aforementioned critical readouts (except GSK3, see below). We may thus hypothesize that—in contrast to the assumption in our model—PI3K does not influence the small G-proteins Rac and CDC42 in primary hepatocytes and in HepG2 cells.

- In the model, the inhibitory phosphorylation of GSK3 can be induced by a MEK1/2-dependent pathway (via p90RSK) and by a PI3K-dependent pathway (via AKT). Figures 3.6 and 3.7 (lines 9 and 13) show that the phosphorylation of GSK3 in response to TGF α is independent of the MEK inhibitor and the PI3K inhibitor, both in HepG2 and in primary cells. As TGF α stimulation leads to a strong phosphorylation of GSK3 in both cell types, there must be another signaling route, not involving MEK and PI3K. One possible candidate is PKC which has already been reported to inhibit GSK3, however, not in response to ligands of the EGF family (Grimes and Jope 2001).
- According to the data, both GSK3 and p90RSK are influenced by JNK inhibitor after TGF α stimulation in primary hepatocytes (see Figure 3.6, line 18). This seems to support another possible mechanism, where JNK activates p90RSK, which may then phosphorylate GSK3. However, the JNK inhibitor affected much more proteins than expected, both in HepG2 and in primary cells. As these unexpected influences also occurred in absence of ligand stimulation, this strongly suggested a minor specificity of the JNK inhibitor. This is in accordance with observed off-target effects of the JNK inhibitor in Alexopoulos et al. (2010).
- Similar as for GSK3 phosphorylation, data analysis with our model provided useful insights into the activation mechanism of CREB in response to TGF α : the proposed effect of the p38-dependent kinase MK2 on CREB could not be observed both in HepG2 and in primary cells (see Figures 3.6 and 3.7, line 11). The positive effect of MEK on CREB phosphorylation after TGF α stimulation could be seen in HepG2 (Figure 3.7, line 9), but not in primary hepatocytes (Figure 3.6). This indicated that there must be an alternative pathway for CREB activation in primary hepatocytes.

A summary of the above mentioned results is given in Table 3.2. Changing the model accordingly, we could increase the number of comparisons that agreed explicitly from 73.6% to 83.2% for primary hepatocytes, and from 44.6% to 65.7% for HepG2 cells. Moreover, the number of entries where we assumed a change in the data but could not detect a significant increase or decrease was reduced from 13.2% to 2.5% (primary hepatocytes), and from 45.7% to 22.2% (HepG2 cells), albeit at the expense of a minor increase in the number of contradictions (primary hepatocytes: increase from 13.2% to 14.2%; HepG2: increase from 9.7% to 12.1%).

Table 3.2: Proposed ERBB model changes to improve agreement between interaction graph model structure and experimental data.

	Primary human hepatocytes	HepG2 Cells
Removed interactions		
erbb11 → stat3	x	x
pip3 → vav2	x	x
pi34p2 → vav2	x	x
pip3 → sos1_eps8_e3b1	x	x
p90rsk ⊣ gsk3	x	x
akt ⊣ gsk3	x	x
mk2 → hsp27		x
mk2 → creb	x	x
p90rsk → creb	x	
Added interactions		
mtor_rap → p70s6k_1	x	x
pkc ⊣ gsk3	x	x
tgfa → creb	x	x

3.3.2.3 Analysis Based on Sign Consistency

In the previous section, we checked for data from primary human hepatocytes and from HepG2 cells whether the measured responses to certain ligands and inhibitors were consistent with the causal dependencies in the ERBB network topology. Resulting from this analysis, we proposed for both cell types changes in the network structure to improve the agreement between experimental data and model predictions. These changes were derived solely by inspection. Here, we apply the ILP approach introduced in Section 3.2 to compare experimental data with the ERBB network structure. This can be seen as a step forward, as the ILP approach adapts the model structure to the experimental data in an automatic way and searches for all possible solutions that resolve discrepancies between model and data. We restricted ourselves here to analyze the data from primary human hepatocytes (Figure 3.3(a)). As the JNK inhibitor data from the data set showed off-target effects (see Section 3.3.2.2 and Alexopoulos et al. (2010)), we decided to exclude this data.

Data and Model Preprocessing. The data analysis using the concept of sign consistency is based on the assumption that the system moves from one steady state to another upon imposing the perturbations (Siegel et al. 2006). However, this does not necessarily mean that we have to wait until the system has reached its new steady state completely; instead, we can take the measurements if we can assume

that the *sign* of the state variations will not change anymore. Thus, it is important to determine a suitable time point where all relevant state changes induced by the perturbation have become visible in the measurements. For example, if measurements are taken too early, a signal has possibly not yet been propagated to all downstream nodes resulting in inconsistencies with the predictions made from the interaction graph. For our data set under study, the 30 minutes time point has been chosen in such a way that the perturbation-induced changes in the phosphorylation level of the proteins are well-reflected by the measurements (Alexopoulos et al. 2010). Data discretization was performed as described in Section 3.3.2.1.

We used the same network as for the analysis in the previous section, that is, we removed eight interactions from the ERBB interaction graph that represent feedback mechanisms or downregulating feed-forward loops (indicated in Figure 3.2). In the following, we show the application of the SCEN_FIT, MCoS, OPT_SUBGRAPH and OPT_GRAPH algorithms. To be able to make meaningful conclusions, we sought to find, for each of the problems, all optimal solutions. However, enumerating all solutions for OPT_SUBGRAPH and OPT_GRAPH in the full model structure becomes quickly intractable as the highly branched network structure (e.g., various feedforward routes running over different combinations of ERBB dimers and adapter proteins connect TGF α with PI3K) leads to an immense number of different optimal solutions. Thus, we used a loss-free network compression technique by which compressed solutions were computed from a smaller network. Other advantages of network compression are that differences between the original and the compressed network structure may indicate non-identifiabilities in the original network and that obtained optimal solutions can be represented in a condensed manner, not explicitly displaying all combinatorial solutions existing due to non-uniqueness. By keeping track of the made compression steps it is, in principle, possible to decompress solutions found in the compressed network. However, it is often useful to discuss instead the found solutions directly for the compressed network, thereby avoiding the interpretation of a typically much larger number of decompressed solutions arising due to non-uniqueness. For example, instead of listing all possible parallel pathway combinations connecting A with B, one might conclude that “at least one pathway between A and B must exist”, what can easier be represented in a compressed network. We used four simple compression rules, which are given in Appendix B. For the SCEN_FIT and MCoS analysis, we only removed non-observable and non-controllable nodes according to Rule 1 of the model compression (Appendix B); the resulting graph is shown in Figure 3.8(a). The full compression was applied before

searching for optimal subgraphs and graphs. As can be seen in Figure 3.8(b), the model structure could be compressed substantially from 39 nodes and 67 edges to 14 nodes and 18 edges. Strikingly, *Rac_CDC42* remained as the only latent node in the compressed structure. The compressed interaction graph reflects the essential dependencies in the original network structure that can be addressed by the given set of perturbed/measured nodes. For example, parallel signaling paths leading from a perturbed node to a measured node without passing any other measured/perturbed node cannot be distinguished in the analysis performed herein and were therefore condensed to one single edge in the compressed graph.

Note that in Figure 3.8, compared to the original network (Figure 3.2), the interactions influencing GSK3 are activating. This is due to the fact that the measured phosphorylation of GSK3 is inhibitory; in order to directly map the measurements, we decided here to interpret the interactions as “positively influencing the phosphorylation” instead of “inhibiting the activity” as in the original network.

Applying SCEN_FIT and Minimal Correction Sets. In Figure 3.9, the discretized measurements and, for each scenario, the corresponding SCEN_FIT solution are depicted. Recall that the SCEN_FIT algorithm determines, for a given scenario, a sign consistent node labeling that is closest to the measurements and can thus best explain how the ERBB network topology in Figure 3.8(a) induces the measured node changes for the respective scenario. Deviations between the determined optimal sign pattern and the measured state changes (as indicated in Figure 3.9) uncover inconsistencies between network structure and observed behavior. For example, scenario 1 reflects the influence of the ligand $TGF\alpha$, that is, $TGF\alpha$ is the perturbed node and its state is fixed to 1. As depicted in Figure 3.9, the SCEN_FIT solution for this scenario shows a fitting error of 1: in the optimal sign consistent node labeling, all measured nodes have sign 1 as they are connected to $TGF\alpha$ by positive paths only. This is in accordance with the measured state of all nodes except STAT3: the latter shows no significant change in response to $TGF\alpha$ inducing thus a fitting error. Scenarios 2–6 reflect the influence of $TGF\alpha$ in presence of different inhibitors. We assume that an inhibitor completely blocks the signal flow through the inhibited species; thus, we define these scenarios by fixing the state of $TGF\alpha$ to 1 and of the inhibited node to 0. The remaining scenarios reflect the influence of the inhibitors in presence (scenarios 8, 10, 12, 14, and 16) and absence (scenarios 7, 9, 11, 13, and 15) of $TGF\alpha$. In each of these scenarios, the perturbed node is the respective inhibitor and its state is fixed to -1 . Importantly, by using the enumeration algorithm for

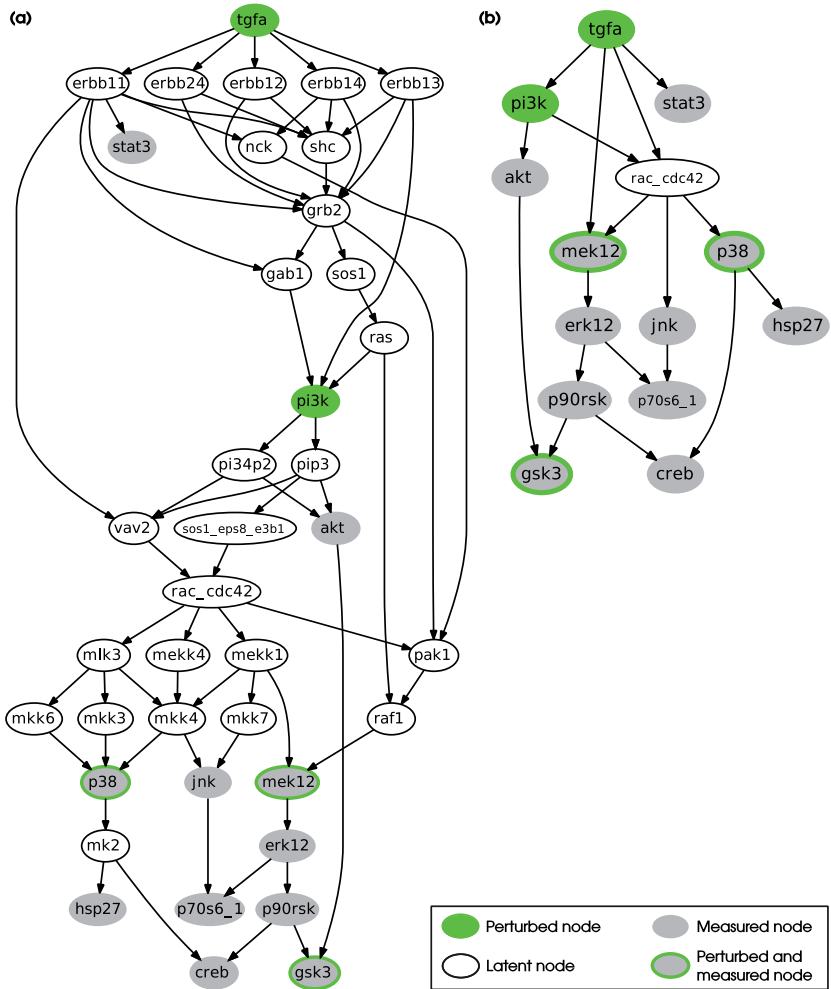


Figure 3.8: Interaction graph model of the ERBB signaling network, processed for ILP approach. (a) The full network as shown in Figure 3.2 after removal of eight interactions (see text) and of non-observable and non-controllable nodes. All edges are activating edges having positive signs. (b) The compressed model obtained after applying the compression rules given in Appendix B to the model shown in (a). Figure reprinted from Melas, Samaga et al. (2013).

SCEN_FIT we could prove that, for each scenario, the found solution for the optimal fit is unique, hence, no other optimal solutions need to be considered.

We also assessed the sensitivity of the SCEN_FIT results with respect to the chosen thresholds $t_{up} = 1.5$ and $t_{down} = 0.66$ for data discretization (significant increase: fold-increase $> t_{up}$, significant decrease: fold-decrease $< t_{down}$). Therefore, we screened 80 different values for t_{up} (from 1.025 to 3.025) and 40 values for t_{down} (from 0.1 to 1.0) and plotted the resulting cumulative fitting error of the SCEN_FIT problems from all 16 scenarios (Figure 3.10). The larger t_{up} and the smaller t_{down} became, the more over-responsive the discretized data set was. We observed that there was a relatively large range ($0.2 \leq t_{down} \leq 0.7$, $1.5 \leq t_{up} \leq 1.9$) where the total fitting error showed its lowest value (between 40 and 50), and which included the chosen thresholds. Outside that area, the fitting error increased rapidly.

The SCEN_FIT results shown in Figure 3.9 are in high accordance with the results derived with the dependency matrix (Figure 3.6). Please note that due to the removal of JNK inhibitor data for the SCEN_FIT analysis, there are three scenarios in Figure 3.6 which are not included in Figure 3.9.

Figure 3.9 shows that there are several inconsistencies between experimental data and the SCEN_FIT solutions derived from the initial network topology (fitting error over all scenarios is 43). In order to understand where these inconsistencies are induced in the network, we address the identification of minimal correction sets (MCoS). We recall that MCoS are minimal sets of artificially enforced changes of node states (e.g., from increased to decreased), which make an inconsistent scenario consistent. Exemplarily, we focus on scenario 11 of Figure 3.9 (where PI3K inhibitor is added without presence of TGF α), whose SCEN_FIT solution produced a total error value of 6.

As shown in Table 3.3, five MCoS were identified, each containing three corrections (virtual perturbations) rendering the experimental scenario 11 sign consistent. Common trend in all MCoS is to remove the downregulating effect of PI3K on signals downstream of Rac_CDC42 by setting Rac_CDC42 to unchanged (0) or one of the nodes SOS1_EPS8_E3B1, VAV2, PtdIns(3,4)P₂ or PIP₃ to increased (1). Introducing this change, the states of p38, JNK, MEK1/2, HSP27, CREB, and p90RSK were in accordance with the measurements (i.e., they show now response upon adding PI3K inhibitor). However, by this modification, the states of ERK1/2 and p70S6K_1 would change their predicted level from “decreased” to “unchanged”, which is not in agreement with the measured state. This was corrected in all MCoS by setting ERK1/2 to -1 . Again, this correction implied an undesired effect, namely chang-

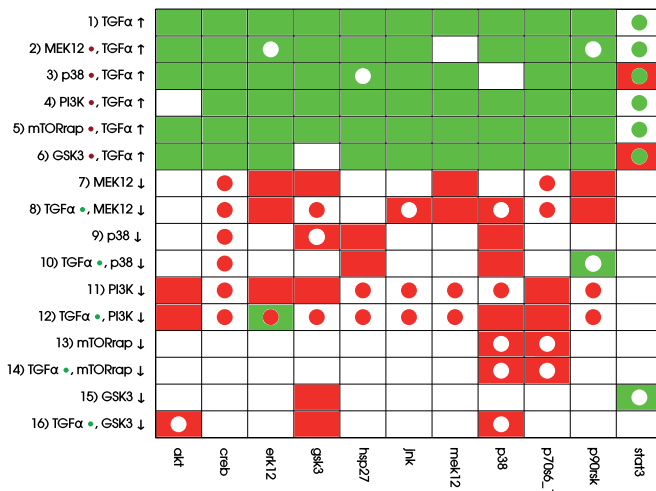


Figure 3.9: Discretized measurements of the 16 considered experimental scenarios and the resulting SCEN_FIT solutions computed from the ERBB interaction graph model. Each row corresponds to one experimental scenario, each column contains the measured state changes of the readout species. An arrow behind a species name in the row indicates that the perturbation (increase or decrease according to arrow pointing up or down) of this species is considered. Species names followed by a red or green filled circle in the rows indicate that this species was added as ligand (green) or that an inhibitor for this species was added (red). For example, scenario 2 considers an increase of TGFα in presence of an inhibitor for MEK1/2. The discretized measurements are mapped to the fill color of the respective fields: if a node is increased in the respective scenario, the corresponding field is filled green, if it is decreased, the field is filled red, and if it shows no significant change, it is filled white. Accordingly, the color of the added circle shows the sign of the node in the closest sign consistent node labeling derived by SCEN_FIT: green circles correspond to sign 1, red circles to sign -1, and white circles to sign 0. Note that circles only appear if the measurement is not in accordance with the respective state in the sign consistent labeling. Figure adapted from Melas, Samaga et al. (2013).

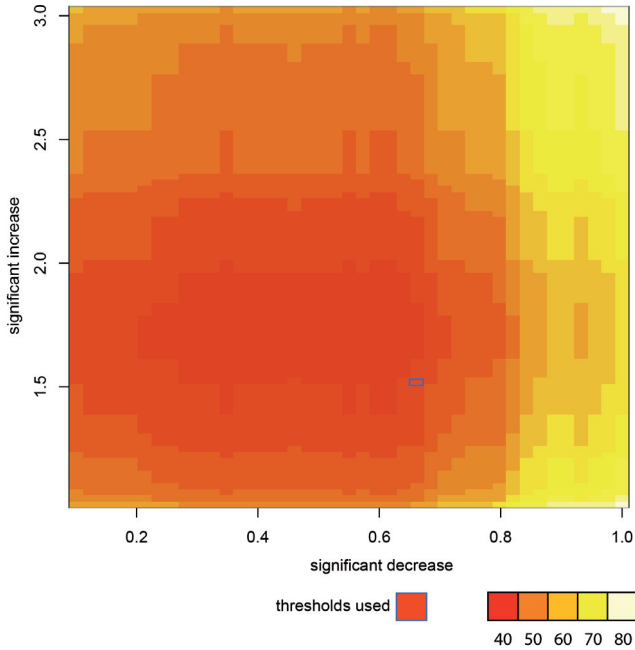


Figure 3.10: Cumulative fitting error of optimal SCEN_FIT solutions over all 16 scenarios in the ERBB network as a function of the two discretization thresholds. The figure shows the cumulative fitting error of optimal SCEN_FIT solutions over all 16 scenarios in the ERBB network (Figure 3.8) as a function of the two thresholds for significant increase and significant decrease. The blue rectangle indicates the combination of thresholds used for all analyses in this section. Figure reprinted from Melas, Samaga et al. (2013).

Table 3.3: MCoS for scenario 11 in Figure 3.9.

Node id	MCoS 1			MCoS 2			MCoS 3			MCoS 4			MCoS 5		
	B_i^+	B_i^-	Val	B_i^+	B_i^-	Val	B_i^+	B_i^-	Val	B_i^+	B_i^-	Val	B_i^+	B_i^-	Val
rac_cdc42	1		0												
p90rsk	1		0	1		0	1		0	1		0	1		0
erk12		1	-1		1	-1		1	-1		1	-1		1	-1
sos1_eps8_e3b1				1		1									
vav2							1		1						
pi34p2									1		1				
pip3													1		1

Five MCoS were identified for the ERBB network model (Figure 3.8(a)) with respect to scenario 11 in Figure 3.9. Each MCoS would lead to a perfect fit for this scenario, and all five MCoS contain three nodes to be enforced to a certain value. Nodes p90rsk and erk12 are common in all MCoS. In addition, one of the nodes rac_cdc42, sos1_eps8_e3b1, vav2, pi34p2 and pip3 is perturbed in each MCoS. In columns MCoS 1-5, three sub-columns are shown: sub-column Val shows the corrected state of the node, the entry 1 in sub-column B_i^+ indicates that a positive input edge is added to the node in order to alter its state, and the entry 1 in sub-column B_i^- indicates that a negative input edge is added to the node. Table adapted from Melas, Samaga et al. (2013).

ing p90RSK from 0 to -1 , which is countered by assigning p90RSK the value 0 in all MCoS. Clearly, three required corrections indicate that the observed behavior for this scenario was not well-reflected by the network topology. It would therefore be useful to consider all scenarios at the same time to detect common points of errors produced in all or many scenarios.

Applying OPT_SUBGRAPH. We used the OPT_SUBGRAPH algorithm to find—by appropriate edge removals—optimal subgraphs of the compressed ERBB network structure (Figure 3.8(b)), each minimizing the fitting error over all experimental scenarios.

The computation of all optimal subgraphs of the compressed network resulted in six solutions having the same minimum fitting error of 26 which has thus reduced much in comparison to 43 in the original model. Figure 3.11(a) shows a combined view of the six optimal solutions; the single solutions are shown in Table 3.4. In more detail, a positive influence of TGF α on STAT3 is not reflected in the measurements (see Figure 3.9); consequently, the edge TGF $\alpha \rightarrow$ STAT3 is removed in all optimal solutions. Another edge that is removed in all solutions is PI3K \rightarrow Rac_CDC42, as a number of signals downstream of Rac_CDC42 did not show the expected decreased response to the PI3K inhibitor in the measurements (this is consistent with the re-

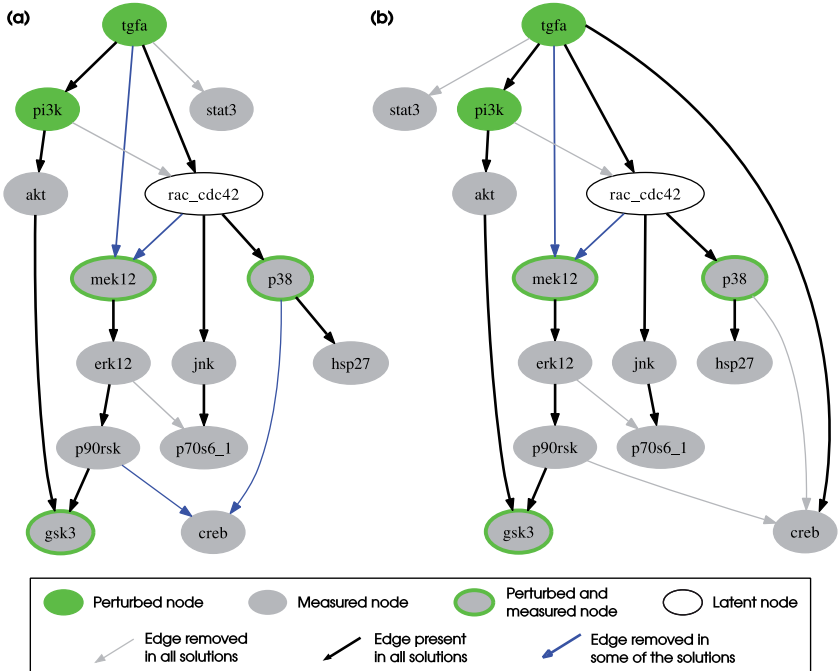


Figure 3.11: Optimal model structures derived from the compressed ERBB model by applying OPT_SUBGRAPH and OPT_GRAPH procedures. (a) Combined view of the six optimal model structures derived from the compressed ERBB model (Figure 3.8(b)) by applying the OPT_SUBGRAPH procedure with enumeration. **(b)** Combined view of the three optimal sub-graphs resulting when adding $tgfa \rightarrow creb$ to the initial model structure. Figure adapted from Melas, Samaga et al. (2013).

Table 3.4: Optimal subgraphs that fit the discretized data shown in Figure 3.9.

Removed edges	Solution 1	Solution 2	Solution 3	Solution 4	Solution 5	Solution 6
pi3k → rac_cdc42	1	1	1	1	1	1
rac_cdc42 → mek12	1		1			
rac_cdc42 → p38						
tgfa → rac_cdc42						
tgfa → pi3k						
tgfa → mek12				1	1	
pi3k → akt						
p90rsk → creb	1	1		1		
p90rsk → gsk3						
akt → gsk3						
rac_cdc42 → jnk						
p38 → hsp27						
p38 → creb			1		1	1
mek12 → erk12						
jnk → p70s6k_1						
erk12 → p70s6k_1	1	1	1	1	1	1
erk12 → p90rsk						
tgfa → stat3	1	1	1	1	1	1

The first column shows the complete list of edges included in the compressed network (Figure 3.8(b)). Columns 2–7 denote whether the respective edge was removed (1) or not (empty cell) by the optimization procedure. Table adapted from Melas, Samaga et al. (2013).

sults of the MCoS discussed in the previous paragraph. Finally, by removing the edge $\text{ERK1/2} \rightarrow \text{p70S6K}_1$ in all solutions, the missing influence of MEK inhibitor on p70S6K_1 is accommodated. The edges $\text{TGF}\alpha \rightarrow \text{MEK1/2}$ and $\text{Rac_CDC42} \rightarrow \text{MEK1/2}$ are only removed in some of the solutions. This is an example for two parallel routes that cannot be distinguished: the model structures containing both routes or either route give rise to the same sign consistent labeling. In contrast, removing either of the edges $\text{p90RSK} \rightarrow \text{CREB}$ and $\text{p38} \rightarrow \text{CREB}$ results in different sign consistent labelings, both showing the same number of discrepancies to the measurements: the phosphorylation state of CREB is neither affected by MEK inhibitor nor by p38 inhibitor. However, removing both edges at the same time would interrupt all routes from $\text{TGF}\alpha$ to CREB what is contradictory to the observed positive effect of $\text{TGF}\alpha$ in scenarios 1–6. Thus, in this case, allowing only the removal of edges is not sufficient to fully explain the observed measurements. This can be seen in Figure 3.12, where the two possible optimal sign consistent labelings that SCEN_FIT would find for the six pruned model structures are shown in comparison to the discretized measurements: in each solution, there are three different remaining errors in the CREB column. The errors for STAT3 as well as the errors in response to PI3K inhibitor (scenarios 11 and 12) could be significantly reduced by removing the respective edges.

Applying OPT_GRAPH. Next, we used the OPT_GRAPH procedure to identify edges that may be missing from the ERBB network and whose addition would therefore improve the goodness of fit to the data. Table 3.5 displays the edges that lead to the highest improvement as determined by OPT_GRAPH. All these edges have in common that they give rise to an additional route from $\text{TGF}\alpha$ to CREB not running over p38 or MEK1/2. By adding any of these edges to the model structure before reapplying the OPT_SUBGRAPH procedure, we could further reduce the fitting error to 23, compared to 26 if only edge removals were allowed.

As an example, we show the optimized model structures when adding the edge $\text{TGF}\alpha \rightarrow \text{CREB}$. A combined view of the three optimal solutions that can be found by OPT_GRAPH after adding this edge is shown in Figure 3.11(b). As it was the case for the optimization in the original network, the edges $\text{TGF}\alpha \rightarrow \text{STAT3}$, $\text{PI3K} \rightarrow \text{Rac_CDC42}$, and $\text{ERK1/2} \rightarrow \text{p70S6K}_1$ were removed in all solutions, while the edges $\text{TGF}\alpha \rightarrow \text{MEK1/2}$ and $\text{Rac_CDC42} \rightarrow \text{MEK1/2}$ are two alternative routes (either both are present or at least one of both; this gives the three optimal subgraphs). With the added edge $\text{TGF}\alpha \rightarrow \text{CREB}$ the model structure contains an activation route

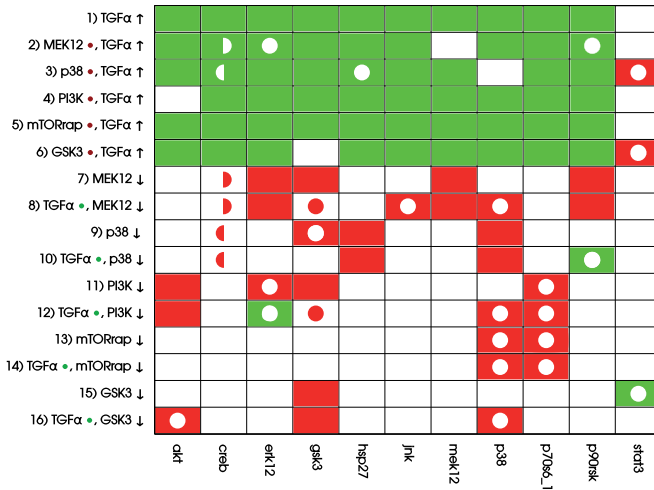


Figure 3.12: Discretized data and the two SCEN_FIT solutions that result from the optimal subgraphs given in Figure 3.11(a). The color coding is the same as in Figure 3.9. All six optimal subgraphs shown in Figure 3.11(a) gave rise to the same SCEN_FIT solution, except for the CREB column. Here, three subgraphs showed a mismatch in scenarios 3, 9, and 10 (indicated by the left semicircles), while the other three showed a mismatch in scenarios 2, 7, and 8 (indicated by the right semicircles). Figure adapted from Melas, Samaga et al. (2013).

Table 3.5: Suggestions for new single edges as computed by OPT_GRAPH.

- tgfa → creb
- jnk → creb
- p70s6k_1 → creb
- rac_cdc42 → creb
- tgfa → erk12
- jnk → erk12
- rac_cdc42 → erk12

Adding any of these edges to the model structure leads to a decrease of the fitting error from 26 to 23. Table adapted from Melas, Samaga et al. (2013).

from TGF α to CREB that is independent of p38 and p90RSK, and removing both the p90RSK \rightarrow CREB and p38 \rightarrow CREB edge in all solutions was now optimal.

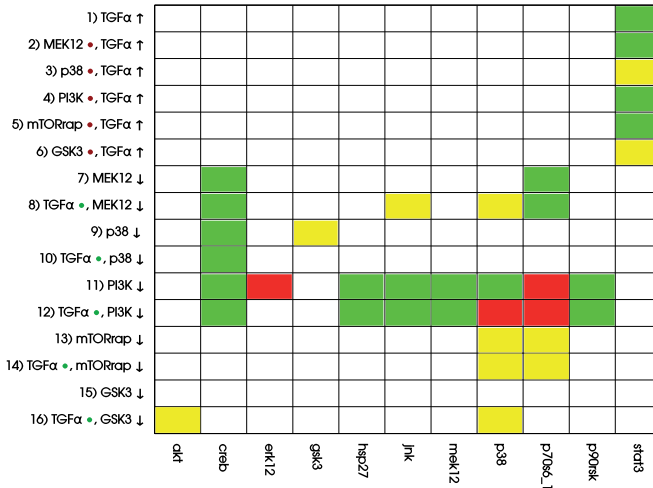


Figure 3.13: Comparison of the fitting errors of the initial model structure (see Figure 3.8) and of the optimal interaction graph shown in Figure 3.11(b). Green fields indicate an error that has been present in the original model structure, but could be removed by optimizing the model structure. Yellow fields refer to errors that could not be resolved, and red fields indicate errors that have not been present in the original model structure, but were introduced by the optimization. Figure adapted from Melas, Samaga et al. (2013).

All three solutions induced the same optimal sign consistent node labeling. Figure 3.13 shows the mismatches of the experimental data in the optimal graph (Figure 3.11(b)) versus the mismatches in the initial model structure (Figure 3.8). The measurements for CREB were now in full accordance with the model structure, and the errors for STAT3 could be significantly reduced. Furthermore, a number of errors in scenarios 11 and 12 showing the influence of PI3K inhibitor could be eliminated, although at the same time a few mismatches for some nodes have been introduced. Finally, the influence of MEK inhibitor on p70S6K_1 was now predicted correctly. Here, we considered only the addition of a single edge to improve the fit to data. In principle, one could remove all remaining discrepancies by adding further edges. However, in particular if the measurements show inconsistencies (e.g., the different effect of PI3K inhibitor on ERK1/2 with/without TGF α), some errors can only be

removed by introducing a positive and a negative edge between a pair of nodes. Furthermore, edges leading only to a minor improvement of the fitting error are unlikely to represent a real effect. We also emphasize that proposed new edges may often indicate *indirect* rather than direct effects, thus representing hidden paths in the network. In any case, dedicated experiments are required to confirm or prove the suggested causal links.

To summarize, essential findings of the new network structure optimization via sign consistency in the ERBB network—which may indicate important specifics of this signaling pathway in primary human hepatocytes—are:

1. STAT3 is not activated by TGF α .
2. Phosphorylation of the autoinhibitory domain of p70S6K (termed p70S6K_1 in the model) is independent of ERK1/2.
3. The activation of CREB in response to TGF α is likely to be caused by a p38 and MEK1/2 independent route.
4. The activation of Rac/CDC42 is independent of PI3K activity.

These results, generated in an automated way, confirm several of the conjectures formulated in Section 3.3.2.2, which were derived by inspection only. In detail, findings 1, 3, and 4 agree explicitly with the proposed changes in network structure derived with the dependency matrix (see Table 3.2). Finding 2 shows the advantage of an automated approach: the proposed change in the model structure would have also improved the agreement between model and data using the dependency matrix, but was simply overlooked. In contrast, some proposed model changes in Table 3.2 were not identified with the ILP approach, as, for this application, we only considered the addition of one edge to the initial model structure. This applies for the added edges mTOR_Rap \rightarrow p70S6K_1 and PKC \rightarrow GSK3, but also for p90RSK \rightarrow GSK3 and AKT \rightarrow GSK3, whose removal only reduces the fitting error if, in return, PKC \rightarrow GSK3 (or another way of TGF α -induced GSK3 phosphorylation) is added. Not only considering single edge additions for OPT_GRAPH might have resulted in the identification of these changes.

3.4 Discussion

In this chapter, we demonstrated how interaction graphs can be applied to detect and remove inconsistencies between measurements and signaling network topolo-

gies. The concept of the dependency matrix (Section 2.4.1.2) can be used in a straightforward way to predict the effect of perturbations in stimulus–response experiments. As an application example, we studied a large-scale model describing signaling through the four ERBB receptors. We carefully constructed this model based on the stoichiometric pathway map of Oda et al. (2005) and intense literature research. The model can be seen as a “master model”, which assembles the current qualitative knowledge of the main and also some side routes of ERBB signaling derived from different cell types. The analysis of high-throughput phosphoproteomic data from primary human hepatocytes and the hepatocarcinoma cell line HepG2 revealed systematic inconsistencies, from which we could derive suggestions for possible cell-type-specific changes of the network structure. Another example where we applied data analysis based on the dependency matrix to interrogate interaction graph network structures is that of Interleukin 1 and 6 signaling in primary human hepatocytes (Ryll et al. 2011).

If the measurements to be analyzed were taken at a time point where all relevant state changes induced by the perturbation have become visible, that is, if a steady state is reached in that sense that the signs of the state variations will not change anymore, stronger constraints on the possible network behavior can be derived from the concept of sign consistency. The approach we presented herein uses integer linear programming (ILP) to encode these constraints and to predict the possible changes (increase, decrease, no change) of the activation levels of the involved players for a given experiment. Based on this ILP formulation, we presented four basic optimization routines that not only enable the detection of inconsistencies between measurements and predicted behaviors (introduced as SCEN_FIT problem), but also their removal (problems MCoS, OPT_SUBGRAPH, and OPT_GRAPH). While with minimal correction sets (MCoS) places in the network are identified that might cause the observed inconsistencies for a single experimental scenario, OPT_SUBGRAPH and OPT_GRAPH enable the identification of changes in the network structure to minimize the mismatch between data and model predictions over all experimental scenarios. We also provided enumeration algorithms to find multiple or all solutions that solve the optimization problem equally well. For example, one can determine all optimal subgraphs that minimize the number of inconsistencies between measurements and predictions. However, the enumeration of optimal solutions may quickly become prohibitive in larger networks. A possible workaround is to apply network compression techniques to deal with the combinatorial complexity arising in large-scale networks. This not only allows the enumeration of solutions, but is

also a way to indicate non-identifiabilities in the original network and to represent the obtained solutions in a condensed manner.

Closely related to our sign consistency approach are the methods introduced in Gebser et al. (2011), Guziolowski et al. (2009), and Siegel et al. (2006). This framework is also based on interaction graphs and uses a similar consistency rule as we did herein. However, there are a number of key differences. First, we explicitly allow a “0” change to mark non-affected node states. Second, the four basic problem formulations presented herein go beyond the techniques introduced in Gebser et al. (2011), Guziolowski et al. (2009), and Siegel et al. (2006). In particular, the training of the topology, that is, the identification of inactive or missing interactions based on a library of stimulus–response experiments, was not considered in these works. A third key difference is that we formulated the constraints resulting from the consistency rules as an ILP problem, in contrast to answer set programming (ASP), which was used in Gebser et al. (2011). A follow-up study of the work presented herein, which seeks to derive ASP formulations of the training and enumeration problems formulated herein, has already been initiated in the research group of Steffen Klamt. Within this study, approaches to enable the analysis of networks containing positive feedback loops are also under consideration.

As for the dependency analysis, we used the interaction graph model of ERBB signaling together with a set of experimental high-throughput data from primary human hepatocytes as an application example. Using the newly presented algorithms, we could systematically uncover all inconsistencies between measurements and network topology and give possible explanations for them. Basically, these findings were in accordance with our findings derived with the dependency matrix. Furthermore, novel biological insights for this important signaling pathway could be revealed by listing interactions that are likely to be inactive in hepatocytes and by giving suggestions for possibly missing interactions that, if included, would significantly improve the goodness of fit. Clearly, these predictions await experimental validation.

The ERBB study gave a proof of principle for the presented methodologies, showing the flexibility of both approaches and that they can be applied to a wide range of problems arising when confronting signaling network topologies with experimental data. A key advantage of these interaction-graph-based methods is that only fairly accessible biological knowledge is required. A critical point might be the necessary data discretization. Here, assessing the results with respect to sensitivities of the chosen discretization parameters is one possible way to account for the uncertainties

of the parameter choice. Probably the most serious shortcoming of the presented methods is their poor predictive power in highly interlinked networks. As demonstrated in the ERBB example, it might be necessary to neglect certain influences that are assumed to have a minor impact on the question under study. Although such simplifications are often justified, one should be aware of them when interpreting the results. After all, we will show in Chapter 5 a study where a hybrid network inference approach making use of the dependency matrix can very well be applied to a network containing feedback loops and cross-regulating mechanisms.

4 | Logical Models of Signal Transduction Networks

In logical models, the information that is contained in an interaction graph is extended by rules defining how the discretized state of a node is governed by the states of other nodes. This enables to compute the qualitative input–output behavior of signaling pathways as well as the identification of intervention strategies (see Section 2.4.2). Here, we present some further developments of the methods introduced in Klamt et al. (2006) for the static analysis of logical models. As in the previous chapter, we exemplify the proposed methods by application to ERBB signaling.

First, in Section 4.1 we present with species equivalence classes a new technique for the analysis of logical models. Species equivalence classes were introduced in Samaga et al. (2009), from which Section 4.1 was derived. Section 4.2 deals with the computation of minimal intervention sets (MISs). The concept of MISs has been introduced in Klamt et al. (2006). Here, we present techniques for search space reduction facilitating the enumeration of MISs in networks of realistic size as introduced in Samaga et al. (2010). In Section 4.3, we describe how the predicted binary network response derived from a logical model can be compared to experimental stimulus–response data; this section is adapted from Samaga et al. (2009). Finally, we describe in Section 4.4 the logical model of ERBB signaling derived from the interaction graph model given in Section 3.3.1. Besides analyzing the model with the methods described in Section 2.4.2, we show how the newly proposed techniques apply to it. Again, this is a revision of the descriptions given in Samaga et al. (2009).

4.1 Species Equivalence Classes

Here, we present a new analysis technique for logical networks: we search for *equivalence classes* of network nodes whose activation pattern is completely coupled in logical steady state: species A and B are elements of the same equivalence class if it either holds that their values in steady state are always the same ($A = 0 \Leftrightarrow B = 0$, $A = 1 \Leftrightarrow B = 1$; positive coupling) or always the opposite ($A = 0 \Leftrightarrow B = 1$,

$A = 1 \Leftrightarrow B = 0$; negative coupling), irrespective of the chosen inputs (e.g., ligands). In other words, the state of one species in the equivalence class determines the states of all other species in this class. The relation given above holds for logical steady states where both A and B are determined and where no intervention was made in the network except for the inputs.

An efficient algorithm for computing the equivalence classes can be constructed as follows:

1. Equivalence classes can be computed for a given scenario, so we first define a specific, possibly empty, set of fixed states, typically from some input nodes.
2. For this given scenario, we test systematically for each species whether the respective node is strongly coupled with other nodes or not, independently of external stimuli. For each species A we compute (i) the logical steady states of all other species that result when fixing the state of A to 1 and (ii) the logical steady states of all other species that result when fixing the state of A to 0. A node B whose logical steady state can be determined in both cases and is 1 in one case and 0 in the other case is known to be in one equivalence class with species A: B is positively coupled with A if the two resulting logical steady states of B are 1/0 (it then holds $A = 1 \Rightarrow B = 1$, $A = 0 \Rightarrow B = 0$ and, thus, according to contraposition also $B = 0 \Rightarrow A = 0$, $B = 1 \Rightarrow A = 1$) and negatively coupled if the two logical steady states are 0/1 (it then holds $A = 1 \Rightarrow B = 0$, $A = 0 \Rightarrow B = 1$ and, thus, according to contraposition also $B = 0 \Rightarrow A = 1$, $B = 1 \Rightarrow A = 0$). The case that the logical steady state of a species B is 0/0 or 1/1 (for fixing $A = 1/A = 0$) indicates that this species B can never be activated or never be inhibited, respectively, and thus indicates a semantic problem in the model.

We note that if a species A is coupled with species B, and species B is coupled with species C, we can subsume all three species into one equivalence class.

Composing the equivalence classes in this way it may also happen that species that cannot influence each other, that is, no directed path between both exists, are in one equivalence class due to a common upstream regulator. Consider a network that only contains the interactions $A \rightarrow B$ and $A \rightarrow C$. Fixing the state of B or C to 1 or 0 we cannot conclude any equivalence relations as no further states can be determined. Fixing A to 1 and 0 we find that A is equivalent to B and A is equivalent to C, thus—according to the rule given above—A, B, and C form one equivalence class.

We exemplify the equivalence classes by means of the logical model of the EGF/NRG1 network example given in Figure 2.3. If the states of all input nodes were left open, the model has two equivalence classes: {SOS, Ras, MEK, ERK} and {PI3K, PIP₃, AKT}. If we fix the states of all receptor monomers to 1, we get the two additional classes {EGF, ERBB11} and {NRG1, ERBB23}.

Species equivalence classes are very similar to enzyme (or reaction) subsets known from metabolic networks (Burgard et al. 2004; Pfeiffer et al. 1999), and they help to uncover functional couplings embedded in the network structure. We anticipate that the concept of equivalence classes also provides a basis for model reduction (e.g., when computing logical steady states), similar as it has been employed in metabolic networks (see, e.g., Gagneur and Klamt (2004)).

4.2 Computation of Intervention Strategies and Failure Modes

The identification of combinatorial intervention strategies and the elucidation of failure modes that may cause aberrant behavior of cellular signaling networks are highly relevant topics in cell biology, medicine, and pharmaceutical industry. As already stated in Section 2.4.2, the concept of minimal intervention sets (MISs) allows to tackle those questions within a logical modeling framework. The enumeration of MISs is computationally expensive: each MIS can be represented by an n -element vector where each element (node) is either -1 (constitutive inhibition), 0 (no intervention), or 1 (constitutive activation), such that in a network with n nodes, 3^n possible combinations of interventions exist. Therefore, in order to facilitate the computation of MISs in realistic large-scale problems, algorithmic approaches are required to cope with the combinatorial complexity of this problem. In this section, we introduce several techniques that reduce the search space without losing any solution. We start with definitions giving a rigorous framework of MISs.

4.2.1 Definition of Minimal Intervention Sets and General Algorithm

The idea of intervention sets is to search for combinations of constitutive activations and deactivations of species in the network that provoke a desired network response. In other words, we search for a pattern of species values to be fixed (the intervention set), so that certain species reach a predefined value in the resulting logical steady state. Please note that in the following, logical steady state denotes the logical steady state in three-valued logic as introduced in Section 2.4.2. In the

logical framework, interventions (constitutive activations/deactivations) are treated as nodes whose values are clamped to 1 or 0, respectively. In the biological context, examples for constitutive activations are mutations that lead to constitutively activated species or a continuous stimulation with external signals (e.g., by a ligand), whereas a constitutive inhibition may correspond to gene knock-outs, specific blocking of the activity of a certain species, or—approximately—the knock-down of certain nodes using RNA interference techniques.

Given a logical network with species v_1, \dots, v_n , the desired network response is specified by the intervention goal $x^G = (x_1^G, \dots, x_n^G)^T$, $x_i^G \in \{0, 1, *\}$. Those species v_i whose desired value x_i^G is specified as 0 or 1 are the *target species*, and by $I_T = \{i : x_i^G \in \{0, 1\}\}$ the index set of the target species is given. To distinguish between nodes that shall be activated and nodes that shall be inhibited, we introduce the index sets $I_{T+} = \{i : x_i^G = 1\}$ and $I_{T-} = \{i : x_i^G = 0\}$. In the examples given below, we will sometimes simply write $\{v_i = x_i^G\}_{i \in I_T}$ to declare the intervention goal. Optionally, we may define natural or external side constraints for a set of nodes $\{v_i\}_{i \in I_E}$ by introducing the state vector $x^C = (x_1^C, \dots, x_n^C)^T$ satisfying $x_i^C \in \{0, 1\}$ if $i \in I_E$, and $x_i^C = *$ if $i \in \{1, \dots, n\} \setminus I_E$. This enables us to state the conditions under which the intervention goal should be fulfilled, for example, in the presence/absence of a ligand. The intervention goal together with the given side constraints forms an intervention scenario $P = (x^G, x^C)$. In general, one might also consider intervention problems that are composed of several intervention scenarios and search for intervention sets that fulfill for each scenario the intervention goal considering the respective side constraints (Samaga et al. 2010). However, here, we restrict ourselves to intervention problems with one intervention scenario. An intervention set for the intervention problem $P = (x^G, x^C)$ can now be defined as a set of states $x_i^\# \in \{0, 1\}$, $i \in I_\# \subseteq \{1, \dots, n\}$, satisfying $x_i^S(x^0) = x_i^G$ for all $i \in I_T$, where $x_i^S(x^0)$ denotes the logical steady state of species v_i that results from the fixed values given by the interventions and side constraints, that is, $x_i^0 = x_i^\#$ for all $i \in I_\#$, and $x_i^0 = x_i^C$ for all $i \in \{1, \dots, n\} \setminus I_\#$. As for the intervention goals, we will sometimes use the simplified notation $\{v_i = x_i^\#\}_{i \in I_\#}$ to specify an intervention set. For practical reasons we are only interested in (*support-minimal* intervention sets (MISs), in which no subset of the involved interventions fulfills the intervention goal. Furthermore, we concentrate on the nodes of the network although, in principle, we could also search for interventions at the hyperedges. However, in cell signaling, the species rather than the interactions can be externally controlled—in contrast to metabolic networks, where typically the reactions are subject to interventions (Klamt 2006).

The basic algorithm for computing MISs, as outlined in Klamt et al. (2006) and Klamt et al. (2007), checks systematically combinations of interventions (first of size 1, then of size 2, and so on) whether or not they lead to a fulfillment of the intervention goal when the network reaches the logical steady state (see also pseudocode in Figure 4.1). For minimality, supersets of found MISs need not be further considered. Clearly, this brute-force approach leads quickly to an explosion of possible candidates for higher cardinalities. Therefore, strategies are required to make this computation also feasible in larger networks. One simple yet very useful strategy is to limit the number of allowed interventions per MIS. In fact, one is typically interested in the small MISs (see e. g. Saez-Rodriguez et al. (2007)). Other strategies try to reduce the search space without losing any solution, and in the following section, we shall present several of those techniques.

4.2.2 Methods for Search Space Reduction

The primary goal of search space reduction is to reduce the number of candidates that have to be tested for the MIS property. Although the computation of the logical steady state for a given MIS is polynomial¹, these tests turn out to be the most time-consuming parts of the whole algorithm. We use two different strategies for search space reduction:

1. In a preprocessing step, one seeks to safely discard some of the $2n$ possible single interventions, either because they can be shown not to be part of any MIS (Section 4.2.2.1), or because there are interventions being equivalent with respect to the intervention goal(s); in the latter case, only one representative needs to be considered, and in a post-processing step, the MISs with the equivalent interventions can be easily obtained (Section 4.2.2.2).
2. A second technique, applied during the main algorithm, is to discard certain MIS candidates without explicit testing by employing some simple rules (Section 4.2.2.3).

4.2.2.1 Exploiting Dependencies

A suitable technique to discard uninfluential single interventions is to consider only those interventions that can affect the target species in the desired direction. The

¹For logical steady state calculation see Section 2.4.2.3.

required information is captured in the *dependency matrix* (Klamt et al. 2006) that can be derived from the interaction graph underlying the logical network (see Section 2.4). In Klamt et al. (2006), the possibility of such a search space reduction was already mentioned, however, just for calculations in interaction graphs, not in logical models.

From the dependency matrix, we cannot only deduce whether a certain species influences at least one target node at all (if not, then neither constitutive activation nor deactivation needs to be considered for this node), but also whether it acts as activator and/or inhibitor. Recall that a node v_i is an activator of v_j if only positive paths lead from v_i to v_j . If v_j is a target to be switched on, a constitutive inhibition of v_i cannot contribute to achieve this goal and, thus, does not need to be considered. Analogously, if v_i is an inhibitor for v_j , a constitutive activation of v_i cannot contribute to switch v_j on. Note that we do not need to distinguish between weak/strong activators as usually done in the dependency matrix (see Section 2.4.1.2), as we are here only interested in the logical steady state, not in states that can be reached during the transition. Taking those interdependencies into account may drastically decrease the number of interventions and, thus, of MIS candidates to be checked.

To illustrate this, consider the EGF/NRG1 example model given in Figure 2.3. Assume the intervention goal is to activate ERK and inactivate PI3K, that is, $\{\text{ERK} = 1, \text{PI3K} = 0\}$, under the side constraints $\{\text{ERBB1} = 1, \text{ERBB2} = 1, \text{ERBB3} = 1\}$. The dependency matrix of this example model is shown in Figure 2.2. As AKT is a neutral factor both for PI3K and ERK, neither constitutive activation nor inhibition needs to be considered for this node. GAB1 and PIP₃ are activators for PI3K and do not influence ERK, implying that only constitutive inhibitions of GAB1 and PIP₃ are useful. Inhibition of MEK can be excluded, as MEK is an activator for ERK and an inhibitor for PI3K. Inhibition of ERBB1, ERBB2, or ERBB3 need not be considered as these interventions are already fulfilled by the given side constraints. Obviously, constitutive inhibition of ERK and activation of PI3K can never be part of intervention sets that fulfill the intervention goal. Note that, in this example, we allow interventions at nodes with fixed values (side constraints) and directly at the target species; sometimes, it is reasonable to exclude these interventions from the outset. Taken together, the number of possible single interventions reduces in this example from 34 to 24 (six nodes can be set to 0, two nodes set to 1, and eight nodes can be set to 0 or 1), and the number of possible MIS candidates decreases from 3^{17} to $3^8 \cdot 2^8$.

4.2.2.2 Exploiting Fault Equivalence Classes

Another preprocessing step for reducing the number of single interventions exploits the equivalence of certain interventions with respect to the intervention goal. Two interventions W and Z are equivalent if W can be replaced by Z in all MISs where W is a part of yielding a new set of MISs (the same must hold in the other direction). Hence, it suffices to consider only one representative of each class of equivalent interventions. Searching for equivalent interventions with respect to a certain intervention goal is highly related to equivalence fault collapsing in logical networks studied for systems testing in electrical engineering (Abramovici et al. 1990; McCluskey and Clegg 1971). In electrical engineering, one considers stuck-at-0 (s-a-0) faults and stuck-at-1 (s-a-1) faults, which may occur, for example, by physical damages. These faults directly correspond to the constitutive inhibitions (s-a-0) and activations (s-a-1) in our logical framework—though we are interested in *constructing* suitable intervention sets whereas in electrical engineering, one is interested in *identifying* faults—and we can easily adapt the methods from fault collapsing for our application. We restrict ourselves to considering structurally equivalent interventions that can be determined by a local analysis of the network. The more general concept of functional equivalence relations requires a global analysis, and determining whether two arbitrary faults are functionally equivalent is an NP-complete problem (Abramovici et al. 1990).

The occurrence of equivalent interventions in logical networks is based on the fact that the controlling value of an AND gate is 0 and the controlling value of an OR gate is 1; that is, setting one of the input values of an AND gate to 0 determines the gate output to 0, whereas setting one of the input values of an OR gate to 1 determines the gate output to 1 (Abramovici et al. 1990). The intervention $v_j = 1$ where v_j is the output of an OR gate is therefore equivalent to setting any input of the OR gate to 1. In analogy, the intervention $v_j = 0$ where v_j is the output of an AND gate is equivalent to setting any input of the AND gate to 0. To state the conditions under which two interventions in our network are equivalent, recall that the network is represented as logical interaction hypergraph where each hyperedge represents an AND gate and different hyperedges pointing into the same node are OR-connected (see Section 2.4.2.1). Furthermore, the inputs of the AND gates can be negated using the NOT operator. A hyperedge that has only one tail node represents a simple activation or inhibition depending on whether the value of the input node is negated or not. The intervention $v_j = 1$ (s-a-1 fault) is equivalent to $v_i = 1$ (or to

$v_i = 0$) with respect to the set of target species $\{v_i\}_{i \in I_T}$ if the following conditions hold:

- 1.1 v_i is a direct predecessor of v_j ,
- 1.2 v_i is connected to v_j via a hyperedge that has v_i as the only tail node (i.e., via a simple edge),
- 1.3 the value of v_i is not negated (for equivalence with $v_i = 1$) / negated (for equivalence with $v_i = 0$),
- 1.4 the out-degree of v_i is 1 (i.e., one hyperedge is pointing out of v_i),
- 1.5 v_i is not a target species (i.e., $i \notin I_T$),
- 1.6 the value of v_j is not given as external side constraint (i.e., $j \notin I_E$).

The intervention $v_j = 0$ (s-a-0 fault) is equivalent to $v_i = 0$ (or to $v_i = 1$) with respect to the set of target species $\{v_i\}_{i \in I_T}$ if the following conditions hold:

- 2.1 v_i is a direct predecessor of v_j ,
- 2.2 the in-degree of v_j is 1 (i.e., one hyperedge is pointing into v_j),
- 2.3 the value of v_i is not negated (for equivalence with $v_i = 0$) / negated (for equivalence with $v_i = 1$),
- 2.4 the out-degree of v_i is 1,
- 2.5 v_i is not a target species (i.e., $i \notin I_T$),
- 2.6 the value of v_j is not given as external side constraint (i.e., $j \notin I_E$).

Conditions 1.1–1.3 and 2.1–2.3 follow directly from the fact that an OR gate is controlled by any input set to 1, whereas an AND gate is controlled by any input set to 0. Condition 1.4/2.4 is needed to ensure that the upstream node (v_i) may not influence the target nodes via other paths not leading over v_j . A special case for s-a-0 faults in which we can find an equivalence relation between two adjacent nodes although condition 2.4 is not met arises if all hyperedges pointing out of v_i point into the same node v_j (e.g., the logical function $x_j = x_i \cdot x_k + x_i \cdot x_l$). The need for conditions 1.5/2.5 is based on the fact that an intervention at a certain node does not generally lead to the fulfillment of an intervention goal for a preceding species.

For example, $\{\text{PI3K} = 1\}$ is an MIS for the intervention goal $\{\text{PI3K} = 1, \text{AKT} = 1\}$ in the EGF/NRG1 example model (Figure 2.3). Ignoring condition 1.5, we find that $\text{PIP}_3 = 1$ is equal to $\text{PI3K} = 1$. However, $\{\text{PIP}_3 = 1\}$ is not an MIS for the intervention goal as it does not lead to $\text{PI3K} = 1$. Conditions 1.6/2.6 are needed if an intervention at a node whose value is given as side constraint is not excluded.

The rules described above detect the equivalence relations between interventions on the head and tail nodes of a hyperedge. Apart from that, we can also find equivalence relations between interventions on input nodes of the same gate. Some of these relations are indirectly detected by applying the rules above since pairwise equivalent interventions can be combined to equivalence classes. For example, if conditions 1.1–1.5 hold for the two inputs v_i, v_k of an OR gate with output v_j , the equivalent interventions $v_j = 1 \Leftrightarrow v_i = 1$ and $v_j = 1 \Leftrightarrow v_k = 1$ are merged to the equivalence class $\{v_i = 1, v_j = 1, v_k = 1\}$, reflecting also the equivalence between interventions at v_i and v_k . Whereas for OR gates, all equivalence relations are detected in this way, s-a-0 faults at the inputs of an AND gate (hyperedge) can be equivalent although the equivalence with interventions at the output of the gate does not hold. This arises if condition 2.2 is not met, and we therefore have to consider the possible equivalence of interventions at the input nodes for this case separately: let v_1, \dots, v_k be the inputs of an AND gate and let s_i denote whether the value of v_i is negated ($s_i = 1$) or not ($s_i = 0$). It then holds that the interventions $v_i = s_i$ and $v_j = s_j$, $i, j \in \{1, \dots, k\}$, are equivalent if (i) the out-degree of v_i and v_j is 1 and (ii) neither v_i nor v_j is a target species.

In the EGF/NRG1 example model (Figure 2.3), the following classes of equivalent interventions with respect to the target species EGF and PI3K can be found: $\{\text{SOS} = 1, \text{Ras} = 1\}$, $\{\text{MEK} = 1, \text{ERK} = 1\}$, $\{\text{ERBB1} = 1, \text{GRB2} = 1\}$, $\{\text{PI3K} = 1, \text{GAB1} = 1\}$, $\{\text{SOS} = 0, \text{Ras} = 0\}$, and $\{\text{MEK} = 0, \text{ERK} = 0\}$. As we only consider one representative of each class of equivalent interventions as a possible element of an MIS, the number of candidates for single interventions is reduced by six.

We emphasize that fault equivalence classes are specific for a given set of target species. One may exploit fault equivalence classes in combination with the dependency analysis: before computing the MISs for one specific intervention scenario, it is advantageous to first exploit the species interdependencies as described above and then, for the remaining interventions, to determine the fault equivalence classes further reducing the set of possible interventions. As an example, consider again the intervention goal $\{\text{ERK} = 1, \text{PI3K} = 0\}$ with side constraints $\{\text{ERBB1} = 1, \text{ERBB2} = 1, \text{ERBB3} = 1\}$. As described in Section 4.2.2.1, we can ex-

clude the following interventions by exploiting the interdependencies: $\{\text{ERBB1} = 1\}$, $\{\text{ERBB2} = 1\}$, $\{\text{ERBB3} = 1\}$, $\{\text{GAB1} = 1\}$, $\{\text{PI3K} = 1\}$, $\{\text{PIP}_3 = 1\}$, $\{\text{AKT} = 1\}$, $\{\text{AKT} = 0\}$, $\{\text{MEK} = 0\}$, and $\{\text{ERK} = 0\}$. Of the six equivalence classes given above, only $\{\text{SOS} = 1, \text{Ras} = 1\}$, $\{\text{MEK} = 1, \text{ERK} = 1\}$, $\{\text{ERBB1} = 1, \text{GRB2} = 1\}$, and $\{\text{SOS} = 0, \text{Ras} = 0\}$ contain no interventions that have already been excluded. The interaction $\text{PIP}_3 \rightarrow \text{AKT}$ is removed from the interaction graph by exploiting the species dependencies, as AKT has no effect on any of the target species. Hence, considering a s-a-0 fault in GAB1, the species PIP_3 now fulfills conditions 1.1–1.5 (including condition 1.3, which was not fulfilled in the original network) and we have the additional equivalence class $\{\text{PIP}_3 = 0, \text{GAB1} = 0\}$. Thus, the number of single interventions from which MISs are combined in the main algorithm can be further decreased from 24 to 19 by taking into account the equivalent interventions. We finally achieve an enormous reduction of the potential search space from 3^{17} (129 140 160) to $3^6 \cdot 2^7$ (93 312): six species can be set to 0 or 1, six species set to 0, and one species set to 1.

4.2.2.3 Exploiting Dependencies in Intervention Goals with Multiple Targets

If an intervention goal includes more than one target species, we can further reduce the search space by not testing those combinations of interventions (MIS candidates) that do not influence *all* targets in the appropriate direction. In contrast to the reduction methods described in Sections 4.2.2.1 and 4.2.2.2, this is not done in preprocessing, but during the iterations: before we compute the resulting logical steady state for an MIS candidate, we check whether its interventions can influence all the target species with appropriate sign. Consider the intervention goal with target species $T^+ = \{v_1\}$ and $T^- = \{v_2\}$. In this case, each intervention set must contain at least one intervention that has a positive influence on v_1 and one intervention that has a negative influence on v_2 . MIS candidates that do not fulfill this condition will not be checked in the current iteration, but will be combined with further interventions in the next iteration.

We can now provide the scaffold of the full MIS algorithm in pseudocode (Figure 4.1). The calculation of MISs including the extensions described herein was implemented in *CellNetAnalyzer* (Klamt et al. 2007). Results from Benchmarks illustrating the beneficial effect of search space reduction are shown in Section 4.4.2.3.

```

Function MIS  $\leftarrow$  ComputeMIS (Model, TargetsPlus, TargetsMinus, MaxSize)

inputs:
  Model: a logical model in hypergraph representation
  TargetsPlus: all nodes that shall be switched on
  TargetsMinus: all nodes that shall be switched off
  SideConstraints: given side constraints
  MaxSize: maximum cardinality of the minimal intervention sets (MISs) to be computed

output:
  MIS: the set of MISs

AllowedInterventions  $\leftarrow$  DependencyAnalysis (Model, TargetsPlus, TargetsMinus, SideConstraints) ;
(AllowedInterventions, Classes)  $\leftarrow$  FaultEquClasses (AllowedInterventions, Model, TargetsPlus, TargetsMinus) ;
MIS  $\leftarrow$  {};
MIS_CandidatesNew  $\leftarrow$  {{}};

for i from 1 to MaxSize do
  MIS_Candidates  $\leftarrow$  MIS_CandidatesNew;
  MIS_CandidatesNew  $\leftarrow$  {};
  for each M in MIS_Candidates do
    for each K in AllowedInterventions do
      M_new = M  $\cup$  K;
      if M_new is support-minimal w.r.t. MIS and MIS_CandidatesNew then
        if M_new can affect all TargetsPlus and all TargetsMinus in required directions then
          GoalsOK  $\leftarrow$  true;
          LSS  $\leftarrow$  ComputeLogicalSteadyState (M_new, Model, SideConstraints) ;
          if LSS does not fulfill the intervention goal then
            GoalsOK  $\leftarrow$  false;
            break;
          end
          if GoalsOK == true then
            append M_new to MIS;
          else
            append M_new to MIS_CandidatesNew;
          end
        end
      end
    end
  end
end
MIS  $\leftarrow$  ExpandFaultEquClasses (MIS, Classes);

```

Figure 4.1: Pseudocode for the computation of minimal intervention sets.

4.3 Analysis of Experimental Data

The logical model can in a similar way be used for the analysis of experimental stimulus–response data as its underlying interaction graph. Where the interaction-graph-based analysis reveals whether the measured changes (ups and downs) induced by external perturbations are in accordance with the causal dependencies in a given network structure, the logical model enables to predict the binary network response in steady state resulting from a given set of stimuli and to compare this response with the measured species activities (on and off). Again, we assume we are given a certain “master model”, here, a logical model in hypergraph form, and a set of experimental scenarios. Each scenario is characterized by a set of species (inputs) whose state is fixed to 0 (e.g., a species whose activity is blocked by an inhibitor) or 1 (e.g., an added ligand) and shows the measured response, usually the phosphorylation state, of a set of readout species. In order to compare the measurements with predictions from the logical model, the measured states have to be discretized to 0 or 1, that is, one has to decide for each measured species whether it was activated (1) or not (0) in response to the stimulation. General rules for data discretization are hard to define, as data structure and characteristics of the system play an important role. We describe the data discretization for the ERBB example in Section 4.4.3. To derive the model predictions, we set for each experimental scenario the states of the input nodes to the fixed values and compute the resulting logical steady state. As described in Section 2.4.2, negative feedback loops hamper the computation of the logical steady state. Thus, data analysis based on the logical steady state is especially suited to uncover discrepancies between experimental results and our current qualitative knowledge for the initial response, for example, to analyze which species can be activated at all in certain experimental scenarios.

Following the data analysis, one would try to reduce the identified discrepancies by changing the model accordingly. In the application example we will show herein (Section 4.4.3), we retrieved such model changes manually. However, similar to our introduced integer linear programming (ILP) approach for interaction graphs (Section 3.2), one can also train a given network structure to a set of experimental data based on logical steady state analysis (Saez-Rodriguez et al. 2009): Based on an interaction graph master model built from prior knowledge, a superstructure of logical models is created that contains all possible logical gates. In a subsequent optimization step, model structures whose predicted logical steady states show the smallest discrepancies to the experimental data are identified. Thereby, the objective function

balances goodness of fit with model complexity. Finally, new links that improve the fit to data and are not present in the initial model structure are added. The original implementation of the method was based on genetic algorithms (Saez-Rodriguez et al. 2009). In subsequent studies, also ILP (Mitsos et al. 2009) and Answer Set Programming (Guziolowski et al. 2013) formulations were derived, which show a much better performance.

We note that the comparison of the measured ups and downs based on interaction graphs (performed in Chapter 3) and the comparison of the discretized data with the predicted logical response are naturally correlated. However, they do not necessarily lead to exactly the same results. An example: assume you have an input stimulus (ligand L), which may activate a target species S via two independent pathways, one of both leading over an intermediate species A, for which we have an inhibitor I. If we compare the scenario *stimulation with L and adding inhibitor I* against *stimulating with L* via, for example, dependency analysis, we would expect a decrease in the activation level of S since the inhibited species A is an activator for S. However, the phosphorylation state of S might show no significant change in the dependency analysis (i.e., leads to a gray entry as in Figures 3.6 and 3.7) due to the alternative pathway not affected by the inhibitor. In contrast, if the two pathways from L to S are OR-connected in the logical model, the latter would still predict S to be active. Another difference in the data analysis based on interaction graphs versus logical models is that the former compares species changes obtained from two different experiments (e.g., experiment with/without inhibitor), whereas the logical model gives for each experiment one prediction for the state of each species.

4.4 The Logic of ERBB Signaling

In this section, we introduce a logical model of ERBB signaling as an example for a large-scale logical model of a signal transduction pathway. The interaction graph underlying this logical model (cf. Section 2.4.2) is the ERBB model presented in Section 3.3.1. As already stated above, we built the ERBB model based on a stoichiometric pathway map (Oda et al. 2005) and additional information from the literature. In the following, we discuss some technical problems that arise when converting a stoichiometric model into a logical one and propose some general guidelines how to deal with them (Section 4.4.1). In Section 4.4.2, we analyze the logical model by applying the methods introduced in Section 2.4.2 as well as the new extensions presented in Sections 4.1 and 4.2. Finally, we show in Section 4.4.3 the analysis

of the experimental data introduced in Section 3.3.2.1 according to the methods in Section 4.3.

4.4.1 From a Stoichiometric Model to a Logical Model

Translating a stoichiometric model into a logical model is not a trivial task and requires additional information. Whenever a species is only influenced by one upstream molecule, the interpretation as a Boolean function is straightforward: the downstream species is active (state 1) if and only if the upstream species has a positive influence and its state is 1, or if and only if the upstream species has a negative influence and its state is 0 (Figure 4.2(a)). In some other cases it is clear how to code the dependency in a logical function—for example, the formation of a complex (e.g., the heterodimerization of c-Jun and c-Fos to the transcription factor AP1 as shown in Figure 4.2(b), or binding of a ligand to a receptor), where all involved proteins have to be present to trigger downstream events and are thus connected with an AND gate. Furthermore, we use an OR gate whenever a protein can be recruited and activated through different receptors or adapter proteins (Figure 4.2(c)).

However, in many cases the stoichiometric information is not sufficient to approximate the activation level of a species as a logical function of the states of its upstream effectors, and one requires additional—mainly qualitative—information, which can often be obtained from the literature. The two main cases that can arise are the following:

- A species is positively influenced by two (or more) upstream molecules, for example, a protein that can be phosphorylated by different kinases (see Figure 4.2(d)). Here, the decision whether both kinases are necessary or if one suffices, that is, whether to use an AND or an OR, cannot be made on the basis of the information that is contained in a stoichiometric model. However, the necessary information can often be obtained from related literature (e.g., from knock-out studies where one of both effectors has been removed, or if an inhibitor is available for an upstream species).
- A species is positively influenced by one species, for example, by a kinase, and negatively influenced by another, for example, by a phosphatase. In this case, we cannot be sure what happens when both the kinase and the phosphatase are present; it will depend on the respective strength, described as kinetic parameters in a quantitative model, and may differ in different cell types. However,

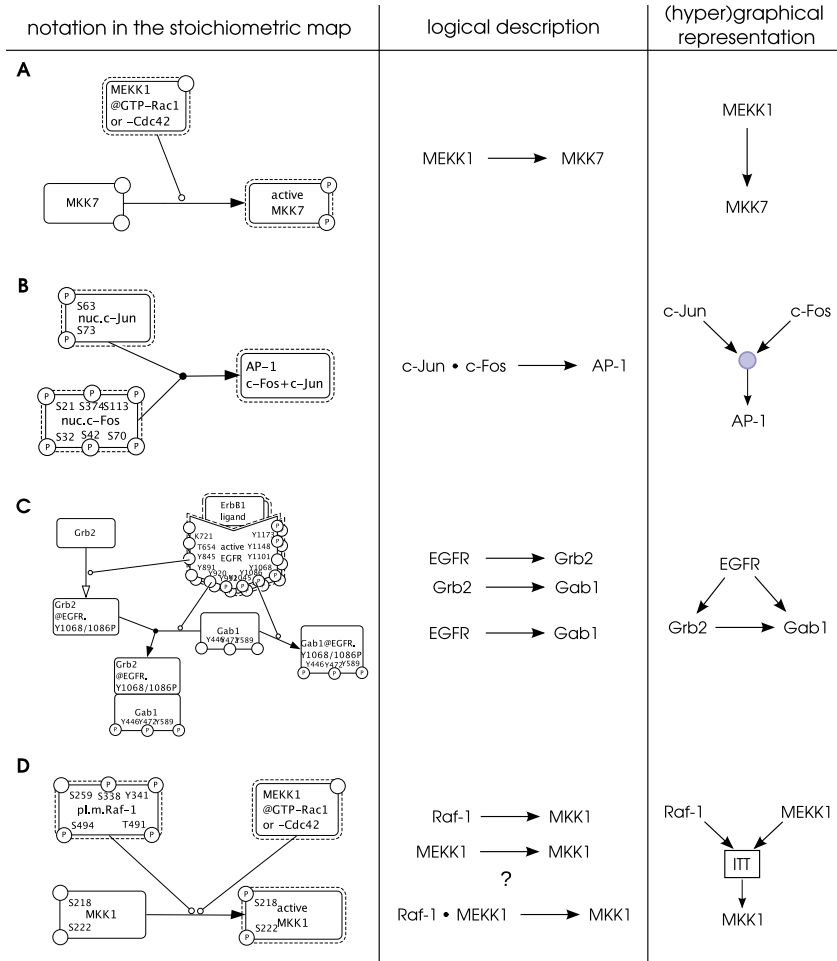


Figure 4.2: Examples illustrating the translation of the stoichiometric EGFR model into a logical description. The examples are taken from the stoichiometric map of Oda et al. (2005). **(a)** The activation level of MKK7 is only influenced by one upstream molecule (active MEKK1). **(b)** c-Jun and c-Fos form the transcription factor AP1. Accordingly, both species are combined with an AND gate (denoted by \cdot in the logical equations). **(c)** GAB1 can bind directly to EGFR homodimers or via receptor-bound GRB2. For the activation of downstream elements, the activation mechanism of GAB1 does not make a difference, what results in a logical OR connection represented by two (independent) activation arrows: GRB2 \rightarrow GAB1 OR EGFR \rightarrow GAB1. **(d)** In this example, we cannot immediately decide whether both Raf1 and MEKK1 are necessary for the activation of MKK1, or if the activation of one of these two kinases suffices. Further information is required, or an ITT gate can be used (in model M1 we used an OR based on facts published in the literature). Figure reprinted from Samaga et al. (2009).

the activation of phosphatases often occurs as a temporarily secondary event upon stimulating a signaling pathway (required for switching off the signal). Therefore, they may be neglected when considering the early events, that is, the initial response of the network that follows upon stimulation.

We also have to keep in mind that, in all cases, the logical description is only a discrete approximation of a quantitative reaction. As already stated in Section 2.4.2.1, we can use incomplete truth tables in those cases where neither an AND nor an OR is a good approximation (Klamt et al. 2006). This operator, herein after referred to as *ITT gate*, returns 1 if and only if all positive arguments are 1 and all negative arguments are 0, and returns 0 if and only if all positive arguments are 0 and all negative arguments are 1. In all other cases, no decision can be made and the response of the molecule remains undefined, that is, the state of the respective output species is *. Using ITT gates may limit the determinacy of the model: when performing stimulus–response simulations, it can happen that some states cannot be determined uniquely. However, it allows for a safer interpretation of the results. To illustrate this concept and to discuss uncertainties in our reconstructed logical model—in the following referred to as model M1—we consider a model variant M2 where the activation mechanisms of 14 proteins are described with ITT gates reflecting the uncertainties in the logical description of M1 (see Appendix C.4). In this way, model M2 accounts explicitly for the uncertainties in the logical concatenation of different signals; however, it cannot account for uncertainties that are captured in the wiring diagram itself.

We set-up models M1 and M2 with ProMoT (Mirschel et al. 2009) and exported the mathematical description as well as the graphical representation to the analysis tool *CellNetAnalyzer* (Klamt et al. 2007). Model M1 is shown in Figure 4.3; it contains 104 nodes and 202 interactions (hyperedges). All species and interactions are listed in Appendix C.1 and C.2, respectively. Two interactions of the ERBB interaction graph model (Figure 3.2) are not part of the logical model, as they do not change the logical function of their target node, or as the exact mechanism of the interaction is unknown (see Appendix C.3). The model includes 28 input nodes, whose regulation is not explicitly considered in the model, but which can be used to simulate different scenarios. Besides ligands and receptors these include some phosphatases with unknown activation mechanism. For all input nodes, a default value is given in Appendix C.2 and indicated in Figure 4.3; this value is used for the logical analyses unless otherwise specified. The logical models contain two types of

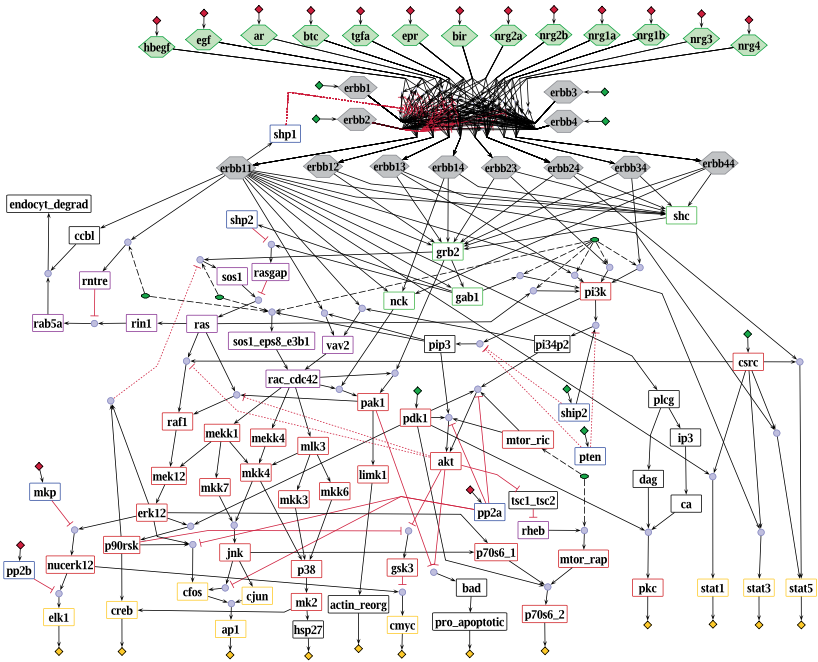


Figure 4.3: Logical model of the ERBB receptor signaling pathway represented in ProMoT. Blue circles symbolize AND connections. Inputs with default value 0 (1) are indicated by red (green) diamonds. Yellow diamonds stand for the outputs of the model. Gray hexagons represent receptor monomers, homodimers, and heterodimers, and green hexagons stand for the 13 different ligands. The remaining species (symbolized with rectangles) are colored according to their function: red: kinases; blue: phosphatases; yellow: transcription factors; green: adapter molecules; violet: small G proteins and GAPs/GEFs; black: other. Green ellipses symbolize reservoirs, dummy species are not displayed (see Appendix C.1). The box in the upper part of the network contains binding of the ligands to the receptor and receptor dimerization, showing the high combinatorial complexity. Black arrows indicate activations, red blunt-ended lines stand for inhibitions. Dotted lines represent “late” interactions (with attribute $\tau = 2$) that are excluded when studying the initial network response. Dashed lines indicate connections from reservoirs. Figure adapted from Samaga et al. (2009).

auxiliary species that have no biological correspondence, but are needed for modeling purposes: *dummy* species and *reservoirs*. Both types of species are explained in Appendix C.1.

4.4.2 Theoretical Analysis of the Logical ERBB Model

4.4.2.1 Qualitative Input–Output Behavior

Once the network construction has been completed, one may start to perform discrete simulations. We will start by computing the logical steady state that follows from exposing the network to a certain input stimulus, that is, by propagating the signals from the input to the output layer. Feedback loops, especially negative feedback loops, may hamper this kind of analysis of the input–output behavior (cf. Section 2.4.2.3). However, herein we will focus on the initial response of the network nodes induced by external stimulations or perturbations. Assuming that the system is in a pseudo-steady state at the beginning, the initial response of a node is governed by the paths connecting the inputs with this node, whereas feedback loops are secondary events that can only be activated at a later time point when each node in the loop has exhibited its initial response. Although path/cycle length is no precise measure for the velocity of signal transduction, the comparable average length of input/output paths (19) and feedback loops (17) in the ERBB model supports the assumption that the initial response of the network nodes is dominated by the input/output paths, whereas feedback loops may overwrite the initial response of the network nodes only after a certain time period with significant length—again, feedback loops can causally not be activated before the initial response occurred. To decouple the initial response from the activity of the feedback loops, we proceeded as follows: we assigned to each reaction a time variable τ determining whether the reaction is active / available during the initial response (i.e., is an early event; $\tau = 1$) or not (late event; $\tau = 2$). In each negative feedback loop we identified the node Z that has the shortest distance to the input layer. This node Z can be considered as the initialization point of the feedback loop, and we then assigned $\tau = 2$ to the “last” interaction of the feedback loop closing the cycle in node Z (i.e., points into Z). For example, in a causal chain $\text{Input} \rightarrow A \rightarrow B \rightarrow C \dashv D \rightarrow B$, we would consider $D \rightarrow B$ as a late event. In this way we interrupt the feedback loop and the logical steady states computed in the network reflects the initial response of the nodes. Strikingly, it was sufficient to consider only two interactions as late event to break all feedback loops in the logical ERBB network (the negative influence of SHP1 on phosphory-

lation of ERBB1 dimers, and the inhibitory phosphorylation of SOS1 by ERK1/2 or p90RSK). With this acyclic network, a unique logical steady state in which all states are either 0 or 1 followed for any set of input values in model M1. The assignment “late” was not only reasonable for selected interactions in feedback loops, but also for three interactions involved in negative feed-forward loops downregulating the signaling after a certain time. The time variables for each reaction can be found in Appendix C.2. It is important to mention that the logical steady state computed for a given set of input states does not necessarily reflect the activation pattern in the cell at one particular point of time. Instead, it reflects for each species the initial response to the stimulus. The time range in which this initial response takes place can differ for each molecule—typically, a species situated in the upper part of the network (e.g., a receptor) responds faster to the stimulus than a species of the output layer (e.g., a transcription factor).

We determined the network response in model M1 upon stimulation with the different ligands, focusing on the early events, that is, the interactions with $\tau = 2$ were set to 0. We found that the outputs can be divided into two groups: the majority of the output elements can be activated by all possible dimers. However, PKC, STAT1, STAT3, and STAT5 can only be activated through ERBB1 homodimers (PKC, STAT1, STAT3), or ERBB1 homodimers and ERBB2/ERBB4 dimers (STAT5). Accordingly, stimulation with neuregulins does not result in activation of the protein kinase PKC and the transcription factors STAT1 and STAT3, in contrast to stimulation with the other ligands that activate all output molecules except the pro-apoptotic effect of BAD, which is repressed. This is due to the fact that the neuregulins, unlike the other ligands, do not bind to ERBB1 and, thus, cannot activate ERBB1 homodimers.

Strikingly, despite of the 14 ITT gates in model M2, the logical steady state in response to ERBB1 homodimers could still be determined in model M2 and did not differ from M1. This observation reflects a high degree of redundancy in at least some parts of the network. The state of each of the different kinases phosphorylating p38 or MKK4 is, for example, only dependent on the activity of Rac/CDC42, implying that these kinases are always activated together. Thus, the input–output behavior of the network can be uniquely predicted for all ligands that activate ERBB1 homodimers (i.e., for all ligands except neuregulins). In contrast, model M2 failed to predict the response for some nodes if other dimers (in absence of the ERBB1 homodimer) are stimulated. This concerns, in particular, most of the output nodes. Only the states of PKC, STAT1, STAT3, and STAT5 could be determined: as in model

M1, these proteins can only be activated by ERBB1 homodimers, except STAT5 that is “on” in response to ERBB2/ERBB4-dimers. The indeterminacy of M2 with respect to stimulations of dimers others than ERBB1 homodimers can be explained by the uncertainty (ITT gate) in the activation of Rac/CDC42.

4.4.2.2 Species Equivalence Classes in the ERBB Model

Studying the input–output behavior of model variant M2, we already realized that certain species in the network show strongly coupled behavior. These species can be detected in a systematic way by computing the species equivalence classes as introduced in Section 4.1.

Figure 4.4 and Table 4.1 show the equivalence classes in the ERBB model variants M1 and M2 for early signal propagation, where the states of all ligands and receptors were left open. The states of the other inputs were fixed to their default value as given in the model description (Appendix C.2). In model M1, we found six equivalence classes (numbered 1–6 in Table 4.1), the largest comprising 24 species. The latter includes parts of PI3K signaling and the Rac-induced parts of the MAPK cascades reflecting the strong coupling of these two major pathways. In model M2, this equivalence class (number 1) splits into three smaller ones (1a, 1b, 1c; see Table 4.1), because the ITT gates introduce uncertainties that lead to a decoupling of the two pathways (see Figure 4.4). The other equivalence classes of M2 hardly differ from the ones in M1, again indicating that alternative pathways contribute rather to a higher degree of redundancy than to a higher degree of freedom regarding the potential input–output behavior. The new concept of species equivalence classes thus helps to identify coupled parts of the pathway and to point to redundant network structures.

4.4.2.3 Computation of Minimal Intervention Sets in the Logical ERBB Model

An application of minimal intervention sets (MISs) is to determine failure modes in the network that lead to an activation of elements of the output layer without any external stimulation of the cell. For example, we can search for failures that activate ERK1/2 and AKT, two central players controlling growth, proliferation, and apoptosis, when no ligand is present. Thus, we defined the intervention goal $\{\text{ERK1/2} = 1, \text{AKT} = 0\}$ with the default values given in Appendix C.2 as side constraints. Considering only the early events and not allowing interventions at the target species or at any species with fixed value, 15 MISs of size 1 fulfilled the intervention goal: activation of any of the eight ERBB dimers, or of Ras, SOS1, SHC,

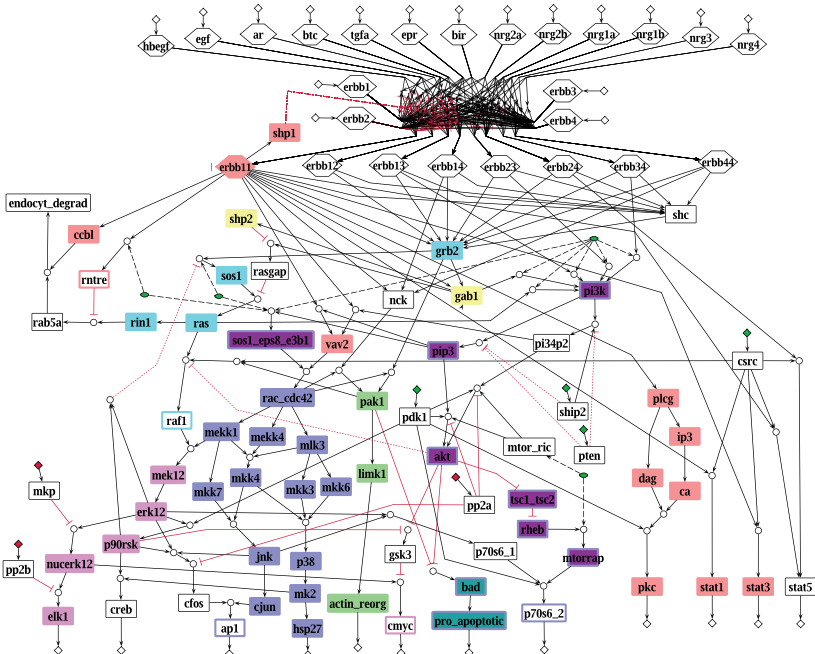


Figure 4.4: Species equivalence classes in the logical ERBB model. Each color represents one equivalence class. The equivalence classes found in model M1 are mapped on the border color, while equivalence classes found in model M2 are mapped on the node coloring. White-colored species are not part of any equivalence class. The states for the ligands and the four receptor monomers were left open, all other inputs were fixed to their default value, which is indicated by the red (0) and green (1) diamonds. Late events (dotted lines) were excluded. Figure adapted from Samaga et al. (2009).

Table 4.1: Species equivalence classes in the Logical ERBB models M1 and M2.

Species	Model M1	Model M2
tsc1_tsc2	1	1a
sos1_eps8_e3b1	1	1a
rheb	1	1a
pip3	1	1a
pi3k	1	1a
mtorrap	1	1a
akt	1	1a
rac_cdc42	1	1b
p38	1	1b
mlk3	1	1b
mkk7	1	1b
mkk6	1	1b
mkk4	1	1b
mkk3	1	1b
mk2	1	1b
mekk4	1	1b
mekk1	1	1b
jnk	1	1b
hsp27	1	1b
cjun	1	1b
pro_apoptotic	1	1c
bad	1	1c
ap1	1	-
p70s6_2	1	-
sos1	2	2
rin1	2	2
ras	2	2
grb2	2	2
raf1	2	-
vav2	3	3
stat3	3	3
stat1	3	3
shp1	3	3
plcg	3	3
pkc	3	3
ip3	3	3
erbb11	3	3
dag	3	3
ccb1	3	3
ca	3	3
rntr	3	-
pak1	4	4
limk1	4	4
actin_reorg	4	4
p90rsk	5	5
nucerk12	5	5
mek12	5	5
erk12	5	5
elk1	5	5
cmyc	5	-
shp2	6	6
gab1	6	6

PIP₃, PI3K, GAB1, or GRB2 suffices to activate ERK1/2 and AKT. Furthermore, six MISs of size 2 were found (Table 4.2). Strikingly, no MISs of higher cardinality exist for this problem. In model M2, 94 MISs up to size 3 exist that provoke the above mentioned response. The set {ERBB11 = 1} is the only MIS containing only one element; furthermore, 33 MISs of cardinality 2 (e.g., activating any of the other receptor dimers and VAV2), and 60 MISs of cardinality 3 were found.

These findings showed that the network has fragile points where a mutated protein, for example, one that is constitutively active, may support uncontrolled growth and proliferation. However, besides ERBB signaling, various other pathways are important for the regulation of growth and apoptosis and a failure in one pathway might be compensated by another, what makes it important to include these pathways step by step into our model. Additionally, when building up the model we did not focus on one certain cell type, but collected species and interactions that have been detected in different kinds of cells leading to a kind of “master model”. A model that describes only one cell type would probably include less interactions (Saez-Rodriguez et al. 2009), implying that a constitutive signal has not such a global, that is, network-wide, influence as in the master model.

Table 4.2: Minimal intervention sets of size 2 to activate ERK1/2 and AKT in the ERBB model.

<i>Intervention goal: erk12 = 1, akt = 1</i>							
<i>Side constraints: default values according to Appendix C.2</i>							
	sos1_eps8_e3b1	vav2	raf1	rac_cdc42	mekk1	mek12	pi34p2
1	1						1
2		1					1
3			1				1
4				1			1
5					1		1
6						1	1

Shown are the computed MISs of cardinality 2 to activate erk12 and akt without external stimuli. For MISs of cardinality 1 see text. An entry 1 means constitutive activation. Only early events have been considered. Interventions for species with fixed values and the target species erk12 and akt have been excluded.

In the following, we illustrate the power of the introduced reduction techniques for the computation of MISs (Section 4.2) by the above mentioned intervention problem in the ERBB network. As a second case study, we analyse another intervention problem in a logical model of T cell receptor signaling (Saez-Rodriguez et al. 2007).

For the T cell receptor model, we consider the intervention problem as described

in Saez-Rodriguez et al. (2007), hereafter referred to as the *TCR problem*: we search for MISs with a maximum cardinality of three that lead to sustained T cell activation without external stimuli. The target species (nine to be activated, three to be inhibited) cover the complete output layer of the model, that is, each non-output node influences at least one target node, whereas the three main inputs are set to 0 as side constraints. The intervention problem we consider for the ERBB model (the *ERBB problem*) is to activate AKT and ERK1/2 without external stimuli, as described above.

Table 4.3 shows the results of the reduction techniques together with the achieved reduction in computation time. We first consider the results when applying each reduction technique separately. All three techniques yield a decrease of the computation time by at least 49 % up to 99 %. In the ERBB problem, in contrast to the TCR problem, the target species AKT and ERK/2 are both situated in the intermediate layer of the ERBB network, and many species or even whole pathways do not affect them. According to these different structures of the intervention problems, the methods for search space reduction yield varying profit (Table 4.3). For the ERBB problem, exploiting the interdependencies in the interaction graph underlying the logical model (which is exactly the one analyzed in Section 3.3) is very beneficial: the number of possible single interventions can be reduced from 138 to 38 when only those single interventions are permitted that influence the target species properly (i.e., help activating AKT or/and ERK1/2). In comparison, in the TCR problem, dependency analysis reduces the 130 possible single interventions only by 36 down to 94. In both models, we find a similar number of fault equivalence classes each containing at least two equivalent interventions (19 in TCR, 20 in ERBB); however, the fault equivalence classes for ERBB are larger leading to a higher number of single interventions that can be excluded in the main algorithm. In most cases, the number of tested MIS candidates (for which the logical steady state was calculated) is higher for the TCR problem, although more single interventions are possible in ERBB. This is due to the fact that many MISs of cardinality 1 and 2 exist for the ERBB problem decreasing the search space considerably since supersets of these MISs need not be considered in later iterations. In contrast, no MIS with cardinality smaller than 3 exists in the TCR problem.

When combining the three reduction techniques, as already mentioned in Section 4.2.2.2, we first exploit the dependencies and then compute the fault equivalence classes in the reduced model providing the final set of permitted single interventions. During the main algorithm, the number of MIS candidates that are checked (i.e., for

Table 4.3: Benchmark tests showing the power of reduction techniques for the computation of minimal intervention sets.

	Naive algo- rithm	Exploiting			sDep and EqClass	Algorithm exploiting all
		sDep	EqClass	mDep		
ERBB problem						
Computation time [s]	750	4	307	257	2	2
(% of naive algorithm)	(100%)	(0.5%)	(41%)	(34%)	(0.3%)	(0.3%)
Checked candidates	302975	1951	122817	93483	970	960
Possible single interventions	138	38	105	138	32	32
Species that can						
be set to 1 and 0	69	11	42	69	10	10
be set to 1	0	15	14	0	11	11
be set to 0	0	1	7	0	1	1
Single interventions excluded						
by sDep	-	100	-	-	100	100
by EqClass (# EqClass)	-	-	33 (20)	-	6 (4)	6 (4)
Candidates excluded						
by mDep	-	-	-	209492	-	10
TCR problem						
Computation time [s]	655	249	212	336	91	66
(% naive algorithm)	(100%)	(38%)	(32%)	(51%)	(15%)	(10%)
Checked candidates	357890	135626	114577	169189	49061	34991
Possible single interventions	130	94	89	130	67	67
Species that can						
be set to 1 and 0	65	31	34	65	17	17
only be set to 1	0	27	15	0	26	26
only be set to 0	0	5	6	0	7	7
Single interventions excluded						
by sDep	-	36	-	-	36	36
by EqClass (# EqClass)	-	-	41 (19)	-	27 (13)	27 (13)
Candidates excluded						
by mDep	-	-	-	188701	-	14070

All computations were made on an Intel^(R)Core^(TM)2 Duo CPU E6850 at 3.00 GHz.

sDep, dependency analysis to exclude single interventions; mDep, dependency analysis to exclude minimal intervention set candidates not influencing all targets; EqClass, fault equivalence class. Table adapted from Samaga et al. (2010).

which the logical steady state is computed) is then further reduced by excluding those candidate MIS whose interventions do not influence *all* target species properly (see Section 4.2.2.3). Both problems benefit when incorporating the fault equivalence classes, for the TCR problem even more than for the ERBB problem, because in the latter many interventions were already excluded by dependency analysis. Exploiting the dependencies for multiple targets finally leads to an appreciable improvement only for the TCR problem—what can be explained by the fact that there are many target species in this setting.

In conclusion, the examples show that the methods for search space reduction lead to a considerable decrease of the computation time, whereby the actual benefit and the contribution of each of the three techniques is dependent on the structure of the intervention problem and the network structure itself.

4.4.3 Data Analysis

In Section 3.3.2, we confronted the interaction graph model of ERBB signaling with high-throughput experimental data from primary hepatocytes and HepG2 cells. Here, we will show the analysis of the same data sets based on the logical ERBB model as explained in Section 4.3.

As stated above, the comparison of experimental data with the binary network response of the logical model requires a discretization of the data to 0 (off) and 1 (on). To obtain the discretized values, we used DataRail, a MATLAB toolbox that facilitates the linkage of experimental data to mathematical models (Saez-Rodriguez et al. 2008). DataRail provides a variety of methods for data processing, including algorithms to convert continuous data into binary values and to create convenient data structures for the analysis in *CellNetAnalyzer*. The discretization depends on three thresholds p_1 , p_2 , p_3 . If all thresholds are exceeded, the measured signal is discretized to 1, otherwise to 0. The first threshold is for the *relative significance*, that is, the ratio between the value at time 1—in our case after 30 minutes—and the value at time 0. The second threshold ensures the *absolute significance*, that is, the ratio between the signal and the maximum value for this signal from all measurements. The third threshold ascertains that the signal is above experimental noise. The choice of the threshold is quite difficult as no reference data exist that define when a molecule is “on”, that is, when it is sufficiently activated to induce its downstream events. Most likely, the required level of activation differs from protein to protein and from cell to cell. However, since no information on these differences is available and to

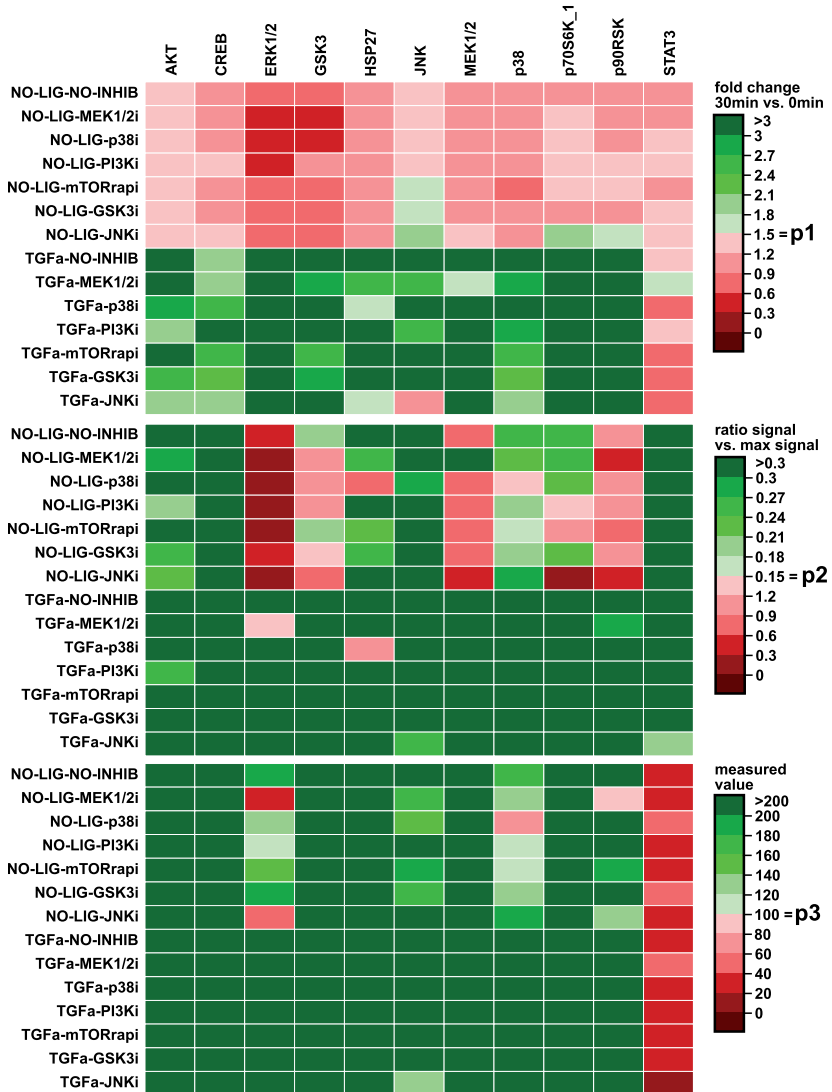


Figure 4.5: Sensitivities of the binarization to the chosen parameters (primary human hepatocytes). Upper panel: ratio between 30 and 0 minutes time point lies above (green) or below (red) chosen threshold for relative significance ($p_1 = 1.5$). Middle panel: ratio between signal and maximum value for this signal over all conditions lies above (green) or below (red) chosen threshold for absolute significance ($p_2 = 0.15$). Lower panel: measured signal lies above (green) or below (red) the chosen threshold for experimental noise ($p_3 = 100$). For all panels: the darker a field is colored, the larger the distance to the chosen threshold is, that is, the binarization is less sensitive on the parameter. Figure adapted from Samaga et al. (2009).

avoid unnecessary degrees of freedom, we decided to define the same threshold for all molecules and both cell types ($p_1 = 1.5$, $p_2 = 0.15$, $p_3 = 100$). Figure 4.5 shows the sensitivities of the binarization with respect to these three parameters for the data from primary human hepatocytes. The sensitivities for the HepG2 data are shown in Appendix D.

We computed for each measured scenario the binary network response of our model and compared it with the discretized data. All input species whose values were not specified in the particular scenario were set to their default values (see Appendix C.2), and late reactions with time variable $\tau = 2$ (see Figure 4.3 and Appendix C.2) were removed prior to the computation of the logical steady state.

As in the case of the dependency analysis, the measured data agreed reasonably well with the predictions of the model M1 (HepG2: 77% correct predictions; primary cells: 90% correct predictions; see Figure 4.6).

In Figure 4.7, the comparison of model M2 with the experimental data is shown. For primary cells, only 7% of the states could not be determined due to the ITT gates, for HepG2 21%. 83% of the predictions for primary cells and 59% for HepG2 cells were correct. In all cases where a state could be predicted by M2, it naturally coincided with the prediction from M1 since the latter is only one special case of all possible behaviors in model M2. In some cases where we used an ITT gate in model M2, the logical function could be uniquely determined with the experimental results confirming some of the deterministic logic gates used in model M1: for example, the transcription factor CREB can be activated through the MEK-dependent kinase p90RSK AND/OR through the p38-dependent MK2. As CREB was still activated both with MEK inhibitor and with p38 inhibitor, this pointed to an OR-connection achieving a match between model predictions and data in this node. In the same way, we could verify an AND connection for the two negative modulators of GSK3 and an OR for the phosphorylation of the auto-inhibitory domain of p70S6K.

Again, using ITT gates, we can only reflect uncertainties regarding the logical combination of different paths and not whether a species influences another at all. This is why some of the discrepancies between the predictions of model M1 and the data also appeared for model M2.

The data analysis with the logical model confirmed the following findings from the interaction graph analysis (Section 3.3.2):

- no activation of STAT3 by the tested ligands, both in HepG2 and in primary

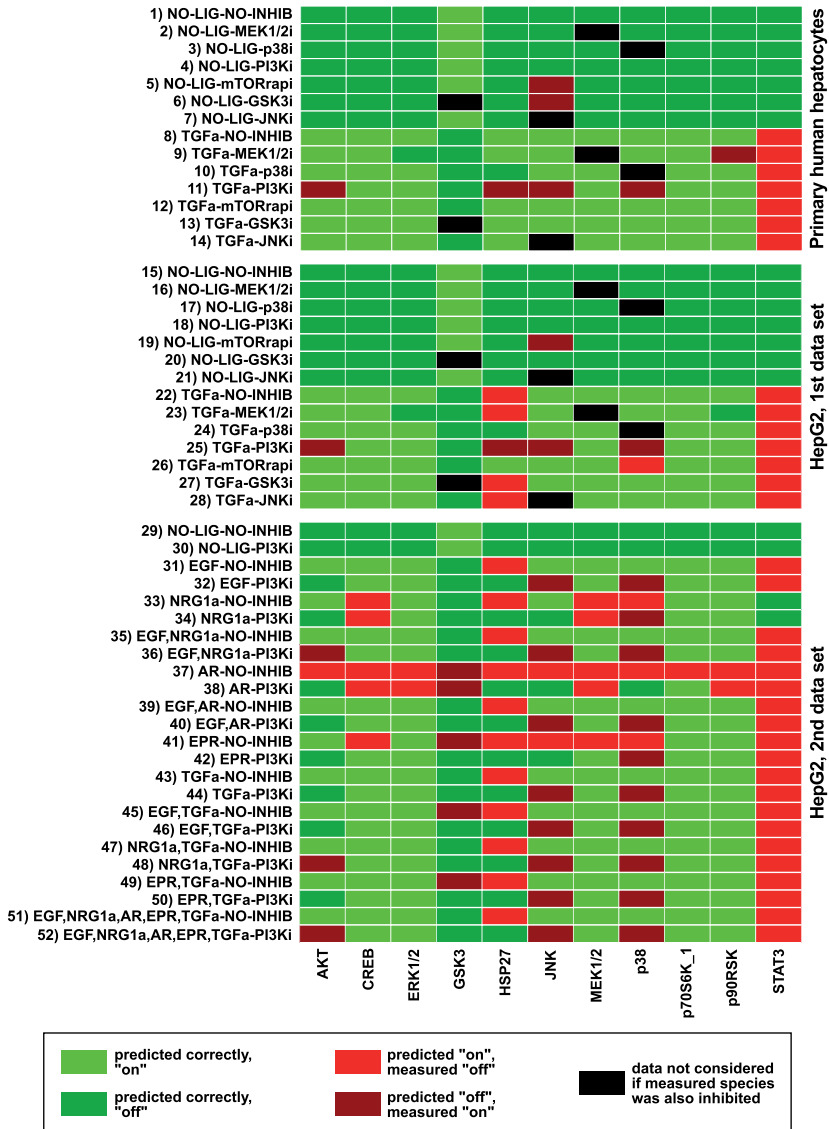


Figure 4.6: Comparison of the discretized data with predictions from the logical model. Each row represents one treatment, readouts are shown in the columns. Figure adapted from Samaga et al. (2009).

4 Logical Models of Signal Transduction Networks

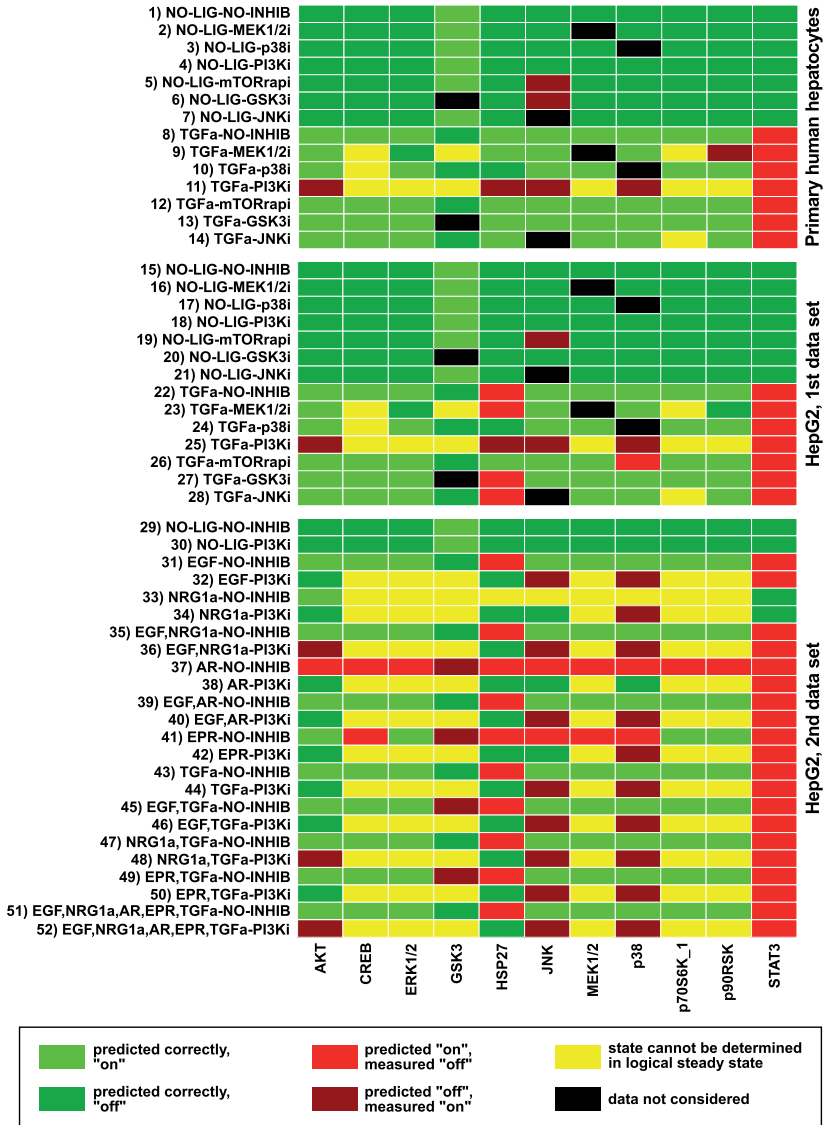


Figure 4.7: Comparison of the discretized data with predictions from model M2. Each row represents one treatment, readouts are shown in the columns. Figure adapted from Samaga et al. (2009).

cells,

- no activation of the measured proteins by amphiregulin in HepG2 cells (Figure 4.6, lines 37/38),
- no activation of HSP27 in HepG2 cells,
- activation of JNK (HepG2 and primary cells), p38 (HepG2 and primary cells), and HSP27 (primary cells) in presence of PI3K inhibitor.

In addition, the analysis with the logical model showed that phosphorylation of JNK in response to NRG1 α is—in contrast to the response to any of the other ligands—sensitive on PI3K inhibitor (Figure 4.6, line 34). This was also reflected in the analysis with the dependency matrix (although we did not explicitly state it), where an increase of NRG1 α only increased the phosphorylation of JNK in absence of PI3K inhibitor (Figure 3.7, lines 28–33) and decreasing the level of PI3K, that is, adding the inhibitor, after NRG1 α stimulation also lead to a decreased phosphorylation state of JNK (Figure 3.7, lines 52 and 59). Thus, NRG1 α must use a different, PI3K-dependent signaling path for activating JNK than the other ligands, probably due to the fact that NRG1 α only activates ERBB1/ERBB3 dimers, whereas the other tested ligands also bind to ERBB1 homodimers. Taking these findings together, we propose the following alternative mechanism: VAV2 is the major GEF for Rac/CDC42 in hepatocytes and activates Rac/CDC42 in a PI3K-independent way. NRG1 α , which cannot bind to ERBB1 homodimers and, accordingly, is not able to activate VAV2 (see Appendix C.2), provokes the activation of JNK independently of the Rac/CDC42-induced MAPK cascade through a different, PI3K-dependent pathway.

As already stated above, with the logical data analysis, some of the chosen logical gates could be confirmed or falsified. In contrast, there are also some findings that only showed up in the analysis with the interaction graph model. This is, for example, the case for the result that interactions on p38 are missing in the model or that the JNK inhibitor showed a minor specificity. Furthermore, additive effects when adding a second ligand—or a missing additive effect, as it was the case for adding TGF α and EGF—could only be detected with the interaction graph. The analysis with the interaction graph model revealed that the proposed mechanism of GSK3 phosphorylation was not in accordance with the data. This could not be detected with the logical model: in the model, we assumed that GSK3 is phosphorylated by AKT or p90RSK, independently of each other (see interaction 176 in Appendix C.2).

The experimental data showed that GSK3 is still phosphorylated when blocking either p90RSK (MEK1/2 inhibitor) or AKT (PI3K inhibitor), what was in accordance with the model predictions. For the interaction graph based analysis, we expected a decreased signal for both single inhibitor treatments; as this was not observed, we concluded that GSK3 phosphorylation in response to TGF α must be triggered by another (unknown) player (see Table 3.2). However, the possibility that inhibiting one upstream species is compensated by another species cannot be ruled out. This example shows that the different modeling formalisms enable different views on the signaling system that should be combined in order to come closer to the truth. In this particular example, combined inhibitor treatments would be necessary to disentangle the phosphorylation mechanism.

A list of proposed model changes considering both the results from the analysis with the logical model and the interaction graph is given in Table 4.4. Changing the logical model accordingly, we could improve the agreement of model predictions and data in the logical analysis from 90 % to 97 % for the primary cells and from 74 % to 94 % for HepG2 cells.

4.5 Discussion

Logical modeling of biological systems has a long tradition (Kauffman 1969) and several logical models of gene regulatory and signaling networks showed the value of this modeling formalism for systems biology (Albert and Othmer 2003; Christensen et al. 2009; Giacomantonio and Goodhill 2010; Huard et al. 2012; Li et al. 2004; Naldi et al. 2010; Ryll et al. 2011; Schlatter et al. 2009; Zhang et al. 2008). In addition to methods that enable dynamic simulations of logical networks, static methods have been developed, which are particularly suited for the analysis of large-scale networks. In this chapter, we presented some extensions of the static formalism introduced by Klamt et al. (2006).

With species equivalence classes, we introduced a new technique for the analysis of logical networks. These classes reveal coupled activation patterns in the logical model, thus providing valuable insights into the correlated behavior of network elements. As a possible future development, the concept of species equivalence classes might provide a basis for model reduction.

Minimal intervention sets (MISs) can be used to gain important insights into signaling networks. They can be used for drug target identification, to design experiments leading to a desired result, or for diagnosis, that is, for generating hy-

Table 4.4: Proposed ERBB model changes to improve agreement between model structure and experimental data.

	Primary human hepatocytes	HepG2 Cells
Removed interactions		
erbb11 · csrc → stat3	x ^{(1),(2)}	x ^{(1),(2)}
erbb11 · pip3 → vav2	x ^{(1),(2)}	x ^{(1),(2)}
erbb11 · pi34p2 → vav2	x ^{(1),(2)}	x ^{(1),(2)}
pip3 → sos1_eps8_e3b1	x ^{(1),(2)}	x ^{(1),(2)}
!p90rsk · !akt → gsk3	x ⁽¹⁾	x ⁽¹⁾
mk2 → hsp27		x ^{(1),(2)}
mk2 → creb	x ⁽¹⁾	x ⁽¹⁾
p90rsk → creb	x ⁽¹⁾	
Added interactions		
mtor_rap → p70s6k_1	x ⁽¹⁾	x ⁽¹⁾
erbb11 → vav2	x ^{(1),(2)}	x ^{(1),(2)}
!pkc → gsk3	x ⁽¹⁾	x ⁽¹⁾
tgfa → creb	x ⁽¹⁾	x ⁽¹⁾
pi3k → jnk		x ⁽²⁾

AND gates are depicted by a dot (A · B refers to AND(A,B)), NOT operators by an exclamation mark (!A means NOT(A)). An x in the respective column indicates that the change is proposed for the model describing the indicated cell type. Entries marked with (1) denote findings derived with interaction graph analysis, entries marked with (2) findings derived with the logical model.

potheses that explain experimental data in the context of a given network (Franke et al. 2008; Saez-Rodriguez et al. 2007). Due to the complex combinatorial problem, determination of MISs is computationally challenging, what calls for algorithmic improvements. We introduced three search space reduction techniques, which may significantly speed up the calculation procedure. As the benchmark demonstrated, the achievable performance gain and the contributions of the different reduction techniques depends on the actual intervention problem and network size. The computation times for typical intervention problems show that realistic problems can be treated in reasonable time. With fault equivalence classes, we adopted a well-known technique from electrical engineering illustrating relationships between electrical circuits and cellular networks with information flow. The concept of fault equivalence classes is not only relevant for computing MISs: these classes also reveal sets of failures in the network that are indistinguishable with a given set of readouts. The two other reduction methods are based on the dependency matrix, again showing the importance of this concept for elucidating network-wide activatory and inhibitory interdependencies (cf. Section 3.1).

With the advances of experimental techniques, it becomes more and more essential to provide tools that allow for the analysis and exemplification of the huge amount of data that arise. Here, we showed how the predicted binary network response derived from a logical model can be compared to experimental data in an automated way. This method is well suited to analyze large data sets, in particular stimulus–response data that stem from combinatorial experiments, that is, where cells were treated with systematic combinations of different ligands and inhibitors.

The newly developed methods were applied to a large-scale logical model of signaling through the four ERBB receptors. The underlying interaction graph of this model was studied in Section 3.3. Our ERBB model is based on the stoichiometric pathway map of Oda et al. (2005), and we proposed a general guideline how to deal with technical problems that arise when converting a stoichiometric model into a logical one.

One possibility to deal with uncertainties concerning the correct logical combination of different influences on a certain node is the usage of gates with incomplete truth tables (ITT gates). Thus, we replaced the deterministic logical gates for the activation of 14 species of our model with ITT gates and repeated all logical analyses with this modified model. Surprisingly, the predictive power of the model with these ITT gates is still high, highlighting the redundant structure of major parts of the signaling pathway and showing that many properties of the network do not rely

on the assumptions we made when choosing the logical functions. These redundant structures also became apparent by computing the species equivalence classes.

The analysis of high-throughput experimental data from primary human hepatocytes and the HepG2 cell line confirmed some of the findings derived with the interaction-graph-based analyses (Chapter 3). However, some disagreements between model structure and experimental results only showed up in the analysis with the interaction graph model, while for other aspects, the logical analysis was more revealing. This shows that different modeling formalisms can complement each other, and that it is indispensable to investigate a system under study from different perspectives in order to get a holistic understanding.

5 | From Structure to Dynamics: Combined Interaction Graph and ODE Modeling Unravels Network Structure of HGF Signaling

As already demonstrated in the previous chapters, signaling pathways cannot simply be modeled as linear chains, but involve extensive crosstalk, feedforward, and feedback loops resulting in complex signaling networks. Network topologies are often context specific and altered in diseases (Pawson and Warner 2007), and the knowledge of cell-context-specific feedback and crosstalk mechanisms is, for example, important to gain insights into mechanisms that cause undesired effects in targeted therapy. Considering different potential crosstalk, feedforward, and feedback mechanisms easily results in an enormous number of possible network structures. Thus, due to the high combinatorial complexity, there is the need for systematic methods facilitating an unbiased identification of the network wiring.

In the previous chapters, we presented computational methods to infer and analyze signaling networks. These methods are based on qualitative modeling formalisms that can deal with large networks, but are limited in their capacity to describe dynamic properties such as signal duration or amplitude. Dynamic modeling approaches using coupled ordinary differential equations (ODEs) are well suited to analyze the impact of crosstalk and feedback regulation and allow quantitative insights. However, consideration and systematic analysis of several potential mechanisms results in a high combinatorial complexity with many degrees of freedom and is therefore often not feasible with ODEs. Hence, it is desirable to exploit the advantages of both qualitative and quantitative modeling and to develop strategies to combine both approaches.

Here, we present a hybrid approach (Figure 5.1) which combines qualitative and quantitative modeling techniques. The approach is introduced using the example of HGF-induced activation of MAPK and PI3K signaling in primary mouse hepatocytes (see Section 2.3.2). First, we built an interaction graph master model containing

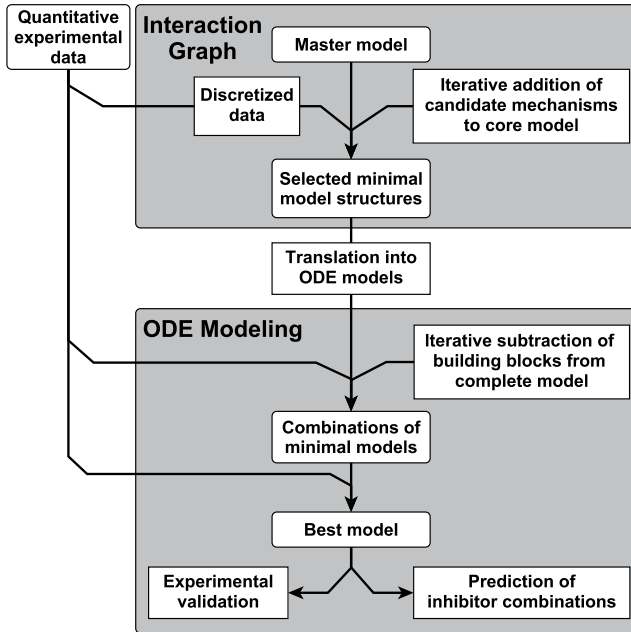


Figure 5.1: Workflow of model selection strategy. Quantitative time-resolved data is discretized for the selection of submodels of the interaction graph master model. The interaction graph master model is based on literature knowledge and consists of the *core model* and reported interactions between the signaling pathways of interest, the *candidate mechanisms*. Ordinary differential equation (ODE) modeling, based on preselected model structures, utilizes the entire information of the time-resolved data. Model selection based on parameter estimation permits the selection of the best model structure. Figure adapted from D’Alessandro, Samaga, Maiwald et al. (2015).

previously reported crosstalk, feedback, and feedforward mechanisms (Section 5.1). Based on experimental data (Section 5.2), we selected minimal substructures of this master model that can explain the qualitative characteristics of the data (Section 5.3). In this way, the search space of model candidates was vastly reduced. Subsequently, we translated the selected minimal model structures into ODE models (Section 5.4) and identified the model structure representing the experimental system best (Section 5.5). The identified model structure was experimentally validated by verifying the presence of a particular negative crosstalk mechanism in the cellular model system under study (Section 5.6.1). To further validate the model structure, we analyzed how the system responds to different perturbation conditions (Section 5.6.2).

The work presented in this chapter has been published in D'Alessandro, Samaga, Maiwald et al. (2015) and results from a collaboration with co-workers of the groups of Ursula Klingmüller (DKFZ Heidelberg) and Jens Timmer (University of Freiburg). Within this thesis, we focus on the qualitative modeling part of the project. The author reconstructed the HGF interaction graph master model from literature information and proposed and developed the method to select minimal model structures from the master model as model candidates for subsequent ODE modeling. This also included discretization of experimental data and the development of rules how to derive an ODE model reflecting a given interaction graph model structure. The experiments were conducted by Lorenza D'Alessandro (DKFZ Heidelberg), and Tim Maiwald (University of Freiburg) was responsible for the ODE modeling part. Thus, the detailed descriptions of the experimental data and of the ODE modeling are not included in this work, but can be found in D'Alessandro, Samaga, Maiwald et al. (2015).

5.1 An Interaction Graph Model of HGF Signaling

As discussed in previous sections, interaction graphs as one example for qualitative models can be used to make predictions on the possible qualitative behavior of a signaling network, and these predictions can be compared with experimental data. Resulting inconsistencies between data and network structure provide then a basis to identify missing or inactive interactions in the network structure (see Chapter 3). In order to describe crosstalk mechanisms, interaction graphs take preference over logical models: While the latter are well suited for studying the input–output behavior of large signaling pathways (see Chapter 4), an appropriate description of crosstalk mechanisms within the logical formalism is often difficult. This is due

to the fact that, in contrast to the main activation routes of a signaling pathway, crosstalk mechanisms typically enhance or downgrade certain effects, rather than being necessary for, for example, activation of a certain protein, or completely blocking a species' activity.

To unravel the crosstalk between the MAPK and PI3K signaling pathways upon HGF stimulation in hepatocytes, we built an interaction graph model of HGF-induced activation of these pathways based on literature knowledge (Figure 5.2). We distinguish between interactions that form the main activation routes (*core model*; black edges in Figure 5.2) and interactions describing feedback, feedforward, and crosstalk mechanisms (*candidate mechanisms*; turquoise edges) that have been reported for special cell types or under certain experimental conditions and whose presence in hepatocytes is uncertain (Appendix E). The full graph including the core model and all candidate mechanisms is considered as a non-cell-type-specific *HGF master model*.

5.2 Experimental Data

To analyze which of the candidate mechanisms are active in our cellular system, primary mouse hepatocytes were treated with HGF in combination with specific Met inhibitor (PHA 665752), MEK inhibitor (U0126), PI3K inhibitors (LY294002, Wortmannin, and PI-103), and PDK1 inhibitor (BX-912), alone or in combination. Additionally, siRNA targeting AKT and ERK1/2 was employed. The phosphorylation of AKT, MEK1/2, ERK1/2, and p90RSK was analyzed by time-resolved quantitative immunoblotting and by protein array (D'Alessandro, Samaga, Maiwald et al. 2015). Activation of p90RSK occurs in two-steps, and we denote with RSK_s the form of p90RSK phosphorylated at a single residue, while RSK_d refers to the double phosphorylated, fully active form (Appendix E). The dynamics of SOS1 activation upon MEK inhibitor treatment were also measured (data shown in D'Alessandro, Samaga, Maiwald et al. (2015)). We calculated the fold change of protein phosphorylation for each treatment condition in comparison to the respective control and between treatment conditions (Figure 5.3). For SOS1 activation, the band shift was quantified as an indicator for its activation status. Treatment with PI3K inhibitor and with siRNA targeting AKT had to be excluded from our study due to the high variability observed in our experimental results (D'Alessandro, Samaga, Maiwald et al. 2015).

To relate the data to the interaction graph model (Figure 5.2), we discretized the measured responses to *increase*, *decrease*, and *not measured / not conclusive* (Figure 5.4): If a certain effect was consistently observed for the replicates within the same time

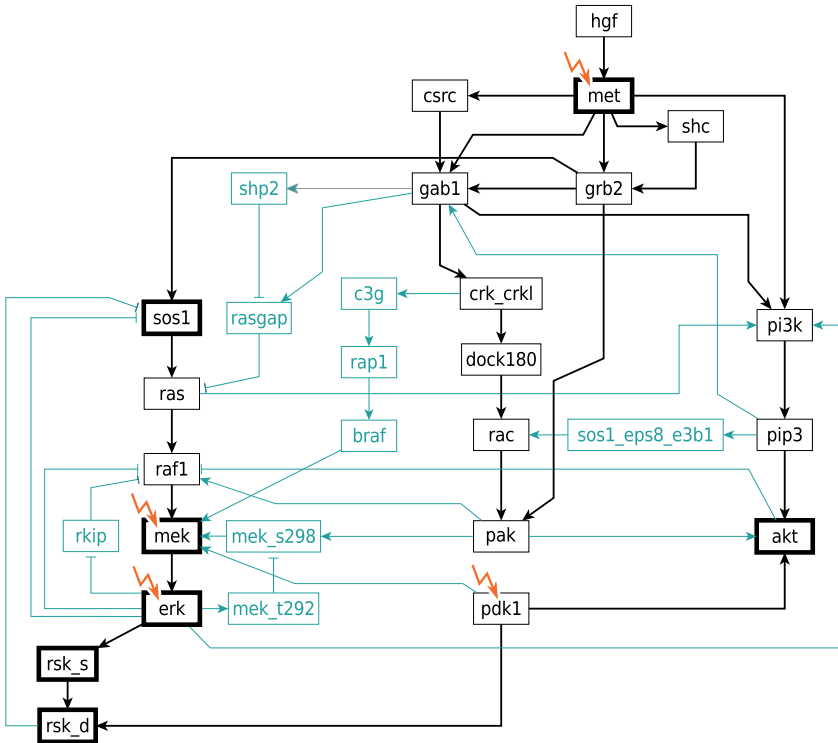


Figure 5.2: Interaction graph master model. The interaction graph master model was built from literature information. Detailed model documentation can be found in Appendix E. The core model is given in black, candidate mechanisms are depicted in turquoise. Arrows represent activating (positive) interactions, blunt-ended lines indicate inhibitory (negative) interactions. The measured species are marked with bold borders. The lightning symbol indicates that the respective species was experimentally targeted with a chemical inhibitor or siRNA. Figure adapted from D’Alessandro, Samaga, Maiwald et al. (2015).

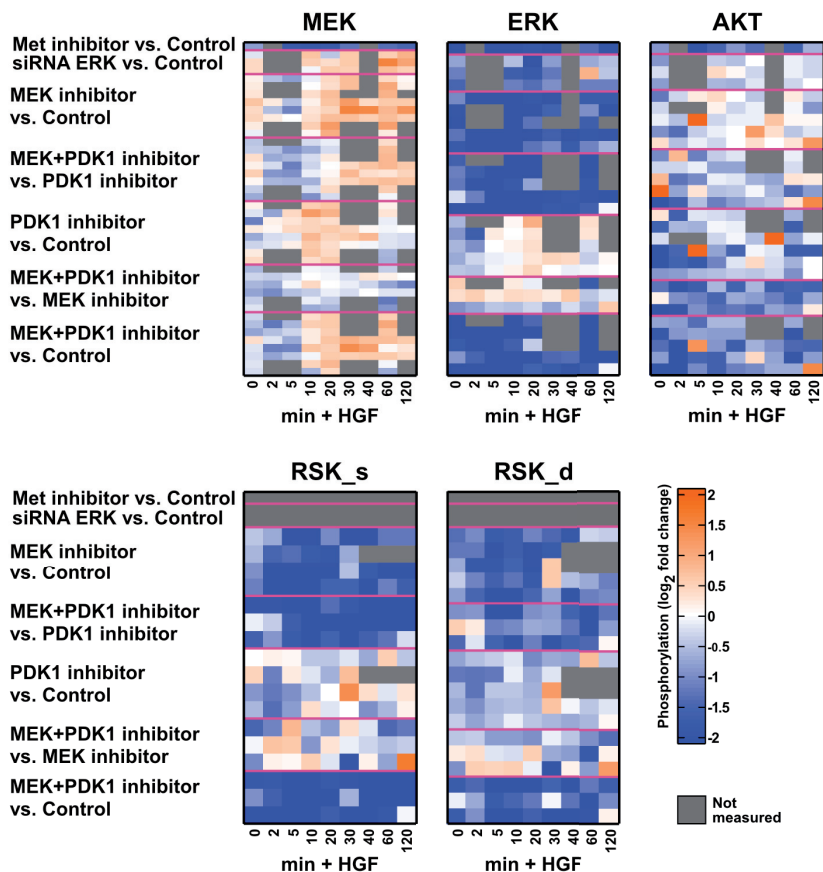


Figure 5.3: Experimental results. For each indicated protein, the fold change of the phosphorylation state measured by quantitative immunoblotting of two different experimental conditions is shown on a logarithmic scale at the indicated time points after HGF stimulation. Each row refers to one experiment; same experimental conditions are grouped with magenta lines. Figure adapted from D’Alessandro, Samaga, Maiwald et al. (2015).

points, it was included as an observed effect in the scheme of the discretized data. Otherwise, the effect was considered *not conclusive* and not taken into account. The fourth possible value *no change* was not observed in our data set. Additionally, we included the timing of the response in the discretized data, classifying an observed effect as an early or late response. *Early* refers to the initial qualitative response of a node after HGF stimulation. A response was termed *late* if the qualitative behavior at any successive time point was different from the initial (early) response. Thus, the late response characterizes the first effect that is opposite to the early response, or indicates that the qualitative response is identical for all time points.

	MEK	ERK	AKT	RSK_s	RSK_d	SOS1	
Met inhibitor vs. Control	↓↓↓	↓↓↓	↓↓↓	///	///	///	Timing: early late
siRNA ERK vs. Control	///↑	///↓	///	///	///	///	
MEK inhibitor vs. Control	↓↓↑	↓↓↓	///	↓↓↓	↓↓↓	↑↑↑	Direction: ↑ Increase ↓ Decrease / Not measured or not conclusive
MEK+PDK1 inhibitor vs. PDK1 inhibitor	↓↓↑	↓↓↓	↓↓↓	↓↓↓	↓↓↓	///	
PDK1 inhibitor vs. Control	↓↓↑	↑↑↑	↓↓↓	↓↑	↓↓↓	///	
MEK+PDK1 inhibitor vs. MEK inhibitor	↓↓↓	///	↓↓↓	///	///	///	
MEK+PDK1 inhibitor vs. Control	↓↓↓	↓↓↓	↓↓↓	↓↓↓	↓↓↓	///	
MEK+PDK1 inhibitor vs. Control	↓↓↓	↓↓↓	↓↓↓	↓↓↓	↓↓↓	///	

Figure 5.4: Discretized experimental data. The discretization is based on Figure 5.3 and additional SOS1 measurements shown in D’Alessandro, Samaga, Maiwald et al. (2015). A slash shows that no measurements were taken, or that the response was not conclusive. Figure adapted from D’Alessandro, Samaga, Maiwald et al. (2015).

5.3 Selection of Minimal Model Structures

Given an interaction graph model, we can predict the possible qualitative response of the considered proteins for the given experimental conditions using the concept of the dependency matrix (see Section 3.1). Exploitation of the dependency matrix is well suited for the analysis of transient effects—in contrast to the methods introduced in Section 3.2, which rely on the concept of sign consistency and require steady-state shift experiments. A comparison between model predictions and discretized data tells us whether a given model structure is able to reflect the experimental results. We used the discretized fold change of two experimental conditions, C1 and C2. Condition C1 represents the control condition, whereas in condition C2,

one or two additional inhibitors compared to C1 were applied (Figure 5.4). For each comparison, we checked in the dependency matrix of the model how the additional inhibitors in C2 act on the measured species leading to qualitative predictions for the early (initial) and late response. The detailed description how to derive these qualitative predictions is given in Section 3.1.

The effect of some inhibitors can often not be mimicked by perturbing a certain node in the network. The applied MEK inhibitor, for example, blocks MEK kinase activity, thus inhibiting the outgoing edges from MEK, but not the activation state of MEK. Therefore, we introduce a “dummy node” that is activated by MEK and itself activates all downstream nodes of MEK (in our particular case only ERK). The effect of the MEK inhibitor is then reflected by this dummy node.

Figure 5.5 shows the possible qualitative responses predicted from the interaction graph core model. Comparing model predictions with the discretized data, the majority of observed behaviors were not represented by the core model. Hence, we concluded that some of the candidate mechanisms must be active in primary mouse hepatocytes. To identify minimal substructures of the HGF master model consistent with the data, we added single candidate mechanisms and combinations thereof to the core model and tested the resulting model structures for their ability to represent the observed effects. As the master model was consistent with the experimental data, it was guaranteed that consistent submodels exist—at least one “submodel”, the master model itself. In order to select all minimal submodels of the interaction graph master model that can explain the observed effects from the experimental data (Figure 5.4) and that contain the core model, we proceeded as follows: starting from the core model, we added one candidate mechanism at a time (that is, all edges making up the mechanism, see Appendix E.3) and derived the model predictions for the resulting interaction graph structure as explained above. If all experimental observations were in accordance with these predictions, the structure was added to the list of selected model structures. If not, we combined the respective structure with other candidate mechanisms whose sole addition to the core model was not able to explain the data. Again, we checked the ability of each resulting model structure to explain the data and either considered the structure as selected minimal model or added further candidate mechanisms. We ensured that only *minimal* model structures were selected, that is, no submodel of a selected model can explain the data. Furthermore, the procedure assures that all minimal model structures that are able to explain the data are found. The described algorithm was implemented as a MATLAB script making use of API functions of *CellNetAnalyzer* (Klamt et al. 2007).

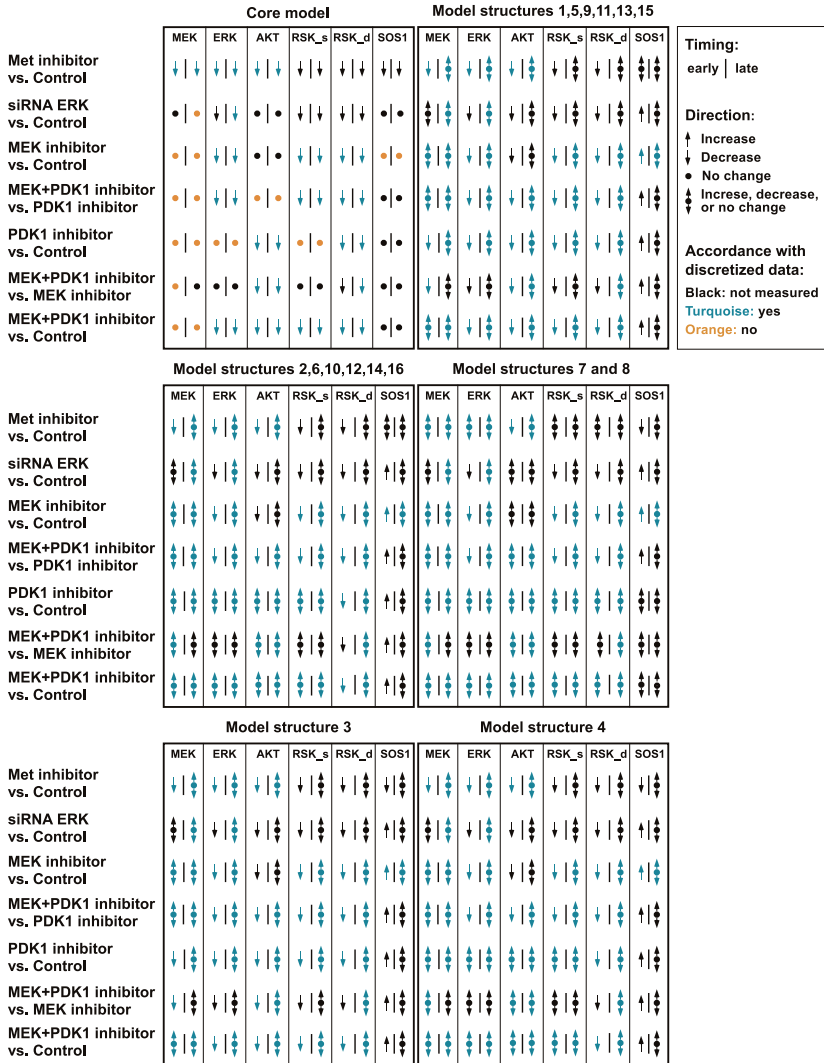


Figure 5.5: Predictions by interaction graph models. Predictions from the core model and from the selected minimal model structures (Figure 5.8) are shown. Arrow pointing up/down: the inhibition can only cause an increased/decreased activation of the measured protein. Filled circle: the inhibition does not affect the measured protein. Combined up/down arrow and filled circle: the model does not restrict the response to the inhibition. Figure adapted from D'Alessandro, Samaga, Maiwald et al. (2015).

The pseudocodes of the algorithm are given in Figures 5.6 and 5.7.

With the described procedure, we selected all possible model structures, each being composed of the core model and a building block, which is a subset of the candidate mechanisms. Thus, among the possible candidate edges that may be relevant for the other cell types and conditions, we selected candidate mechanisms that are specific for the HGF-induced MAPK and PI3K signaling pathways in primary mouse hepatocyte. Candidate edges that were not selected by our approach might play a role in other cell types and for other conditions. In total, we identified 16 minimal model structures that can equally well explain our experimental data (Figure 5.5). Notably, in some cases, the qualitative response is not restricted by the model structure: if a path between the inhibited and measured species includes a negative feedback loop, or if paths of both signs exist between the two nodes, the actual response depends on the strength of the different mechanism and, thus, cannot be predicted by a purely qualitative model.

To develop corresponding ODE models with an appropriate number of parameters, we compressed all selected minimal model structures by removing parallel mechanisms with the same sign and by compressing linear chains (cf. compression techniques given in Appendix B). Each candidate mechanism is represented by a single edge. Hence, each selected minimal model structure is composed of the compressed core model and a combination of candidate edges, each corresponding to one candidate mechanism. We define as *building block* the characteristic combination of candidate edges of a selected minimal model structure. Figure 5.8 shows the compressed 16 selected minimal model structures, the compressed core model and a model containing all building blocks (*complete model*). In contrast to the master model, the complete model contains only those candidate mechanisms that are included in at least one selected minimal model structure. The following candidate mechanisms were not selected: (i) the negative effect of GAB1 on Ras via RasGAP, (ii) the negative effect of ERK on Raf1, (iii) the negative effect of ERK on MEK, and (iv) the positive effect of PAK on Akt (all candidates mechanisms are described in detail in Appendix E.3). By comparing the 16 minimal model structures, we observed that all candidate models include (i) the edge from ERK to PI3K, (ii) a negative feedback from ERK to SOS1, either directly or indirectly via RSK_d, (iii) a positive route from ERK to MEK through various mechanisms, and (iv) a positive route from PDK1 to MEK. The latter route is included as a direct edge from PDK1 to MEK in all models but model 8, whereas model 8 contains a longer path via RSK_d, Ras, and AKT.

```

Function BuildingBlocks  $\leftarrow$  SelectModels (DataEarly,DataLate,Inhibitors,CoreModel,CandEdges)
inputs:
  DataEarly: for each measured protein, discretized fold change of different pairwise comparisons of
  experiments, early response
  DataLate: for each measured protein, discretized fold change of different pairwise comparisons of
  experiments, late response
  Inhibitors: for each pairwise comparison of experiments, applied inhibitor(s) making up the difference of the
  two experiments
  CoreModel: Interaction graph core model in CellNetAnalyzer format
  CandEdges: candidate edges
output:
  BuildingBlocks: building blocks of selected minimal model structures

/* compute dependency matrix of core model */
DepMat  $\leftarrow$  CNAcomputeDepMat (CoreModel);

call DepMatCompare (DataEarly,DataLate,Inhibitors,DepMat) returning Mismatch;
if Mismatch = 0 then
  | print 'Core model in accordance with data';
  | return;
end
BuildingBlocks  $\leftarrow$  empty set;
for i from 1 to number of CandEdges do
  | [Combi_1, ..., Combi_K]  $\leftarrow$  all i-combinations of CandEdges;
  | for j from 1 to K do
  | | if Combi_j is not superset of an element of BuildingBlocks then
  | | | /* compute dependency matrix of core model plus edges from Combi_j */
  | | | Model  $\leftarrow$  CoreModel plus edges from Combi_j;
  | | | DepMat  $\leftarrow$  CNAcomputeDepMat (Model);
  | | | call DepMatCompare (DataEarly,DataLate,Inhibitors,DepMat) returning Mismatch;
  | | | if Mismatch = 0 then
  | | | | append Combi_j to BuildingBlocks;
  | | | end
  | | end
  | end
end

```

Figure 5.6: Pseudocode for the selection of minimal model structures from the interaction graph master model. The described function *SelectModels* returns the building blocks of the selected minimal model structures. During the computation, two subroutines are called: *CNAcomputeDepMat*, a *CellNetAnalyzer* API function (Klamt and von Kamp 2011) that computes the dependency matrix of a given model structure, and *DepMatCompare* (pseudocode given in Figure 5.7), which compares the model predictions derived from the dependency matrix with the discretized experimental data. Figure adapted from D’Alessandro, Samaga, Maiwald et al. (2015).

```

Function Mismatch  $\leftarrow$  DepMatCompare (DataEarly,DataLate,Inhibitors,DepMat,MeasuredSpecies,Comparisons)

inputs:
DataEarly: for each measured protein, discretized fold change of different pairwise comparisons of experiments,
early response
DataLate: for each measured protein, discretized fold change of different pairwise comparisons of experiments, late
response
Inhibitors: for each pairwise comparison of experiments, applied inhibitor(s) making up the difference of the two
experiments
DepMat: dependency matrix
MeasuredSpecies: measured species
Comparisons: pairwise comparisons of experiments

output:
Mismatch: number of mismatches between model predictions and discretized data

for  $i$  from 1 to number of MeasuredSpecies do
  for  $j$  from 1 to number of Comparisons do
    if all Inhibitors of Comparisons( $j$ ) are weak/strong activators for MeasuredSpecies( $i$ ) then /* this
information is taken from DepMat */
      | ModelPredictionsEarly( $i, j$ )  $\leftarrow$  decrease;
    else if all Inhibitors of Comparisons( $j$ ) are weak/strong inhibitors for MeasuredSpecies( $i$ ) then
      | ModelPredictionsEarly( $i, j$ )  $\leftarrow$  increase;
    else if all Inhibitors of Comparisons( $j$ ) are neutral factors for MeasuredSpecies( $i$ ) then
      | ModelPredictionsEarly( $i, j$ )  $\leftarrow$  no change;
    else
      | ModelPredictionsEarly( $i, j$ )  $\leftarrow$  increase, decrease, or no change;
    end
    if all Inhibitors of Comparisons( $j$ ) are strong activators for MeasuredSpecies( $i$ ) then
      | ModelPredictionsLate( $i, j$ )  $\leftarrow$  decrease;
    else if all Inhibitors of Comparisons( $j$ ) are strong inhibitors for MeasuredSpecies( $i$ ) then
      | ModelPredictionsLate( $i, j$ )  $\leftarrow$  increase;
    else if all Inhibitors of Comparisons( $j$ ) are neutral factors for MeasuredSpecies( $i$ ) then
      | ModelPredictionsLate( $i, j$ )  $\leftarrow$  no change;
    else
      | ModelPredictionsLate( $i, j$ )  $\leftarrow$  increase, decrease, or no change;
    end
  end
end
Mismatch  $\leftarrow$  count mismatches between DataEarly and ModelPredictionsEarly and DataLate and
ModelPredictionsLate;

```

Figure 5.7: Pseudocode for the comparison of model predictions with discretized data. This function is called during the computation of minimal model structures from an interaction graph master model (see Figure 5.6). It returns the number of mismatches between discretized experimental data and predictions derived from the dependency matrix of a particular interaction graph model structure. Figure adapted from D’Alessandro, Samaga, Maiwald et al. (2015).

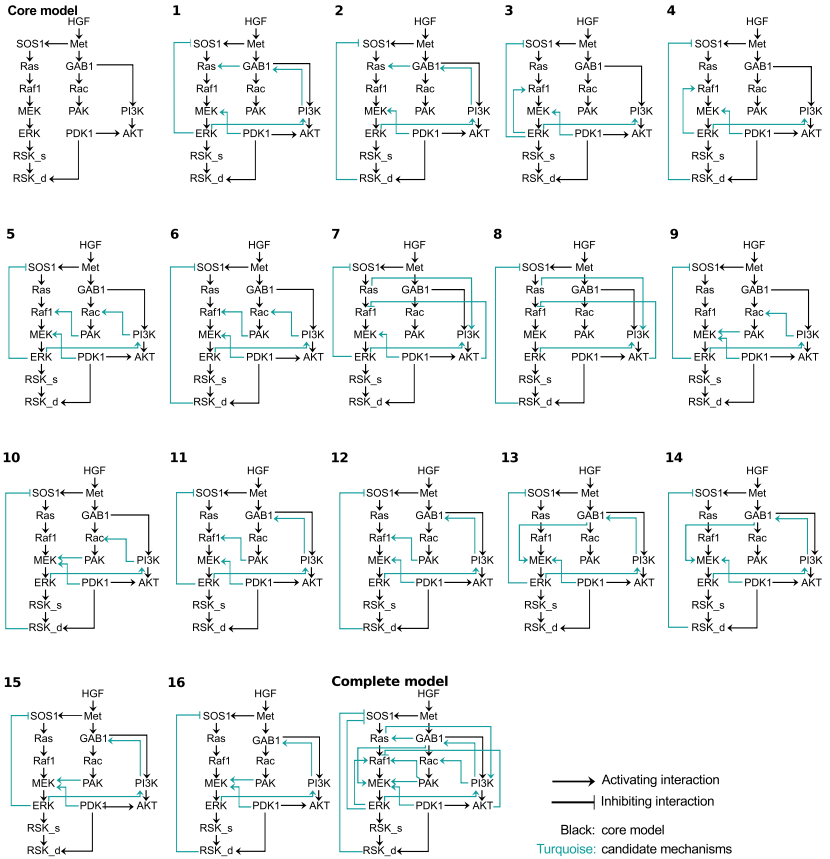


Figure 5.8: Selected minimal model structures, core and complete model. The compressed 16 selected minimal model structures that can explain the discretized data (Figure 5.5) are shown. In addition, the complete model structure (that is, the union of models 1–16) and the compressed core model structure are displayed. Arrows represent activating (positive) interactions; blunt-ended lines indicate inhibitory (negative) interactions. In each model, the core model is colored black, while the building block (the set of added candidate mechanisms) is shown in turquoise. Figure adapted from D’Alessandro, Samaga, Maiwald et al. (2015).

5.4 Translation into Ordinary Differential Equation Models

To test which of the identified structures can quantitatively represent the transient and sustained effects observed in the experimental data, we built for each of the 16 compressed selected minimal model structures as well as the core and the complete model a corresponding ODE model: the underlying interaction graph of the ODE model, which represents the sign structure of the Jacobian matrix (see Section 2.4.3.2), should reflect the respective minimal model structure. To this end, we built the ODE models from the interaction graphs in the following way: For the core model, each edge $pA \rightarrow pB$ from the interaction graph, where pA denotes the activated form of a species A , and pB the activated form of a species B , gives rise to a reaction describing the formation of pB . With mass action kinetics and assuming pA catalyzes the formation of pB , but is not consumed, we got the reaction $B \rightarrow pB$ with kinetic rate law $k_1 [pA] [B]$, where k_1 denotes a kinetic parameter for this reaction, $[pA]$ the concentration of pA , and $[B]$ the concentration of B . In addition, we introduced the deactivation (for example, dephosphorylation) reaction $pB \rightarrow B$ with kinetic rate law $k_2 [pB]$. An exception from this rule is the treatment of the species RSK. The interaction graph model describes activation of RSK as a two-step process: $RSK \rightarrow RSK_s \rightarrow RSK_d$. In this case, we assumed that both activated forms (RSK_s and RSK_d) are deactivated to RSK. For the Met receptor, an additional formation and degradation reaction was necessary to reflect the observed receptor dynamics. PDK1 is considered to be constitutively active (Casamayor et al. 1999; Wick et al. 2002) and was thus included as a constant parameter in the ODE model. To incorporate the effect of inhibitor treatments into the model, inhibitor parameters for PDK1, PI3K, Met, and MEK were introduced. These parameters allow for a potential reduction of the kinetic parameters within a range of 0% to 100%. Furthermore, these parameters are coupled to binary switches, allowing them to be activated only in their respective condition.

For each candidate edge, we added one additional reaction to the ODE model. We distinguished candidate mechanisms that provide an alternative, independent way of activation of the respective species, and those that influence the core activation mechanism. An example for the first is the activation of PI3K by Ras, which is described as $PI3K \rightarrow PI3K_active$ with rate law $k_a [Ras_active] [PI3K]$. An example for the latter is the effect of PAK on Raf1, which is assumed to promote Raf1 activation through Ras, and which is thus included as $Raf1 \rightarrow pRaf1$ with rate law $k_b [Ras_active] [pPAK] [Raf1]$ in the ODE model. Translation of inhibiting edges of

the interaction graph follows the same rules. As an example, deactivation of Raf1 by AKT is described as $pRaf1 \rightarrow Raf1$ with rate law $k_c [pAKT] [pRaf1]$. A list of all ODE reactions is given in Appendix F.

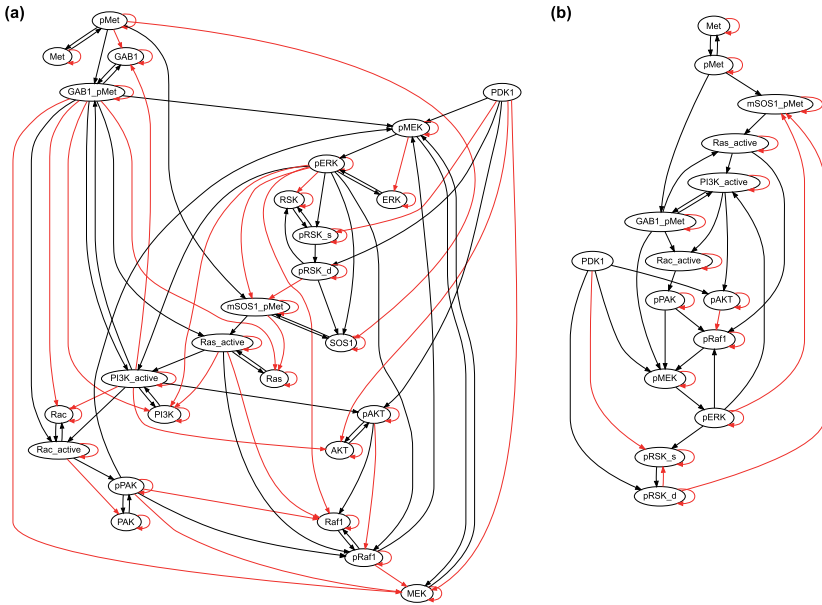


Figure 5.9: Underlying interaction graph of the complete ODE model. (a) Shown is the interaction graph underlying the complete ODE model, thus representing the sign structure of the Jacobian matrix. Black arrows indicate positive, red arrows negative influences. (b) Reduced interaction graph received from the graph in (a) by removing dependent species from conserved moieties. Figure adapted from D’Alessandro, Samaga, Maiwald et al. (2015).

Figure 5.9(a) shows the interaction graph underlying the complete ODE model. In addition to the nodes that are contained in the interaction graph given in Figure 5.8, which represent the activated forms of the proteins, the underlying interaction graph of the ODE model contains for each protein one node representing its inactive form. The reaction $B \rightarrow pB$ with kinetic rate law $k_1 [pA][B]$ (which corresponds to the edge $pA \rightarrow pB$ in the interaction graph) gives rise to the positive edges $pA \rightarrow pB$ and $B \rightarrow pB$, and to the negative edge $pA \dashv B$ in the interaction graph underlying the ODE model. The latter thus induces an additional indirect negative influence of pA on pB ($pA \dashv B \rightarrow pB$). In general, these indirect effects can become visible in the transient

dynamics. However, in our ODE model formulation the active and inactive form of a protein are conserved moieties of the ODE model and thus the interaction graph of its Jacobian can be reduced to the graph given in Figure 5.9(b). This graph contains the same set of nodes as the interaction graph given in Figure 5.8, except for an additional node representing the inactive form of the Met receptor: the moieties of Met and pMet are not conserved as additional uptake and degradation is considered. Still, the paths downstream of the activated receptor are the same as in the interaction graph in Figure 5.8. Compared to Figure 5.9(a), all indirect effects except one could be eliminated: as activation of RSK is modeled as a two-step process, the interaction graph of the ODE model contains with PDK1 \rightarrow pRSK_s and pRSK_d \rightarrow pRSK_d two negative edges that are not contained in the interaction graph in Figure 5.8. However, even if these edges were considered, the selected substructures of the complete interaction graph model would be the minimal structures needed to explain all qualitative effects from the experimental data.

5.5 Ordinary Differential Equation Model Selection

We estimated the kinetic parameters of each of the 16 ODE models as well as of the core and complete model to determine the model performance in relation to the experimental data. While interaction graph modeling was based on discretized data, we utilized the full quantitative information of the data sets for ODE modeling and parameter estimation. In addition to the data sets used for the selection of the 16 model structures, we used time-resolved measurements of HGF-induced activation of the above mentioned proteins and of Met and Ras without inhibitor treatment. Due to its merely qualitative nature, the data set on SOS1 activation was not used for ODE modeling. Overall, the ODE models were calibrated on 2200 data points and 25 experimental conditions including four targeted perturbations.

We applied an adaptation of a likelihood ratio test (LRT) with a threshold of 95%, which takes into account the different degrees of freedom for each model structure (Raue et al. 2013). With this *forward selection*, we received an LRT-based ranking of the model structures (Figure 5.10(a)). As expected, the complete model performed best, whereas the core model performed significantly worse than all selected minimal model structures. The three best performing selected minimal model structures (model 16, 4, and 10) showed a similar likelihood; hence, a clear distinction of a best performing model structure was not possible. Furthermore, a significant gap was observed in the likelihood values between the complete model and the best

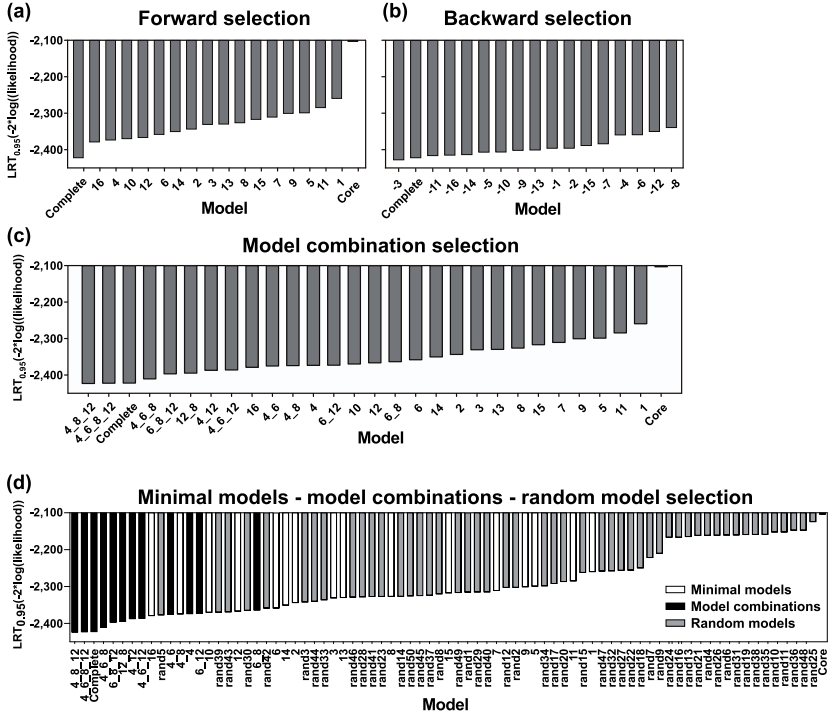


Figure 5.10: ODE model selection. (a) Ranking of the 16 selected minimal model structures and the core and complete model received with the forward selection approach. (b) Backward selection: the building blocks of the respective model structure were removed from the complete model. (c) Ranking of model combinations. (d) Ranking of minimal model structures, model combinations, and random models. All rankings present the negative logarithmic likelihood penalized by parameter difference on the y-axis. Model identifiers are shown on the x-axis. Figure adapted from D’Alessandro, Samaga, Maiwald et al. (2015).

performing selected minimal model structure, model 16, indicating that none of the selected minimal models is a valid simplification of the complete model. This result suggested that models containing combinations of building blocks might perform better than minimal models.

To reduce the complexity of combining all model structures, we applied a *backward selection*: we removed single building blocks from the complete model structure and obtained a new ranking by parameter estimation (Figure 5.10(b)). Strikingly, forward and backward selections revealed very different results. As an example, model 16 was the best performing selected minimal model structure in the forward selection, whereas the removal of the building block of model 16 from the complete model in the backward selection did not lead to a significant loss in likelihood. The highest negative impact on the likelihood was observed by removing the building blocks of any of the model structures 4, 6, 8, or 12, suggesting the importance of their building blocks.

Based on this result from the backward selection, we generated combinations of the minimal model structures 4, 6, 8, and 12 and performed parameter estimation for all eleven possible combinations (Figure 5.10(c)). As expected, the new ranking showed that these model structures filled the gap between the complete model and the best performing minimal model 16. The combinatorial models 4_8_12 and 4_6_8_12 displayed a similar performance as the complete model, indicating that they are valid simplifications of the complete model. Between these models, model 4_8_12 performed best as it contains fewer parameters than model 4_6_8_12. The list of estimated parameter values of model 4_8_12 is given in Appendix F.

Model 4_8_12 contains eight of the thirteen candidate mechanisms included in the complete model. At first glance, it includes three feedback loops (Figure 5.11(a)): Within the MAPK pathway, a positive and a negative feedback emerge from the candidate mechanisms ERK to Raf1 and RSK_d to SOS1, respectively. Within the PI3K pathway, the edge PI3K to GAB1 closes a positive feedback loop. Furthermore, model 4_8_12 is characterized by the presence of several crosstalk mechanisms between PI3K and MAPK. Notably, these mechanisms give rise to two positive and two negative feedback loops, each containing species from both the PI3K and MAPK pathways. As shown in Figure 5.11(b)–(f), model 4_8_12 can reproduce the dynamic behavior of the measured species under diverse experimental conditions. An advantage of the reduced model 4_8_12 is that it has a higher predictive power than the complete model. This is due to parameter non-identifiabilities in the over-parameterized complete model, which resulted in wrong predictions (D’Alessandro,

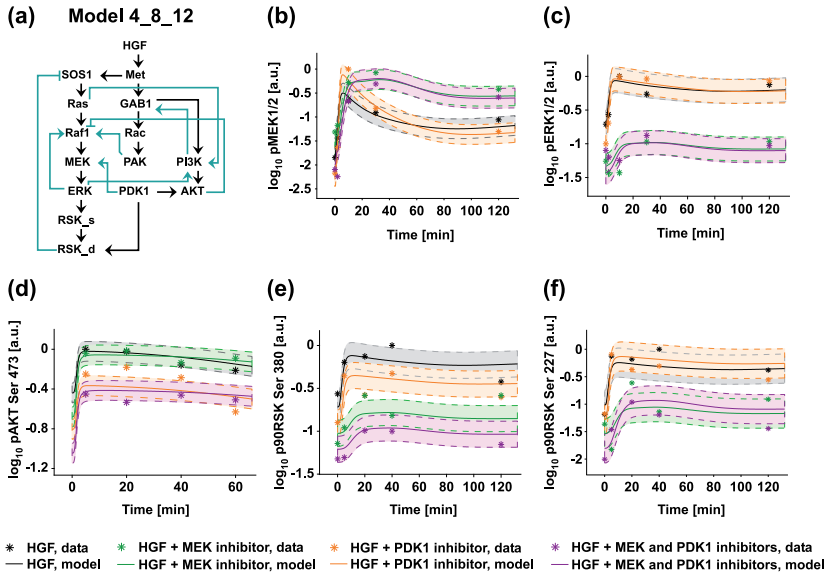


Figure 5.11: ODE model fit. (a) Structure of the best performing model 4_8_12. (b–f) Plots showing representative model trajectories (solid lines) of the phosphorylation kinetic of the indicated proteins measured by quantitative immunoblotting in primary mouse hepatocytes pretreated with the indicated inhibitors and stimulated with 40 ng/ml of HGF for the indicated time (stars). y-axes show the concentration of the respective measured protein in arbitrary units on a logarithmic scale. The shadowed area surrounding the model trajectory represents the confidence interval delimited by the dashed line. Treatments are color-coded as indicated in the figure. Figure adapted from D’Alessandro, Samaga, Maiwald et al. (2015).

Samaga, Maiwald et al. 2015).

To challenge our approach, we generated 50 model structures consisting of randomly selected combinations of candidate edges. We decided to consider random models with five candidate edges (the selected minimal model structures contained either four or five candidate edges). Thus, we randomly selected subsets of size five from all candidate edges present in the complete model and received in this way the building blocks for the random models. Building blocks that were identical to or comprised a building block of the selected minimal model structures were rejected. We performed parameter estimation and derived a ranking of the 50 random models as in the forward selection step (Figure 5.10(d)). This showed that the majority of random models performed worse than our 16 selected models. Random models that performed well had a similar structure as the best performing selected models (Table 5.1 and Figure 5.10(d)). As an example, random model structure 5, which contains three edges that are included both in the building block of model 12 and of model 16, performed similarly as the selected minimal model structure 16. The two additional edges are present in model structure 4 and 8, respectively. Therefore, random model structure 5 is similar to the best performing model structure 4_8_12.

5.6 Experimental Validation and Model Predictions

5.6.1 Experimental Validation of a Negative Crosstalk

To experimentally validate model 4_8_12, we focused on the identified interaction from AKT to Raf1 and predicted the impact of different degrees of AKT inhibition on the inhibitory impact of AKT on Raf1. The model predictions indicate that 3, 6, and 100 fold AKT inhibition results in a 77%, 83%, and 99% reduction of the inhibitory effect of AKT on Raf1 (Figure 5.12(a) and 5.12(b)). The model includes active Raf1, which cannot be directly compared to Raf1 phosphorylated at a specific phosphorylation site as phosphorylation on serine 338 contributes to Raf1 activation, whereas phosphorylation on serine 259, the target of AKT, represents an inactivation signal (Mabuchi et al. 2002). Therefore, we used the effect of AKT inhibition on serine 259 phosphorylation of Raf1 as a proxy to experimentally address the inhibitory impact of AKT on Raf1. Experimentally, we stimulated primary mouse hepatocytes with HGF in the absence and presence of a specific AKT inhibitor and monitored the impact on Raf1 phosphorylation on serine 259. We achieved an inhibition of AKT phosphorylation between 90% and 100% (Figure 5.12(c)) and concomitantly

Table 5.1: Candidate edges of selected models, model combinations, and random models.

	P13K → GAB1	P13K → Rac	ERK → SOS1	RSK_d → SOS1	GAB1 → Ras	PAK → Raf1	ERK → Raf1	AKT → Raf1	PDK1 → MEK	GAB1 → MEK	PAK → MEK	Ras → P13K	ERK → P13K
4_8_12	•	•	•	•	•	•	•	•	•	•	•	•	•
4_6_8_12	•	•	•	•	•	•	•	•	•	•	•	•	•
Complete	•	•	•	•	•	•	•	•	•	•	•	•	•
4_6_8	•	•	•	•	•	•	•	•	•	•	•	•	•
6_8_12	•	•	•	•	•	•	•	•	•	•	•	•	•
12_8	•	•	•	•	•	•	•	•	•	•	•	•	•
4_12	•	•	•	•	•	•	•	•	•	•	•	•	•
4_6_12	•	•	•	•	•	•	•	•	•	•	•	•	•
16	•	•	•	•	•	•	•	•	•	•	•	•	•
rand5	•	•	•	•	•	•	•	•	•	•	•	•	•
4_6	•	•	•	•	•	•	•	•	•	•	•	•	•
4_8	•	•	•	•	•	•	•	•	•	•	•	•	•
4	•	•	•	•	•	•	•	•	•	•	•	•	•
6_12	•	•	•	•	•	•	•	•	•	•	•	•	•
10	•	•	•	•	•	•	•	•	•	•	•	•	•
rand39	•	•	•	•	•	•	•	•	•	•	•	•	•
rand43	•	•	•	•	•	•	•	•	•	•	•	•	•
12	•	•	•	•	•	•	•	•	•	•	•	•	•
rand30	•	•	•	•	•	•	•	•	•	•	•	•	•
6_8	•	•	•	•	•	•	•	•	•	•	•	•	•
rand42	•	•	•	•	•	•	•	•	•	•	•	•	•
6	•	•	•	•	•	•	•	•	•	•	•	•	•
14	•	•	•	•	•	•	•	•	•	•	•	•	•
2	•	•	•	•	•	•	•	•	•	•	•	•	•
rand3	•	•	•	•	•	•	•	•	•	•	•	•	•
rand44	•	•	•	•	•	•	•	•	•	•	•	•	•
rand33	•	•	•	•	•	•	•	•	•	•	•	•	•
3	•	•	•	•	•	•	•	•	•	•	•	•	•
13	•	•	•	•	•	•	•	•	•	•	•	•	•
rand46	•	•	•	•	•	•	•	•	•	•	•	•	•
rand28	•	•	•	•	•	•	•	•	•	•	•	•	•
rand41	•	•	•	•	•	•	•	•	•	•	•	•	•
rand23	•	•	•	•	•	•	•	•	•	•	•	•	•
8	•	•	•	•	•	•	•	•	•	•	•	•	•
rand14	•	•	•	•	•	•	•	•	•	•	•	•	•
rand50	•	•	•	•	•	•	•	•	•	•	•	•	•
rand45	•	•	•	•	•	•	•	•	•	•	•	•	•
rand37	•	•	•	•	•	•	•	•	•	•	•	•	•
rand8	•	•	•	•	•	•	•	•	•	•	•	•	•
15	•	•	•	•	•	•	•	•	•	•	•	•	•
rand49	•	•	•	•	•	•	•	•	•	•	•	•	•
rand1	•	•	•	•	•	•	•	•	•	•	•	•	•
rand29	•	•	•	•	•	•	•	•	•	•	•	•	•
rand40	•	•	•	•	•	•	•	•	•	•	•	•	•
7	•	•	•	•	•	•	•	•	•	•	•	•	•
rand12	•	•	•	•	•	•	•	•	•	•	•	•	•
rand2	•	•	•	•	•	•	•	•	•	•	•	•	•
9	•	•	•	•	•	•	•	•	•	•	•	•	•
5	•	•	•	•	•	•	•	•	•	•	•	•	•
rand34	•	•	•	•	•	•	•	•	•	•	•	•	•
rand17	•	•	•	•	•	•	•	•	•	•	•	•	•
rand20	•	•	•	•	•	•	•	•	•	•	•	•	•
11	•	•	•	•	•	•	•	•	•	•	•	•	•
rand15	•	•	•	•	•	•	•	•	•	•	•	•	•
1	•	•	•	•	•	•	•	•	•	•	•	•	•
rand47	•	•	•	•	•	•	•	•	•	•	•	•	•
rand32	•	•	•	•	•	•	•	•	•	•	•	•	•
rand27	•	•	•	•	•	•	•	•	•	•	•	•	•
rand22	•	•	•	•	•	•	•	•	•	•	•	•	•
rand18	•	•	•	•	•	•	•	•	•	•	•	•	•
rand7	•	•	•	•	•	•	•	•	•	•	•	•	•
rand9	•	•	•	•	•	•	•	•	•	•	•	•	•
rand24	•	•	•	•	•	•	•	•	•	•	•	•	•
rand16	•	•	•	•	•	•	•	•	•	•	•	•	•
rand13	•	•	•	•	•	•	•	•	•	•	•	•	•
rand21	•	•	•	•	•	•	•	•	•	•	•	•	•
rand4	•	•	•	•	•	•	•	•	•	•	•	•	•
rand26	•	•	•	•	•	•	•	•	•	•	•	•	•
rand6	•	•	•	•	•	•	•	•	•	•	•	•	•
rand31	•	•	•	•	•	•	•	•	•	•	•	•	•
rand19	•	•	•	•	•	•	•	•	•	•	•	•	•
rand38	•	•	•	•	•	•	•	•	•	•	•	•	•
rand35	•	•	•	•	•	•	•	•	•	•	•	•	•
rand10	•	•	•	•	•	•	•	•	•	•	•	•	•
rand11	•	•	•	•	•	•	•	•	•	•	•	•	•
rand36	•	•	•	•	•	•	•	•	•	•	•	•	•
rand48	•	•	•	•	•	•	•	•	•	•	•	•	•
rand25	•	•	•	•	•	•	•	•	•	•	•	•	•

Edges that are included in the respective model structure are marked with a filled circle. Model combinations, the complete model, and the 16 selected minimal model structures are highlighted in gray. Rows with white background refer to the 50 random model structures (rand1–50). The models are ordered according to their likelihood value (Figure 5.10).

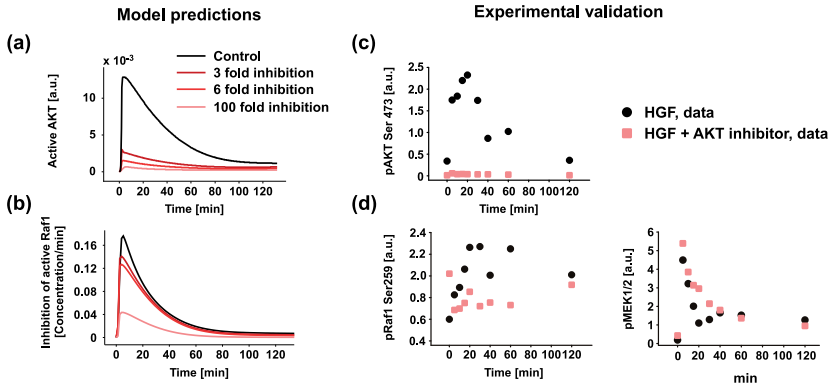


Figure 5.12: Negative crosstalk: experimental validation. (a), (b) Prediction of active AKT and the loss of active Raf1 upon 3, 6, and 100 fold inhibition of active AKT derived with model 4_8_12. (c), (d) Experimental validation of the effect of AKT inhibition in primary mouse hepatocytes treated with 40 ng/ml of HGF alone or in combination with AKT inhibitor. Figure adapted from D’Alessandro, Samaga, Maiwald et al. (2015).

observed a decrease of Raf1 phosphorylation on serine 259 (Figure 5.12(d)). Additionally, a moderate increase of MEK phosphorylation was also observed upon AKT inhibitor treatment (Figure 5.12(d)). The results confirmed the presence of this interaction in our cellular model system upon HGF stimulation.

5.6.2 Inhibitor Combination: Model Predictions and Experimental Validation

We used the inferred model 4_8_12 to identify strategies to efficiently inhibit AKT and ERK signaling. First, we determined the effect of reducing the rate of the outgoing reactions of each protein by 50%. Additionally, we performed simulations for every possible combination of double inhibition. To evaluate the effect of the targeted inhibition on PI3K and MAPK pathway activation, we calculated the area under the curve of AKT and ERK phosphorylation and the sum thereof (Figure 5.13). Single inhibition of PI3K or PDK1 drastically reduced pAKT and, surprisingly, exerted the opposite effect on pERK. This result indicated that the network structure is robust against single inhibitor treatment suggesting that the two signaling pathways compensate each other. The combination of Met and SOS1 inhibitor turned out to be quite promising, as this inhibitor combination reduced both pAKT and pERK activation by about 50% (Figure 5.13).

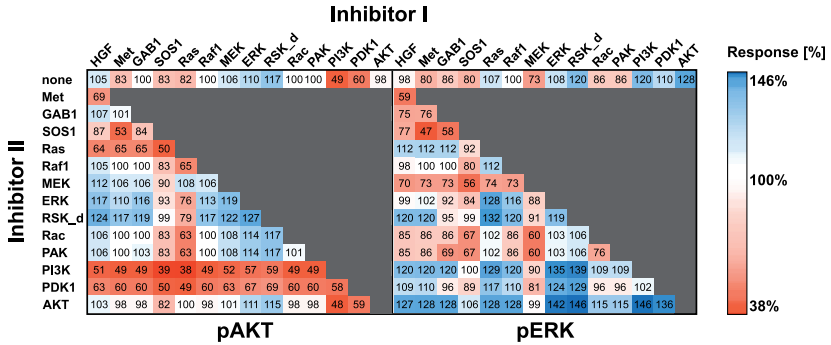


Figure 5.13: Model predictions of inhibitor combinations. Heatmaps showing model simulations of the impact of 50 % inhibitor I individually or in combination with 50 % inhibitor II. As readout, the area under the curve of pAKT and pERK upon inhibitor treatment is compared to the area under the curve of the control condition. The change in the response induced by the inhibitor treatment is indicated as percentage to the control condition. Figure adapted from D’Alessandro, Samaga, Maiwald et al. (2015).

To experimentally validate the model predictions on combinatorial treatments, we first performed single inhibitor treatments targeting PDK1, PI3K, Met, and MEK prior to HGF stimulation and estimated the inhibition strength parameters for each individual inhibitor for model 4_8_12 (Figure 5.14(a)). Based on the estimated inhibitor parameters, we performed predictions of the dynamic behavior of pAKT and pERK upon combinatorial inhibitor treatments. In detail, we simulated the activation kinetics of pAKT and pERK upon combining the PI3K and the MEK inhibitor and upon combining the Met and the PDK1 inhibitor (Figure 5.14(b)). We experimentally validated the model predictions by treating primary mouse hepatocytes with the indicated inhibitor doses and combinations thereof and analyzed the activation kinetics of pAKT and pERK (Figure 5.14(c)). The experimental results indicate that the combination of low dose of the PI3K and the MEK inhibitor slightly increases pAKT and reduces pERK and, therefore, are in good agreement with the model predictions. Interestingly, the combined application of the Met and the PDK1 inhibitor resulted in a reduction of pAKT, while it had only a moderate effect on pERK. Additionally, the calculated area under the curve of pAKT and pERK for the model trajectories and the experimental data are in agreement. In conclusion, model simulations and experimental verifications suggested that the considered signaling network is less sensitive to single interventions, but can be efficiently targeted by

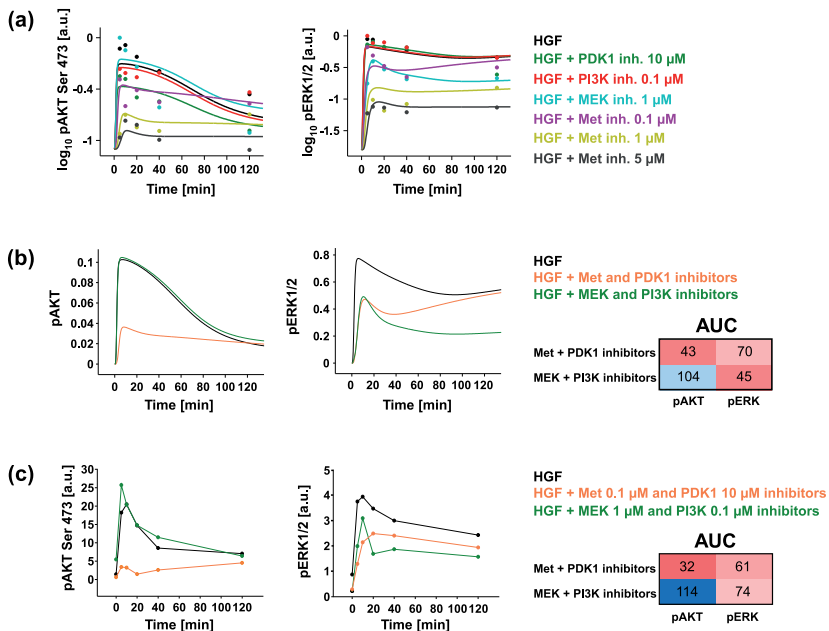


Figure 5.14: Inhibitor combinations: experimental validation. (a) Inhibitor strength parameter estimation. Model 4_8_12 trajectories (solid lines) of the kinetic of pAKT and pERK measured in primary mouse hepatocytes treated with the indicated inhibitor or DMSO prior to HGF 40 ng/ml treatment (filled circles). The experimental data represent the average of two or more replica. (b) Model predictions of pAKT and pERK kinetics. The predictions are based on the inhibitor strength estimated as in (a). In the right panel, the area under the curve (AUC) of the model trajectories of pAKT and pERK is given. (c) Experimental validation of pAKT and pERK kinetics. Primary mouse hepatocytes were treated with the indicated inhibitors or DMSO and subsequently stimulated with 40 ng/ml HGF for the indicated time points. The experimental data is a representative dataset of an experiment performed in biological duplicates. On the right, the quantification of the AUC for pAKT and pERK data is shown. Figure adapted from D’Alessandro, Samaga, Maiwald et al. (2015).

combinatorial treatments. Furthermore, the results regarding inhibitor combinations are an additional validation of the selected model structure.

5.7 Discussion

In this chapter, we presented a network inference approach which benefits from the strengths of both qualitative and quantitative modeling approaches by combining interaction graph and ODE modeling. The key advantage of the presented hybrid approach is that interaction graph modeling enables a preselection of minimal model structures from a vast search space of potential model candidates, which are then translated into ODE models for integrating quantitative details. Several other mathematical modeling approaches also deal with a family of candidate models aiming at identifying the correct wiring. These approaches employ, for example, ensemble modeling (Kuepfer et al. 2007) or Bayesian inference (Xu et al. 2010) and can usually only deal with a limited number (tens to hundreds) of competing ODE models. Other modeling approaches using perturbation data to unravel the network structure rely on modular response analysis (MRA), which requires steady state assumptions and modeling based on linear equations. Therefore, these approaches are more suited to identify feedback and crosstalk mechanisms that determine the medium to long-term behavior (Kholodenko et al. 2002; Klinger et al. 2013). Furthermore, a complete set of perturbation experiments that alter the state of each individual module is required. In contrast, our proposed modeling strategy enables the elucidation of mechanisms influencing the immediate system response and to take into account nonlinear effects such as saturation.

We applied our newly developed hybrid method to HGF signaling in primary mouse hepatocytes. In order to select the minimal model structures, the data analysis approach based on the dependency matrix (Section 3.1) was integrated in an automated routine: Starting from an interaction graph master model of HGF signaling, submodels were derived by iterative addition of candidate mechanisms to the predefined core model structure. For each resulting model structure, model predictions derived from the dependency matrix were compared to qualitative characteristics of time-resolved experimental data. Minimal model structures that were able to explain the data were translated into corresponding ODE models. Importantly, each ODE model was derived in such a way that its underlying interaction graph reflected the respective preselected minimal model structure.

Based on a likelihood ratio test, we received a ranking of the ODE model struc-

tures. Although this “forward selection” did not allow us to identify a single selected minimal model structure as the one explaining the experimental data best, these structures served as important basis for further analysis. The relevance of the selected minimal model structures was also confirmed by a random model analysis, where the majority of random models performed worse than our selected models. By “backward selection”, we identified minimal model structures containing feedback and crosstalk mechanisms relevant for HGF-mediated responses in primary mouse hepatocytes. In this way, only a reasonable number of combinations of minimal model structures had to be tested and we could identify the best performing combinatorial model. Among other feedback and crosstalk mechanisms, the final model structure contained a negative interaction from AKT to Raf1, which we confirmed experimentally in our cellular model system.

Our modeling strategy delivered as output a dynamic ODE model validated with respect to both parameters and network structure. The model can serve as an invaluable tool to not only elucidate cellular signaling pathways, but also to design cell-type- or even patient-specific intervention strategies. In our case study, various interconnections between the PI3K and MAPK pathway in the best performing model structure indicated redundancy within the HGF-stimulated signaling network. Our model predictions suggested that the effects of intervention in one signaling pathway can be compensated by the impact of the other pathway. To maximize the response of inhibition, combinatorial treatments are required as suggested previously (Klinger et al. 2013; Nelander et al. 2008).

The presented study demonstrates that combining modeling approaches of different complexity is a promising strategy to disentangle the complexity of signaling pathways. Despite their simplicity, interaction graphs can be successfully applied to preselect model structures. With increasing complexity of systems under study, the need for such hybrid approaches will become more and more evident. Future developments should also consider the automatization (or semi-automatization) of the different modeling steps to come up with an integrated modeling pipeline.

6 | Prediction of Qualitative Dynamics from Network Structure

In the preceding chapter, we showed how an analysis based on interaction graphs can be used to narrow down the set of potential model candidates that are able to reflect an experimentally observed dynamic behavior. This chapter deals with the identification of mechanisms causing certain dynamic system properties. Interaction graphs again play a central role as they represent the structure of an ODE system given by the sign pattern of its Jacobian matrix. Thus, if one wants to know whether a given ODE network structure may exhibit, for some parameter values, a certain observed dynamic behavior, this can often be answered by inspection of the interaction graph.

Here, we will study structural conditions that allow certain types of complex dynamics in response to a perturbation in steady state. First, we consider perturbation analysis of general ODE systems (Section 6.1) as it has already been studied by Maurya et al. (2003) and Oyeleye and Kramer (1988). We revisit their work and also show how the previous results can be generalized. Second, we consider perturbation analysis of chemical reaction networks and come up with new conditions that take into account the special structure of these networks (Section 6.2). As a main result, we shall see that multistationarity and some dynamic behavior in response to perturbation from steady state are caused by the same mechanisms (Theorem 6.78 and Corollary 6.82). Previous graph-theoretic results for multistationarity in chemical reaction networks were based on bipartite Species–Reaction (SR) graphs. Here, we derive conditions based on interaction graphs and show how these relate to the SR graph results. For certain classes of dynamic chemical reaction networks, the presented conditions are stronger than previous ones (see Example 6.89).

Dietrich Flockerzi and Carsten Conradi both supported the work presented in this chapter with many fruitful discussions.

6.1 Perturbation Analysis for General ODE Systems

Perturbation analysis examines the possible changes of system variables in steady-state-shift experiments: given a system at steady state, one modifies system parameters and observes the resulting change of the variables once the system reaches a new steady state (see Section 2.4.3.2). In Chapter 3, we presented a framework for interrogation and training of interaction graph models based on data from steady-state-shift experiments. Here, we are concerned with the identification of mechanisms in ODE models causing certain dynamic phenomena in response to perturbations in steady state. In particular, we consider the identification of variables that may exhibit an inverse or compensatory response to perturbations. For these variables, propagating the perturbation effect along paths in the interaction graph is not possible (Oyeleye and Kramer 1988). First, we will introduce the problem of perturbation analysis as stated in Oyeleye and Kramer (1988). By using homotopy methods, we derive generalized necessary conditions for inverse and compensatory response, which are not restricted to the linearized system. Finally, we discuss interaction graph topologies of systems allowing inverse or compensatory responses.

6.1.1 General Framework

We consider a system consisting of a set $S = \{S_1, \dots, S_n\}$ of interacting species. The state of the system is given by $x = (x_1, \dots, x_n)$, where $x_i(t)$ denotes the nonnegative concentration of species S_i at time t . For each species S_i , we define a function $f_i : D \rightarrow \mathbb{R}$, $D \supset Q$, $Q = \mathbb{R}_{\geq 0}^n$, describing the rate of production and consumption of the species. It is assumed that the set Q is positively invariant, and that each function f_i is sufficiently smooth on D . The set D is an open neighbourhood of Q .

The system is controlled by an external, constant input, whose strength is given by the function $u : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$. The m -dimensional, nonzero column vector b indicates which of the species are directly affected by the input variable and whether this direct influence of the input on species S_i is positive ($b_i = 1$) or negative ($b_i = -1$):

$$b_i = \begin{cases} \pm 1, & \text{if } i \in I^b \subseteq \{1, \dots, n\}, \\ 0, & \text{if } i \in \{1, \dots, n\} \setminus I^b. \end{cases} \quad (6.1)$$

Thus, the set I^b is the index set of those species that are directly affected by the external input. In most applications, the cardinality of I^b is 1; however, we consider

here the more general case where more than one species can be affected by the external input at the same time, but all with the same strength.

The dynamics of the system is described by a set of ordinary differential equations (ODEs):

$$\dot{x} = H(x, u) = f(x) + b u, \quad x \in D, \quad (6.2)$$

with $H(x, u) = (H_1(x, u), \dots, H_n(x, u))^T$ and $f(x) = (f_1(x), \dots, f_n(x))^T$.

We assume that the system is at steady state for $t = 0$, and, with $u(0) = u^0$, let $x^*(u^0)$ denote this steady state. Given the system at steady state, we change the strength of the external input to another (constant) value:

$$u(t) = u^1 \quad \forall t > 0, \quad u^1 \in \mathbb{R}_{\geq 0}, \quad u^1 \neq u^0. \quad (6.3)$$

As a consequence, the state vector also deviates from the given steady state value. For the sake of convenience, we choose $u^0 = 0$ and $u^1 = 1$.

Thus, in the following, we consider the system

$$\dot{x} = f(x) + b, \quad x(0) = x^*(0), \quad (6.4)$$

where the initial value $x^*(0)$ is the steady state of the unperturbed system $\dot{x} = f(x)$. We will analyze the possible qualitative response of the system, where we focus on initial and ultimate response as defined below.

Remark 6.1. Previous results for perturbation analysis were derived from linearized systems (Oyeleye and Kramer 1988), what makes them valid only for small perturbations. Existing global results were restricted to systems whose structure excludes the existence of multiple steady states (Radulescu et al. 2006; Siegel et al. 2006). In order to derive global results also for systems that have the capacity for multiple steady states, we assume that the perturbation induces a continuous steady-state shift. This will be specified in Section 6.1.4 (Assumption 6.22).

Definition 6.2. Given system (6.4), the *initial response* s_m^I of a state variable x_m is its first nonzero response:

Let $t_1 := \sup_t \{x_m(t) = x_m^*(0) \text{ on } [0, t]\}$. If $t_1 < \infty$ and if there is an $\varepsilon > 0$ with $x_m(t_1 + \varepsilon) \neq x_m^*(0)$, the initial response s_m^I of x_m is given by

$$s_m^I = \text{sgn}(x_m(t_1^+ + \varepsilon) - x_m^*(0)).$$

Definition 6.3. Given system (6.4) and assuming the system reaches a new steady state $x^*(1)$, this new steady state is qualified as the system's *ultimate response*. We specify two types of ultimate response:

- A variable x_m is said to display *inverse response* to perturbations in the input variable if its ultimate response is inverse to its initial response, that is, $\text{sgn}(x_m^*(1) - x_m^*(0)) = -s_m^I$.
- A variable x_m is said to display *compensatory response* to perturbations in the input variable if it returns to the unperturbed state, that is, $x_m^*(1) = x_m^*(0)$.

Figure 6.1 illustrates the definitions.

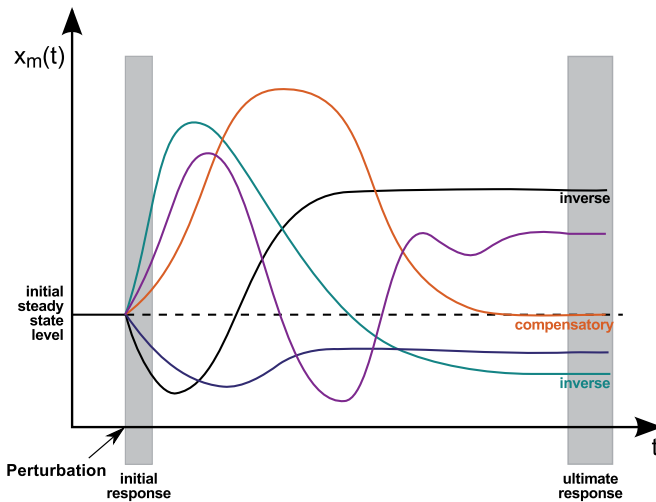


Figure 6.1: Scheme illustrating initial, inverse, and compensatory response. The different-colored trajectories exemplify possible dynamical behaviors of a variable x_m after changing the strength of an external input.

Assumption 6.4. The steady state $x^*(0)$ of system (6.2) is exponentially stable.

Our aim is to come up with criteria that exclude a certain system behavior, and which can be derived solely from the system's structure. For system (6.2), we consider this structure to be given by the sign pattern of the $n \times (n + 1)$ Jacobian matrix

H' of system (6.2), which is given by

$$H' = \left(\frac{\partial H}{\partial x_1}, \dots, \frac{\partial H}{\partial x_n}, \frac{\partial H}{\partial u} \right). \quad (6.5)$$

This sign pattern is determined by the partial derivatives $\partial f_i(x)/\partial x_j$ and the sign pattern of b . We make the following assumption to guarantee that the sign pattern of the Jacobian is constant, in the sense that all entries are either nonpositive, nonnegative, or zero.

Assumption 6.5. For all $i = 1, \dots, n$, the rate function $f_i(x)$ is monotone in each argument, that is, $\frac{\partial f_i}{\partial x_j}(x)$ is either nonpositive, nonnegative, or identically zero for all $x \in D$.

The requested monotonicity (Assumption 6.5) is a common property of many functions describing regulatory functions and chemical kinetics so that a wide range of biological networks can be considered within the framework.

Under Assumption 6.5, we can associate with the Jacobian $J := \frac{\partial H}{\partial x}$ the qualitative matrix $J^q = (J_{ij}^q)$, which is defined as follows.

Definition 6.6. The *qualitative matrix* $J^q = (J_{ij}^q)$ of the Jacobian $J := \frac{\partial H}{\partial x}$ is given by

$$J_{ij}^q = \begin{cases} k_{ij}, & \text{if } \frac{\partial f_i}{\partial x_j}(x) \geq 0 \text{ for all } x \in D, \text{ and } \frac{\partial f_i}{\partial x_j}(x) \neq 0, \\ 0, & \text{if } \frac{\partial f_i}{\partial x_j}(x) \equiv 0 \text{ for all } x \in D, \\ -k_{ij}, & \text{if } \frac{\partial f_i}{\partial x_j}(x) \leq 0 \text{ for all } x \in D, \text{ and } \frac{\partial f_i}{\partial x_j}(x) \neq 0, \end{cases}$$

where $k_{ij} \in \mathbb{R}_{>0}$ for all i, j .

Remark 6.7. We denote J^q as *the* qualitative matrix of J , although it is in fact a family of matrices $J^q(k)$, as we are not interested in the particular values of the k_{ij} . Nevertheless, we will see that choosing $k_{ij} = 1$ for all i, j would not allow to come up with some of our major results, as it will be of importance whether two entries k_{ij} and $k_{i'j'}$ are equal or different.

In the following, we assume that Assumptions 6.4 and 6.5 hold unless otherwise stated.

6.1.2 Determinants and Graphs

In the following, we will see that we can draw conclusions on the behavior of system (6.2) from the signs of certain subdeterminants of its Jacobian. Therefore, we introduce in this section some basic definitions and theorems related to determinants. The classic definition of a determinant is given in terms of permutations and their signs.

Definition 6.8 (Leibniz formula for determinants). Let $\text{Per}(I)$ denote the set of all permutations of $I = \{1, \dots, n\}$, and $\text{sgn}(\sigma)$ the sign of the permutation $\sigma \in \text{Per}(I)$, which is 1 for even and -1 for odd permutations σ . The *determinant* $\det(A)$ of the $n \times n$ matrix $A = (A_{ij})$ is defined by

$$\det(A) = \sum_{\sigma \in \text{Per}(I)} \text{sgn}(\sigma) \cdot A_{1\sigma(1)} \cdot \dots \cdot A_{n\sigma(n)}.$$

The calculation of a determinant can be reduced to the calculation of determinants of lower order. Before we state the well-known recursive formula for evaluating determinants dating back to Laplace, we introduce minors as subdeterminants. Given an $n \times n$ matrix $A = (A_{ij})$ and sets $I \subseteq \{1, \dots, n\}$ and $K \subseteq \{1, \dots, n\}$, we let $A_{I,K}$ denote the submatrix of A with entries A_{ij} with $i \in I, j \in K$.

Definition 6.9. Let A be an $n \times n$ matrix and let I and K be subsets of $\{1, \dots, n\}$ of cardinality p . The determinant of the matrix $A_{I,K}$ is called a *minor of order p* of the matrix A . If $I = K$, this determinant is called *principal minor of order p* , and $(-1)^p \det(A_{I,I})$ a *signed principal minor*.

Theorem 6.10 (Laplace's formula for determinants). Let $A = (A_{ik})$ be an $n \times n$ matrix, $n \geq 2$, and let $I = \{1, \dots, n\} \setminus \{i\}$, $K = \{1, \dots, n\} \setminus \{k\}$. Then, for each $i \in \{1, \dots, n\}$, the determinant of A is given by

$$\det(A) = \sum_{k=1}^n (-1)^{i+k} A_{ik} \det(A_{I,K})$$

(expansion along row i), and, for each $k \in \{1, \dots, n\}$, by

$$\det(A) = \sum_{i=1}^n (-1)^{i+k} A_{ik} \det(A_{I,K})$$

(expansion along column k).

As we want to derive conditions that are only dependent on the sign pattern of the Jacobian, we state here some central definitions for qualitative matrix analysis (Bassett et al. 1968; Brualdi and Shader 1995).

Definition 6.11. A square matrix A is said to have a *signed determinant* if the determinants of all matrices that have the same sign pattern as A are of the same sign.

Definition 6.12. A square matrix A is said to be *sign-nonsingular* (SNS) if A has signed determinant and $\det(A) \neq 0$.

Definition 6.13. A matrix A is said to be *strongly sign determined* (SSD) if each square submatrix of A is either SNS or singular.

It is very well known that the determinant of a matrix can be interpreted in terms of graphs. Therefore, we associate with the $n \times n$ matrix $A = (A_{ij})$ a weighted directed graph $\mathcal{G}(A)$, in this context also known as *flow graph* or *Coates (di)graph* in the literature (Brualdi and Cvetković 2009; Coates 1959). This graph contains n nodes v_1, \dots, v_n , and, for each $i, j \in \{1, \dots, n\}$, we draw an edge from v_j to v_i with edge weight A_{ij} . Edges with zero weight are not drawn. Note that $\mathcal{G}(A)$ may contain selfcycles, corresponding to the diagonal entries of A . Matrix A can be seen as the transposed (weighted) adjacency matrix of the graph $\mathcal{G}(A)$.

As in previous chapters, we define the *interaction graph* associated with a matrix $A = (A_{ij})$ as the signed directed graph containing a positive (negative) edge from v_j to v_i whenever $A_{ij} > 0$ ($A_{ij} < 0$). Thus, the interaction graph contains the same set of nodes and edges as the Coates graph, but, instead of A_{ij} being the weight of the edge from node j to node i , only the sign of A_{ij} is assigned to the edge. We let $\mathcal{G}(A)$ denote both the Coates graph and the interaction graph of a matrix A ; it should be clear from the context which graph we are referring to.

Definition 6.14. Consider the interaction graph associated with the qualitative matrix J^q of system (6.2). This graph contains the nodes S_1, \dots, S_n , which represent the species of the system. Furthermore, we introduce a node U representing the external input and draw, for all $i \in I^b$, an edge from U to S_i with sign $\text{sgn}(b_i)$. The so-derived graph is an interaction graph which represents the matrix that is derived from J^q by appending vector b and will be denoted with $\mathcal{G}(J^q, b)$ accordingly.

With the following definition, we state some terminology from graph theory.

Definition 6.15. Let \mathcal{G} be a weighted directed graph with vertex set V and edge set E .

6 Qualitative Dynamics from Network Structure

- A *walk* in \mathcal{G} is a sequence $v_1 e_1 v_2 e_2 \dots v_k e_k v_{k+1}$ of vertices $v_i \in V$ and edges $e_i \in E$ with $e_i = (v_i, v_{i+1})$. In a *closed* walk, $v_1 = v_{k+1}$.
- The *weight* of a walk is given as product of the weights of the edges it contains.
- The *length* of a walk is given as the number of edges it contains. Repeated edges are counted as many times as they appear.
- A *path* is a walk with pairwise distinct vertices, and a closed walk with pairwise distinct vertices except for the first and last vertex is called *cycle*.
- A *subgraph* of \mathcal{G} is a graph whose set V' of vertices is a subset of V , and whose set E' of edges e'_i is a subset of E , with $e'_i = (v'_i, v'_j)$, $v'_i, v'_j \in V'$. If \mathcal{G}' is a subgraph of \mathcal{G} , the graph \mathcal{G} is called *supergraph* of \mathcal{G}' .
- A *linear subgraph* of \mathcal{G} is a subgraph with vertex set V , which contains exactly one edge into each vertex and exactly one edge out of each vertex.

The following theorem relates the determinant of a matrix to its associated Coates graph (Coates 1959; Harary 1962).

Theorem 6.16. *Let A denote an $n \times n$ matrix and $\mathcal{G}(A)$ the associated Coates graph. The determinant of A is then given by*

$$\det(A) = (-1)^n \sum_{L \in \mathcal{L}(A)} (-1)^{c(L)} \omega(L),$$

where $\mathcal{L}(A)$ denotes the set of all linear subgraphs of $\mathcal{G}(A)$, $c(L)$ the number of cycles contained in L , and $\omega(L)$ the product of the weights of the cycles of L .

Another way to represent the determinant of a matrix based on the associated Coates graph is the following formula, where the determinant is given in terms of cycles in the graph all running over the same node (Maybee and Quirk 1969). Given a set I , we let $\#I$ denote the cardinality of I .

Theorem 6.17. *Let A denote an $n \times n$ matrix and $\mathcal{G}(A)$ the associated Coates graph with nodes indexed by $I = \{1, \dots, n\}$. Then, for all $i \in I$, the determinant of A is given by*

$$\det(A) = \sum_{C \in \mathcal{C}^i(A)} (-1)^{\#I^C - 1} \omega(C) \det(A_{I \setminus \bar{I}^C, I \setminus \bar{I}^C}),$$

where $\mathcal{C}^i(A)$ denotes the set of all cycles in $\mathcal{G}(A)$ containing node S_i , $I^C \subseteq I$ the index set of all nodes of C , $\omega(C)$ the weight of C , and $I^{\bar{C}} = I \setminus I^C$, and where $\det(A_{I^{\bar{C}}, I^{\bar{C}}}) = 1$ if $I^{\bar{C}}$ is the empty set.

6.1.3 A Graph-Theoretic Criterion for Initial Response

A variable's ultimate response is defined as its qualitative change in a new steady state in comparison to its initial response (see Definition 6.3). Here, we restate a graph-theoretic condition for the initial response to disturbances in input variables as introduced in Oyeleye (1990) and Maurya et al. (2003). These authors just considered the linearized system and, thus, only partially proved their statements. We provide a complete proof without linearizing the system in Appendix G.

Theorem 6.18. *Consider system (6.4) and let J_0^q denote the sign structure of the Jacobian at the initial value $x^*(0)$, i.e., $J_0 = f_x(x^*(0))$. If in $\mathcal{G}(J_0^q, b)$ all shortest paths from U to S_m are of the same sign, the initial response of a variable x_m to a change in the input strength u is given by this sign.*

Remark 6.19. For the initial response, we have to consider the qualitative Jacobian at the steady state of the unperturbed system. In general, the Jacobian J^q as defined in Definition 6.6 may contain a non-zero entry where J_0^q is zero. Hence, the shortest path from a species to another might differ in the graphs $\mathcal{G}(J^q)$ and $\mathcal{G}(J_0^q)$. However, the graph-theoretic criteria we derive in this work (Theorem 6.18 as well as criteria for ultimate system response that will be derived in the following) rely on the assumption that all paths between a pair of nodes are of the same sign. In that case, conclusions on the initial response can be drawn from $\mathcal{G}(J^q)$.

6.1.4 Determinant Criteria for Ultimate System Response

In this section, we state conditions to exclude an inverse or compensatory response based on the network structure. Similar conditions are given in Oyeleye (1990) and Maurya et al. (2003); however, again these authors dealt with the linearized system. Here, we show how such criteria can be derived for the nonlinear system by using homotopies (Garcia and Zangwill 1979).

Definition 6.20. Let E^n denote the n -dimensional Euclidean space. Any continuous function $H : D \times [0, 1] \rightarrow E^n$, $D \subset E^n$, is termed a *homotopy*, and the system $H(x, u) = 0$ with $x \in D$, $u \in [0, 1]$, is termed a *homotopy system of equations*. The variable u is called the *homotopy variable*.

For our problem of perturbation analysis, we can interpret the right-hand side of system (6.4) as homotopy: Define the two functions

$$g^0(x) = f(x), \tag{6.6}$$

$$g^1(x) = f(x) + b. \tag{6.7}$$

Then, the homotopy $H(x, u)$ with homotopy variable u describes the deformation from the unperturbed function g^0 to the perturbed function g^1 .

The set H^0 , which we define as

$$H^0 = \{(x, u) \in D \times [0, 1] \mid H(x, u) = 0\}, \tag{6.8}$$

describes all steady states of system (6.4) with $u \in [0, 1]$.

We make the following regulatory assumption:

Assumption 6.21. The Jacobian H' of system (6.4) has full rank n at all $(x, u) \in H^0$.

Both $(x^*(0), 0)$ and $(x^*(1), 1)$ are elements of the set H^0 . Under Assumption 6.21, the set H^0 consists of a finite number of disjoint continuously differentiable paths (Garcia and Zangwill 1979, Theorem 2.1); thus, any $(x, u) \in H^0$ lies on a unique path.

Let us come back to our aim to find necessary conditions for an inverse or compensatory system response after perturbation. We make the following additional assumption:

Assumption 6.22. Given system (6.4) at steady state $x^*(0)$ and switching u from 0 to 1, the system reaches a new steady state $x^*(1)$ that belongs to the path in H^0 with starting point $(x^*(0), 0)$ and end point $(x^*(1), 1)$.

Remark 6.23. Assumptions 6.21 and 6.22 appear simple but are quite restrictive and might be hard to verify for larger systems. However, without these assumptions it is exceedingly difficult to draw general conclusions from the Jacobian matrix, what makes them inevitable at this point.

Definition 6.24. Given an $n \times m$ matrix J , we denote with $J_{i|-b}$ the matrix that is derived from J by replacing column i with the negative of the n -dimensional column vector b .

In the following, let J denote the matrix $\frac{\partial H}{\partial x}$, and $b = \frac{\partial H}{\partial u}$. We parameterize the paths in H^0 by a variable p , that is, we let all variables x_i , $i = 1, \dots, n$, and u be

functions of p . We denote with p_f the first p for which $u = 1$, that is,

$$p_f := \inf\{p : u(p) = 1\}. \quad (6.9)$$

With $k^T(x, u) := (\det(J_{1|b}(x, u)), \dots, \det(J_{n|b}(x, u)), \det(J(x)))$, the parameterization $x(p)$, $u(p)$ will be provided by the ODE system

$$\begin{aligned} \frac{dx_i}{dp} &= (-1)^n \frac{\det(J_{i|b}(x, u))}{1 + \|k(x, u)\|^2}, \quad i = 1, \dots, n, \\ \frac{du}{dp} &= (-1)^n \frac{\det(J(x))}{1 + \|k(x, u)\|^2}, \end{aligned} \quad (6.10)$$

where $\|\cdot\|^2$ denotes the squared Euclidean distance.

Building on previous results (Garcia and Zangwill 1979, Theorem 3.1), the following lemma shows that the system's ultimate response is determined by the ODE system (6.10).

Lemma 6.25. *Assume we are given system (6.4) at steady state $x^*(0)$ and increase u from 0 to 1. Under Assumptions 6.21 and 6.22, the path from $(x^*(0), 0)$ to $(x^*(1), 1)$ is determined by the solution of system (6.10) for $p < p_f$ with initial condition $(x(0), u(0)) = (x^*(0), 0)$.*

Proof. Differentiating $H(x(p), u(p)) = 0$, we get

$$H'(x, u) \begin{pmatrix} \frac{dx}{dp}(x, u) \\ \frac{du}{dp}(x, u) \end{pmatrix} = 0. \quad (6.11)$$

Thus, $(\frac{dx}{dp}, \frac{du}{dp})^T$ must lie in the kernel of $H'(x, u)$.

Introducing $\text{adj}(J)$ as the adjoint of $J = \frac{\partial H}{\partial x}$, the kernel of $H'(x, u)$ is spanned by the vector function

$$\begin{pmatrix} (-1)^{n+1} \text{adj}(J(x)) b \\ (-1)^n \det(J(x)) \end{pmatrix}. \quad (6.12)$$

Observing that the vector function (6.12) equals $k(x, u)$ given by

$$k(x, u) := (\det(J_{1|b}(x, u)), \dots, \det(J_{n|b}(x, u)), \det(J(x)))^T, \quad (6.13)$$

and as Assumption 6.21 guarantees that $k(x, u)$ does not equal the zero vector for any $(x, u) \in H^0$, the vector $k(x, u)$ represents a basis for the kernel of $H'(x, u)$.

The normalization of the right-hand side in (6.10) ensures that the solutions exist

for all p .

Finally, it follows from the assumed exponential stability of $x^*(0)$ (Assumption 6.4) that $(-1)^n \det(J(x^*(0))) > 0$, so that the solution of system (6.10) for $p < p_f$ and with initial value $(x^*(0), 0)$ describes the path from $(x^*(0), 0)$ to $(x^*(1), 1)$. \square

Lemma 6.25 implies the following results about inverse response:

Corollary 6.26. *Consider system (6.4) with initial value $x^*(0)$. Under Assumptions 6.21 and 6.22, a necessary condition for an inverse response of variable x_m in steady state $x^*(1)$ is that there must be a $p < p_f$ with $x(p), u(p)$ satisfying*

$$\operatorname{sgn}\left((-1)^n \det(J_{m|-b}(x(p), u(p)))\right) = -s_m^I,$$

where s_m^I denotes the sign of the initial response of variable x_m .

Proposition 6.27. *Under Assumptions 6.21 and 6.22, an inverse response of variable x_m can be excluded for all systems of the form (6.4) with qualitative Jacobian J^q if the matrix $J_{m|-b}^q$ is SNS and if $\operatorname{sgn}((-1)^n \det(J_{m|-b}^q)) = s_m^I$.*

Proof. If $J_{m|-b}^q$ is SNS, all nonzero terms in its determinant expansion are of the same sign (Brualdi and Shader 1995). According to the definition of J^q , for all x, u , the entries of $J(x, u)$ are either of the same sign as the respective entry in J^q , or $J(x, u)$ has a zero entry where J^q is nonzero. Thus, if $J_{m|-b}^q$ is SNS and if $\operatorname{sgn}((-1)^n \det(J_{m|-b}^q)) = s_m^I$, all nonzero terms in the determinant expansion of $(-1)^n J_{m|-b}(x, u)$ have sign s_m^I , and the necessary condition for inverse response stated in Corollary 6.26 is not satisfied. \square

For compensatory response, we can conclude the following:

Proposition 6.28. *For all systems of the form (6.2) with qualitative Jacobian J^q , assuming Assumptions 6.21 and 6.22 are satisfied, a compensatory response of variable x_m*

- *is excluded if the matrix $J_{m|-b}^q$ is SNS,*
- *is guaranteed if each summand in the determinant expansion of $J_{m|-b}^q$ is zero.*

In the following, we will further examine the structural network properties causing an inverse or compensatory response. The determinant of matrix $J_{m|-b}^q$ can be written as

$$\det(J_{m|-b}^q) = \sum_{i \in I^b} b_i \det(J_{m|-e_i}^q), \quad (6.14)$$

where $J_{m|-e_i}^q$ denotes the matrix derived from J^q by replacing column m with the negative i th unit vector e_i .

Each cycle containing node S_m in $\mathcal{G}(J_{m|-e_i}^q)$ is made up of a path from S_i to S_m and the edge from S_m to S_i with edge weight -1 . Hence, according to Theorem 6.17, the determinant of $J_{m|-e_i}^q$ can be written as summation over all paths from S_i to S_m with weight $-\omega(P)$:

$$\det(J_{m|-e_i}^q) = \sum_{P \in \mathcal{P}_{i \rightarrow m}} (-1)^{\#I^P-1} (-1) \omega(P) \det(J_{I^P, I^P}^q). \quad (6.15)$$

With (6.14), we have

$$(-1)^n \det(J_{m|-b}^q) = \sum_{i \in I^b} \sum_{P \in \mathcal{P}_{i \rightarrow m}} b_i \omega(P) (-1)^{n-\#I^P} \det(J_{I^P, I^P}^q). \quad (6.16)$$

From (6.16) we see that the sign of each summand in the determinant expansion of $J_{m|-b}^q$ is determined by the sign $b_i \omega(P)$ of a path from U to S_m in $\mathcal{G}(J^q, b)$ and by the sign of the determinant of the complementary matrix to that path, which is derived from J^q by removing all rows and columns corresponding to species nodes on the path. Thus, if all paths from U to S_m are of the same sign, it is sufficient to analyze the determinant of smaller subsystems in order to check if an inverse response can be excluded.

Assumption 6.29. In the interaction graph $\mathcal{G}(J^q, b)$ associated with system (6.2), all paths from U to S_m are of the same sign.

Proposition 6.30. Consider system (6.4) with initial value $x^*(0)$, and let Assumptions 6.21 and 6.22 hold. For a path P in $\mathcal{G}(J^q)$, let $I^P \subseteq \{1, \dots, n\}$ denote the set of indices of all species nodes on this path, and let $I^{\bar{P}} = \{1, \dots, n\} \setminus I^P$. If x_m fulfills Assumption 6.29, an inverse response of x_m in steady state $x^*(1)$ is excluded if all paths P from U to S_m in $\mathcal{G}(J^q, b)$ satisfy

$$J_{I^{\bar{P}}, I^{\bar{P}}}^q \text{ has a signed determinant and } (-1)^{n-\#I^P} \det(J_{I^{\bar{P}}, I^{\bar{P}}}^q) \geq 0.$$

6.1.5 Some Graph-Theoretic Implications

In this section, we want to discuss some implications of the conditions for inverse and compensatory response on the interaction graph topology. First of all, this reveals network motifs causing an inverse or compensatory response. In addition, it allows for conclusions about the ultimate system response by interaction graph inspection.

Negative cycle or paths of both signs necessary for inverse and compensatory response. This first property of interaction graphs associated with systems showing an inverse or compensatory response is not derived from the conditions presented above, but results from the fact that inverse and compensatory response are both nonmonotonic dynamical behaviors. A prerequisite for a system variable x_m to be nonmonotonic with respect to changes in the initial value of variable x_i is that in $\mathcal{G}(J^q)$, there are walks from S_i to S_m of opposite signs (Kunze and Siegel 1994). Thus, an inverse or compensatory response of x_m to a perturbation in x_i is only possible if $\mathcal{G}(J^q)$ contains paths of both signs from x_i to x_m , or if a node lying on a path from x_i to x_m is part of a negative cycle.

Positive cycle or paths of both signs necessary for inverse response. Proposition 6.27 implies that an inverse response of x_m is only enabled if at least one of the terms in the determinant expansion of $J_{m|-\mathbf{b}}^q$ has sign $-s_m^I$. If all paths from U to S_m are of the same sign (Assumption 6.29), Proposition 6.30 gives that, for an inverse response, there must be at least one path from U to S_m satisfying $(-1)^{n-\#\mathbb{1}^P} \det(J_{\mathbb{1}^P, \mathbb{1}^P}^q) < 0$. Written in terms of linear subgraphs (Theorem 6.16), we have

$$\sum_{L \in \mathcal{L}(J_{\mathbb{1}^P, \mathbb{1}^P}^q)} (-1)^{c(L)} \omega(L) < 0. \quad (6.17)$$

Thus, under Assumption 6.29, one of the linear subgraphs L of the complementary graph $\mathcal{G}(J_{\mathbb{1}^P, \mathbb{1}^P}^q)$ to a path from U to S_m must contain (i) an even number of cycles with weight $\omega(L) < 0$, or (ii) an odd number of cycles with weight $\omega(L) > 0$. This implies that necessary for inverse response of x_m are either paths of both signs between U and S_m (i.e., Assumption 6.29 is not fulfilled), or the complementary graph to at least one of the paths from U to S_m in $\mathcal{G}(J^q, \mathbf{b})$ must contain a positive cycle.

Positive cycle, paths of both signs, or integrator necessary for compensatory response. Compensatory response of variable x_m can be excluded for all systems with qualitative Jacobian J^q if the matrix $J_{m|-\mathbf{b}}^q$ is SNS (Proposition 6.28). Thus, in order to enable a compensatory response, the determinant expansion of $J_{m|-\mathbf{b}}^q$ must (i) contain terms of both signs, or (ii) be identically zero. For (i), we get the same necessary condition as for inverse response: the complementary graph to at least one of the paths from U to S_m in $\mathcal{G}(J^q, \mathbf{b})$ must contain a positive cycle, or there must be paths of both signs between U and S_m . If none of the two is satisfied, a zero

selfcycle, also known as integrator, is needed for (ii), as otherwise the term referring to the linear subgraph including all selfcycles appears in the determinant expansion.

Example 6.31. As an example for employing perturbation analysis in a regulatory network we consider a dynamical ODE model of T cell receptor signaling (Wittmann et al. 2009). The model was derived from a logical model of T cell activation¹ (Saez-Rodriguez et al. 2007), which was transformed into an ODE model by multivariate polynomial interpolation (Wittmann et al. (2009); see also Section 2.4.3.1). Characteristic of this transformation is that the underlying interaction graph of the original logical model and of the derived ODE model are identical (Wittmann et al. 2009). The interaction graph of the T cell model contains 40 nodes. It has several positive and negative feedback and feedforward loops (Saez-Rodriguez et al. 2007) and, thus, the potential to exhibit inverse or compensatory response. We analyzed for each pair (S_i, S_m) of nodes the possible initial and ultimate response of S_m after perturbing S_i according to the graph-theoretic specifications in Section 6.1.3 and in this section. The initial response is determined for all node pairs, that is, for all pairs (S_i, S_m) , the sign of the shortest path from S_i to S_m is unique. For 1230 of the 1600 node pairs we found that either no path from S_i to S_m exists, or that all paths from S_i to S_m have the same sign, and none of the nodes along these paths is contained in a negative cycle (this type of analysis is facilitated by the notion of the dependency matrix; see Section 2.4.1.2). For the latter, we can thus exclude a compensatory or inverse response as in those cases, the path and feedback structure implies that S_m behaves monotone with respect to changes in S_i . For 235 pairs we found positive and negative paths from S_i to S_m implying that inverse or compensatory response cannot be excluded. Finally, from the 135 cases where only sign-consistent paths from S_i to S_m exist and negative circuits are involved, we found that for just one node pair (TCRphos on itself) inverse and compensatory response is possible; for the other 134 cases inverse and compensatory response can be excluded. Thus, for this type of models, the perturbation analysis is rather conclusive.

6.2 Perturbation Analysis for Chemical Reaction Networks

A large class of biological systems can be described as *chemical reaction networks* (CRNs) that consist of a set of species and a set of reactions between the species.

¹We considered this logical T cell model as Benchmark for the computation of minimal intervention sets in Section 4.4.2.3.

Frequently, the sum of certain species concentrations in a CRN remains constant in the system; as a consequence, the Jacobian of the system does not have full rank, and the criteria for ultimate system response as described in the previous section cannot be applied. For CRNs without conservation relations, the conditions are applicable, but in general of poor predictive power for this class of systems as Assumption 6.29 is usually not fulfilled. In the following, we show how analogous conditions can be derived that take into account the special structure of CRNs and are thus suited for perturbation analysis of these networks. Furthermore, we show the close connection of the derived results to conditions for multistationarity in CRNs (Section 6.2.5).

6.2.1 Dynamic Chemical Reaction Networks

Consider the example reaction



where production of one molecule of species S_3 out of two molecules of S_1 and one molecule of S_2 is described. Following the standard nomenclature for CRNs (see, e.g., Feinberg (1987)), the objects $2S_1 + S_2$ and S_3 are termed the *reactant complex* and *product complex*, respectively, and are represented by the vectors $\mathbf{y} = (2, 1, 0)^T$ and $\mathbf{y}' = (0, 0, 1)^T$. Reaction (6.18) can then be written as $\mathbf{y} \rightarrow \mathbf{y}'$.

Definition 6.32. A *chemical reaction network* \mathcal{N} consists of three finite sets:

1. a set $\mathcal{S} = \{S_1, \dots, S_n\}$ of *species*,
2. a set $\mathcal{C} \subset \mathbb{R}_{\geq 0}^n$ of *complexes*,
3. a set $\mathcal{R} = \{r_1, \dots, r_q\} \subset \mathcal{C} \times \mathcal{C}$ of *reactions* satisfying the following:
 - $(\mathbf{y}, \mathbf{y}) \notin \mathcal{R}$ for all $\mathbf{y} \in \mathcal{C}$,
 - if $\mathbf{y} \in \mathcal{C}$, there exists $\mathbf{y}' \in \mathcal{C}$ such that either $(\mathbf{y}, \mathbf{y}') \in \mathcal{R}$ or $(\mathbf{y}', \mathbf{y}) \in \mathcal{R}$.

Remark 6.33. According to Definition 6.32, we consider a reversible reaction as one single reaction, that is, the information whether a reaction is reversible or irreversible is not contained in the CRN description. We shall see that this information is implicitly given by the associated rate law. However, sometimes it is advantageous to consider a reversible reaction as two single irreversible reactions in a CRN. If we do so in the following, we will denote the respective CRN with $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R}_{\text{ir}})$, where \mathcal{R}_{ir} is the set of irreversible reactions.

A CRN might contain *synthesis reactions* (inflows) and *degradation reactions* (outflows), which are reactions of the form $0 \rightarrow S_i$ and $S_i \rightarrow 0$. A network that contains for each species $S_i \in \mathcal{S}$ a degradation reaction that is only influenced by the species itself is called *fully open* (Shinar and Feinberg 2013).

It is often convenient to describe a CRN by its *stoichiometric matrix*:

Definition 6.34. Given a CRN with n species and q reactions (y_i, y'_i) , $i = 1, \dots, q$, the *stoichiometric matrix* N is defined as the $n \times q$ matrix $(y'_1 - y_1, y'_2 - y_2, \dots, y'_q - y_q)$.

As in the previous sections, we want to study how the concentrations of the species in the CRN change over time. Hence, we let $x_i(t)$ again denote the nonnegative concentration of species S_i at time t . The rate function that describes the rate of production and consumption of each species is now composed of the reaction rates (kinetics) of all reactions influencing this species.

Definition 6.35. Consider a CRN $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ with stoichiometric matrix N .

1. A *kinetics* for \mathcal{N} is an assignment for each reaction $r_i \in \mathcal{R}$ of a differentiable rate function $v_i : D \rightarrow \mathbb{R}_{\geq 0}$, where $D \subseteq \mathbb{R}_{\geq 0}^n$ is an open rectangular domain.
2. The *species formation rate function* for \mathcal{N} with kinetics $v = (v_1, \dots, v_q)$ is given by

$$f(x) = Nv(x). \quad (6.19)$$

The dynamics of a CRN \mathcal{N} with kinetics v and stoichiometric matrix N is described by a set of ODEs:

$$\dot{x} = f(x) = Nv(x). \quad (6.20)$$

The Jacobian $J = \frac{\partial f}{\partial x}$ of system (6.20) is given by the product $J = NV$, where V denotes the $q \times n$ matrix given by $V_{ij} = \frac{\partial v_i}{\partial x_j}$. We will refer to a CRN with stoichiometric matrix N and kinetics v as *dynamic chemical reaction network* $\mathcal{D} = (N, v)$.

Again, we seek to identify criteria by which certain dynamic behaviors can be excluded solely from the system's structure given by the sign pattern of the Jacobian. Instead of demanding a monotone rate function as above (Assumption 6.5), we assume in the following that the kinetics v is monotone, that is, for all $i = 1, \dots, q$ and all $j = 1, \dots, n$, the partial derivative $\frac{\partial v_i}{\partial x_j}(x)$ is either nonpositive, nonnegative, or identically zero for all $x \in D$. Thus, we can associate with the matrix V the qualitative matrix $V^q = (V^q_{ij})$ given by

$$V_{ij}^q = \begin{cases} \alpha_{ij}, & \text{if } \frac{\partial v_i}{\partial x_j}(x) \geq 0 \text{ for all } x \in D, \text{ and } \frac{\partial v_i}{\partial x_j}(x) \neq 0, \\ 0, & \text{if } \frac{\partial v_i}{\partial x_j}(x) \equiv 0 \text{ for all } x \in D, \\ -\alpha_{ij}, & \text{if } \frac{\partial v_i}{\partial x_j}(x) \leq 0 \text{ for all } x \in D, \text{ and } \frac{\partial v_i}{\partial x_j}(x) \neq 0, \end{cases} \quad (6.21)$$

where $\alpha_{ij} \in \mathbb{R}_{>0}$ for all i, j . The qualitative Jacobian J^q is then given by

$$J^q = NV^q. \quad (6.22)$$

Remark 6.36. Again, we denote V^q as *the* qualitative matrix of V , and consequently J^q as *the* qualitative Jacobian, although both are in fact a family of matrices (see Remark 6.7).

Definition 6.37. Given system (6.20) with qualitative matrix V^q , we denote with v^q the set of all possible kinetics v .

Remark 6.38. In contrast to entries of the qualitative matrix considered in the previous section (see Definition 6.6), here, an entry of the qualitative matrix J^q is not necessarily of defined sign, but a sum of summands of fixed sign, for example, $J_{mj}^q = \alpha_{ij} - \alpha_{lj}$, with $\alpha_{ij}, \alpha_{lj} \in \mathbb{R}_{>0}$. See Example 6.40.

If we make further assumptions on the kinetics of a dynamic CRN, it is possible to derive conclusions about the qualitative matrix V^q from the stoichiometric matrix N . Two classes of dynamic CRNs enabling such conclusions are the classes of *weakly monotonic* systems (Shinar and Feinberg 2012) and of *nonautocatalytic* systems (Banaji et al. 2007).

Definition 6.39 (Shinar and Feinberg 2012). A kinetics v for a CRN is *weakly monotonic* if, for each pair of species concentrations x^1 and x^2 , the following implications hold for each reaction $r_i = (y, y') \in \mathcal{R}_{\text{ir}}$ such that $\text{supp } y \subseteq \text{supp } x^1$ and $\text{supp } y \subseteq \text{supp } x^2$:

1. $v_i(x^2) > v_i(x^1) \Rightarrow$ there is a species S_j with $j \in \text{supp } y$ and $x_j^2 > x_j^1$.
2. $v_i(x^2) = v_i(x^1) \Rightarrow x_j^2 = x_j^1$ for all $j \in \text{supp } y$ or there are species S_j, S_k with $j, k \in \text{supp } y$, $x_j^2 > x_j^1$, and $x_k^2 < x_k^1$.

The CRN $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R}_{\text{ir}})$ with kinetics v is said to be *weakly monotonic* if its kinetics v is weakly monotonic.

In less formal terms, the reaction rate of a particular reaction with weakly monotonic kinetics can only increase if the concentration of at least one species of its reactant complex increases. This implies $V_{ij}^q > 0$ if and only if species S_j is a reactant species in reaction $r_i \in \mathcal{R}_{\text{irr}}$, else $V_{ij}^q = 0$.

Example 6.40. Consider the reversible bimolecular reaction $S_1 + S_2 \leftrightarrow S_3$, where S_1 and S_2 are consumed to produce S_3 , and where for each species a separate degradation reaction is considered. The stoichiometric matrix is then given as

$$N = \begin{pmatrix} -1 & -1 & 0 & 0 \\ -1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix}. \quad (6.23)$$

Assuming weakly monotonic kinetics (e.g., mass action kinetics), the matrix V^q is given as

$$V^q = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & 0 & 0 \\ 0 & \alpha_{32} & 0 \\ 0 & 0 & \alpha_{43} \end{pmatrix}, \quad \alpha_{ij} \in \mathbb{R}_{>0}, \quad (6.24)$$

and the qualitative Jacobian $J^q = NV^q$ for this example is

$$J^q = \begin{pmatrix} -\alpha_{11} - \alpha_{21} & -\alpha_{12} & -\alpha_{13} \\ -\alpha_{11} & -\alpha_{12} - \alpha_{32} & -\alpha_{13} \\ \alpha_{11} & \alpha_{12} & \alpha_{13} - \alpha_{43} \end{pmatrix}. \quad (6.25)$$

The class of nonautocatalytic kinetics is wider than that of weakly monotonic kinetics: there, an increase in the concentration of a reactant species, while keeping all other concentrations fixed, cannot cause the reaction rate to decrease, while an increase in the concentration of a product species—again assuming that all other concentrations do not change—cannot result in an increased rate (Banaji et al. 2007). This is formalized as follows:

Definition 6.41 (Banaji et al. 2007). A dynamic CRN $\mathcal{D} = (N, \nu)$ is called *nonautocatalytic (NAC)* if the stoichiometric matrix N and the matrix V^q have opposite sign structure in the following sense:

1. $N_{ij}V_{ji}^q \leq 0$ for all i and j , and
2. For all i and j with $N_{ij} = 0$: $V_{ji}^q = 0$.

A further restriction of CRNs that is usually imposed in connection with NAC networks (Banaji et al. 2007) is the following:

Assumption 6.42. Each species in the CRN occurs only on one side of a reaction.

We end this introduction by giving some characteristics and some further nomenclature of dynamic CRNs that will be of importance later on. The species formation rate function $f(\cdot)$ takes values in the *stoichiometric subspace* $\text{im}(\mathbf{N})$. A trajectory $x(t)$ passing through the vector $x^0 \in \mathbb{R}_{\geq 0}^n$ lies invariably in the set $(x^0 + \text{im}(\mathbf{N})) \cap \mathbb{R}_{\geq 0}^n$, and this invariant linear space is called a *stoichiometric compatibility class*. Two vectors $c, c' \in \mathbb{R}_{\geq 0}^n$ are called *stoichiometrically compatible* if they lie in the same stoichiometric compatibility class, that is, if $c' - c \in \text{im}(\mathbf{N})$. In general, the stoichiometric matrix \mathbf{N} does not have full row rank, so that the stoichiometric subspace is a proper subspace of $\mathbb{R}_{\geq 0}^n$. For s denoting the rank of \mathbf{N} , system (6.20) has $d = n - s$ linearly independent *conservation laws*, each of which is given by a nonzero column vector w fulfilling $w^T \mathbf{N} = 0$. Hence, each conservation law is a vector that lies in the left kernel of \mathbf{N} , and a set of linearly independent conservation laws is given by a basis of the left kernel of \mathbf{N} .

6.2.2 Determinant Criteria for Ultimate Response of Chemical Reaction Networks

The graph-theoretic condition for the initial response (Theorem 6.18) can also be applied to CRNs. For the ultimate response, we state in this section adapted conditions.

Given a system with conservation laws, the determinant criteria for ultimate system response stated in Section 6.1.4 cannot be applied as the singularity of the Jacobian \mathbf{J} conflicts with the stability assumption of the steady state (Assumption 6.4) and, in general, also with the regulatory assumption (Assumption 6.21).

A common approach when analyzing CRNs is to restrict oneself to study the dynamics of the system within stoichiometric compatibility classes (Craciun and Feinberg 2010). We consider the system

$$\dot{x} = f(x) + \mathbf{b}u = \mathbf{N}v(x) + \mathbf{b}u, \quad (6.26)$$

with kinetics v and $n \times m$ stoichiometric matrix \mathbf{N} of rank s . The vector \mathbf{b} and the input u are defined as in Section 6.1.1. Again, we assume that the right-hand side of (6.26) is sufficiently smooth and are interested in the qualitative change of certain

state variables in a new steady state $x^*(1)$ compared to their values in an initial steady state $x^*(0)$.

As the Jacobian $J = \frac{\partial f}{\partial x}$ has $d = n - s$ eigenvalues that are identically zero for all values of x , Assumption 6.4 on the stability of the steady state $x^*(0)$ is substituted by the following assumption.

Assumption 6.43. The Jacobian $J(x^*(0))$ has s nonzero eigenvalues, each having negative real part.

In the following, we want to analyze steady states $x^*(1)$ that lie in the set $x^*(0) + \text{im}(N)$, that is, in the same stoichiometric compatibility class as the steady state $x^*(0)$ we start from. To this end, we let W denote a matrix whose columns w^1, \dots, w^d with $w^i = (w_1^i, \dots, w_n^i)^T$ form a basis of the left kernel of N . Without loss of generality, we order the rows of N in such a way that $W^T = (E_d, -Z^T)$, where E_d denotes the identity matrix of size d . Let N_1 denote the first d rows of N , and let N_2 denote the remaining s rows. As $W^T N = 0$, we have $N_1 = Z^T N_2$, and we can write system (6.26) as

$$\dot{x} = \begin{pmatrix} Z^T N_2 \\ N_2 \end{pmatrix} v(x) + b u. \quad (6.27)$$

In order to restrict ourselves to considering the dynamics within stoichiometric compatibility classes, we have to assure that these classes are not affected by the external input.

Assumption 6.44. The vector b (cf. (6.1)) satisfies $b \in \text{im}(N)$, that is, $W^T b = 0$.

Remark 6.45. If only one species is directly influenced by the external input, Assumption 6.44 is only satisfied if this species is not part of a conservation relation.

Solutions $x(\cdot)$ of system (6.27) that lie in the set $x^*(0) + \text{im}(N)$ satisfy

$$W^T x(t) = W^T x^*(0) \quad \forall t. \quad (6.28)$$

Hence, the steady states of system (6.27) that lie in the set $x^*(0) + \text{im}(N)$ are, under Assumption 6.44, the zeros of

$$\tilde{f}(x) + b u = \begin{pmatrix} W^T x - W^T x^*(0) \\ N_2 v(x) \end{pmatrix} + b u. \quad (6.29)$$

Remark 6.46. The function \tilde{f} is similar to the *associated extended rate function* employed in Wiuf and Feliu (2013). There, the right-hand side reads, in our notation, $(W^T x, N_2 v(x))^T$. In particular, the Jacobian \tilde{J} of \tilde{f} and the Jacobian of the associated extended rate function are equal.

To study the relation between the unperturbed steady state $x^*(0)$ and the steady state $x^*(1)$ reached after setting the external input u from 0 to 1 within stoichiometric compatibility classes, we consider the deformation from the unperturbed function $\tilde{g}^0(x) = \tilde{f}(x)$ to the perturbed function $\tilde{g}^1(x) = \tilde{f}(x) + b u$ described by the homotopy $\tilde{H}(x, u) = \tilde{f}(x) + b u$. The set \tilde{H}^0 describes all steady states of system (6.29) with $u \in [0, 1]$:

$$\tilde{H}^0 = \{(x, u) \in \tilde{D} \times [0, 1] \mid \tilde{H}(x, u) = 0\}, \quad (6.30)$$

where \tilde{D} denotes the set $D \cap \{\text{im}(N) + x^*(0)\}$. Assumption 6.21 is translated as follows:

Assumption 6.47. The Jacobian \tilde{H}' of system (6.29) has full rank n at all $(x, u) \in \tilde{H}^0$.

With $\tilde{J}_{i|-b}$ denoting the matrix that is derived from the Jacobian \tilde{J} of the function \tilde{f} given in (6.29) by replacing column i with the negative of vector b , this gives the following formulation of Lemma 6.25 for CRNs.

Lemma 6.48. *Assume we are given system (6.29) and increase u from 0 to 1. Let $\tilde{k}^T(y, u) := (\det(\tilde{J}_{1|-b}(x, u)), \dots, \det(\tilde{J}_{n|-b}(x, u)), \det(\tilde{J}(x)))$. Under Assumptions 6.22, 6.43, and 6.47, the path from $(x^*(0), 0)$ to $(x^*(1), 1)$ is determined by the solution of*

$$\begin{aligned} \frac{dx_i}{dp} &= (-1)^s \frac{\det(\tilde{J}_{i|-b}(x, u))}{1 + \|\tilde{k}(x, u)\|^2}, \quad i = 1, \dots, n, \\ \frac{du}{dp} &= (-1)^s \frac{\det(\tilde{J}(x, u))}{1 + \|\tilde{k}(x, u)\|^2} \end{aligned} \quad (6.31)$$

for $p < p_f$ with initial condition $(x(0), u(0)) = (x^*(0), 0)$.

As a main result of this work, we can now state conditions for inverse and compensatory response in CRNs, in analogy to Propositions 6.27 and 6.28. With \tilde{J}^q we denote the matrix that is derived from J^q in an analogous way as \tilde{J} is derived from J , that is by replacing certain rows by basis vectors of the left kernel of the stoichiometric matrix.

Theorem 6.49. Under Assumptions 6.22, 6.43, 6.44, and 6.47, an inverse response of variable x_m can be excluded for all systems of the form (6.26) with qualitative Jacobian J^q if the matrix $\tilde{J}_{m|-b}^q$ is SNS and if $\text{sgn}((-1)^s \det(\tilde{J}_{m|-b}^q)) = s_{m'}^1$, where $s_{m'}^1$ denotes the sign of the initial response of variable x_m .

Theorem 6.50. For all systems of the form (6.26) with qualitative Jacobian J^q , assuming Assumptions 6.22, 6.43, 6.44, and 6.47 are satisfied, a compensatory response of variable x_m

- is excluded if the matrix $\tilde{J}_{m|-b}^q$ is SNS,
- is guaranteed if each summand in the determinant expansion of $\tilde{J}_{m|-b}^q$ is zero.

Example 6.51. The following example, taken from Conradi et al. (2005), describes dual phosphorylation and dephosphorylation of a protein A as it occurs, for example, in MAPK signaling. The two activating phosphorylations are mediated by the kinase E_1 , while the phosphatase E_2 dephosphorylates and thus deactivates protein A (see Figure 6.2).

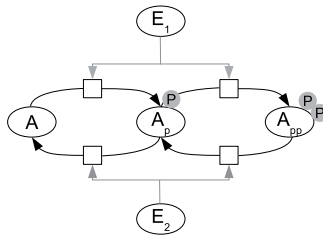
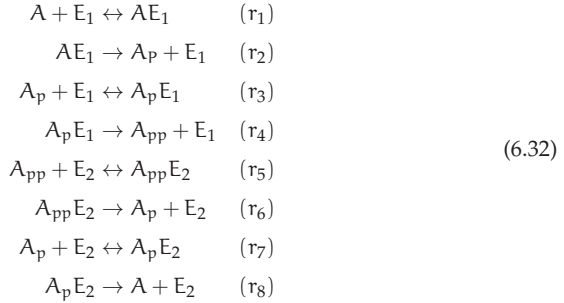


Figure 6.2: Scheme of dual phosphorylation and dephosphorylation mechanism. Enzyme E_1 (a kinase) phosphorylates protein A, enzyme E_2 (a phosphatase) both dephosphorylates the mono- (A_p) and diphosphorylated (A_{pp}) protein. Figure adapted from Conradi et al. (2005).

In Conradi et al. (2005), the authors studied the impact of different phosphorylation and dephosphorylation mechanisms on the possible dynamic behavior of the system. One possibility is that phosphorylation and dephosphorylation follow a distributive mechanism: After binding to A and phosphorylation of the first site, E_1 releases the monophosphorylated protein A_p . The second phosphorylation requires again binding of E_1 , before the dual phosphorylated protein A_{pp} is released. Analogously, the phosphatase E_2 releases the monophosphorylated intermediate A_p before

the final dephosphorylation step occurs after a second binding. This is described by the following network:



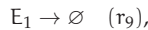
We choose S_1 for A , S_2 for E_1 , S_3 for AE_1 , S_4 for A_p , S_5 for $A_p E_1$, S_6 for A_{pp} , S_7 for E_2 , S_8 for $A_{pp} E_2$, and S_9 for $A_p E_2$.

A basis of the left kernel of the stoichiometric matrix is given by the vectors

$$\begin{aligned}
 w^1 &= (1, 0, 1, 1, 1, 1, 0, 1, 1)^T \\
 w^2 &= (0, 1, 1, 0, 1, 0, 0, 0, 0)^T \\
 w^3 &= (0, 0, 0, 0, 0, 0, 1, 1, 1)^T.
 \end{aligned} \tag{6.33}$$

This shows that all variables are part of a conservation law; consequently, no single perturbation is possible without affecting the stoichiometric compatibility class (cf. Assumption 6.44 and Remark 6.45).

By introducing an outflow in species E_1 ,



that is, by adding the column $-e_2$ to N , the vectors w^1 and w^3 form a basis of the left kernel of N , and we can consider a perturbation in E_1 .

Assuming weakly monotonic kinetics, the matrix V^q is then given by

$$V^q = \begin{pmatrix} \alpha_{11} & \alpha_{12} & -\alpha_{13} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha_{23} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \alpha_{32} & 0 & \alpha_{34} & -\alpha_{35} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha_{45} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \alpha_{56} & \alpha_{57} & -\alpha_{58} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha_{68} & 0 \\ 0 & 0 & 0 & \alpha_{74} & 0 & 0 & \alpha_{77} & 0 & -\alpha_{79} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha_{89} \\ 0 & \alpha_{92} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

and the qualitative Jacobian $J^q = NV^q$ is

$$J^q = \begin{pmatrix} -\alpha_{11} & -\alpha_{12} & \alpha_{13} & 0 & 0 & 0 & 0 & 0 & 0 & \alpha_{89} \\ -\alpha_{11} & -\alpha_{12} - \alpha_{32} - \alpha_{92} & \alpha_{13} + \alpha_{23} & -\alpha_{34} & \alpha_{35} + \alpha_{45} & 0 & 0 & 0 & 0 & 0 \\ \alpha_{11} & \alpha_{12} & -\alpha_{13} - \alpha_{23} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\alpha_{32} & \alpha_{23} & -\alpha_{34} - \alpha_{74} & \alpha_{35} & 0 & -\alpha_{77} & \alpha_{68} & \alpha_{79} & 0 \\ 0 & \alpha_{32} & 0 & \alpha_{34} & -\alpha_{35} - \alpha_{45} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha_{45} & -\alpha_{56} & -\alpha_{57} & \alpha_{58} & 0 & 0 \\ 0 & 0 & 0 & -\alpha_{74} & 0 & -\alpha_{56} & -\alpha_{57} - \alpha_{77} & \alpha_{58} + \alpha_{68} & \alpha_{79} + \alpha_{89} & 0 \\ 0 & 0 & 0 & 0 & 0 & \alpha_{56} & \alpha_{57} & -\alpha_{58} - \alpha_{68} & 0 & 0 \\ 0 & 0 & 0 & \alpha_{74} & 0 & 0 & \alpha_{77} & 0 & -\alpha_{79} - \alpha_{89} & 0 \end{pmatrix},$$

where $\alpha_{ij} > 0$ for all i, j .

The initial response to a perturbation in E_1 can be uniquely determined for all species according to Theorem 6.18. Table 6.1 shows the signs of the initial response for all species assuming E_1 is positively perturbed.

In order to analyze the possible ultimate response, we computed for all $i = 1, \dots, 9$ the determinant of $\tilde{J}_{i|e_2}^q$ and checked the signs of the terms in the determinant expansion (see Table 6.1). For seven out of the nine species, at least one positive and one negative term is contained in the determinant expansion of $\tilde{J}_{i|e_2}^q$, so that an inverse or compensatory response cannot be excluded. The determinant expansion of $\tilde{J}_{5|e_2}^q$, which characterizes the ultimate response of $A_p E_1$, only contains negative terms; thus, as the rank s of N is 7, all terms of $(-1)^s \det(\tilde{J}_{5|e_2}^q)$ are positive. The initial response of $A_p E_1$ is also positive, so that both an inverse and a compensatory response can be excluded. An inverse and compensatory response for $A_{pp} E_2$ can be excluded in the same way.

Table 6.1: Example 6.51, analysis of initial and ultimate response to positive perturbation in E_1

i	1	2	3	4	5	6	7	8	9
Species name	A	E_1	AE_1	A_p	$A_p E_1$	A_{pp}	E_2	$A_{pp} E_2$	$A_p E_2$
Initial response	–	+	+	–	+	+	+	+	–
Terms in det expansion	+/-	+/-	+/-	+/-	+	+/-	+/-	+	+/-
IR excluded	no	no	no	no	yes	no	no	yes	no
CR excluded	no	no	no	no	yes	no	no	yes	no

Row 3: For each species the sign of the initial response to a positive perturbation in E_1 is given. Row 4: Signs of terms in the expansion of $(-1)^s \det(\tilde{J}_{i|-e_2}^q)$. The entry +/- indicates that the determinant expansion contains at least one positive and one negative term. Rows 5 and 6: both inverse response (IR) and compensatory response (CR) to a perturbation in E_1 can only be excluded for the species $A_p E_1$ and $A_{pp} E_2$.

6.2.3 Graphical Representation of Dynamic Chemical Reaction Networks

The determinant criteria for ultimate response derived in the previous section enable to decide from the system structure whether the system may show an inverse or compensatory response. However, they do not provide the information what structural properties cause these responses. Here, conditions that are based on a graphical representation of the network structure can be beneficial.

In Section 6.1, the system under study was represented by an interaction graph. For analyzing the ultimate response of a CRN, this representation is often disadvantageous, as a number of paths and cycles in the interaction graph of a dynamic CRN are not relevant for the steady state behavior and can thus be neglected when analyzing the ultimate response. This can be seen in the following example.

Example 6.52. Consider again the system from Example 6.40 taken with weakly monotonic kinetics. Recall that the qualitative Jacobian J^q for this example reads

$$J^q = \begin{pmatrix} -\alpha_{11} - \alpha_{21} & -\alpha_{12} & \alpha_{13} \\ -\alpha_{11} & -\alpha_{12} - \alpha_{32} & \alpha_{13} \\ \alpha_{11} & \alpha_{12} & -\alpha_{13} - \alpha_{43} \end{pmatrix}, \quad (6.34)$$

where $\alpha_{ij} \in \mathbb{R}_{>0}$.

In the associated interaction graph (Figure 6.3(b)), there is not only the positive edge $S_1 \rightarrow S_3$ (violet edge in Figure 6.3(b)), but also the negative path $S_1 \dashrightarrow S_2 \rightarrow S_3$

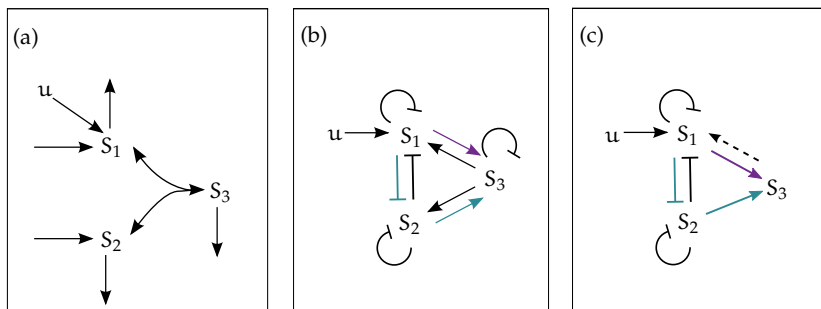


Figure 6.3: Chemical reaction network with bimolecular reaction. (a) Species S_1 and S_2 are consumed in the reversible bimolecular reaction to produce S_3 , and each species can be degraded. We consider an external input directly affecting species S_1 . (b) Interaction graph of the example system in (a). Both a positive (violet) and negative (turquoise) path between S_1 and S_3 are included in the graph. (c) Interaction graph associated with $J_{3|-e_1}^q$ received from the Jacobian J^q of the system in (a). For subfigures (b) and (c): arrows represent positive, blunt-ended lines negative edges.

(turquoise edges in Figure 6.3(b)), as the more S_1 is present, the more S_2 is consumed. In order to analyze the ultimate response of S_3 in response to a perturbation in S_1 , we have to compute the determinant of $J_{3|-e_1}^q$ (cf. Theorems 6.49 and 6.50). Figure 6.3(c) shows the associated interaction graph of $J_{3|-e_1}^q$, also showing the positive and negative path between S_1 and S_3 . Expanding the determinant of $J_{3|-e_1}^q$, we have

$$\det(J_{3|-e_1}^q) = \alpha_{11}\alpha_{12} - \alpha_{11}\alpha_{12} - \alpha_{11}\alpha_{32}. \quad (6.35)$$

The first two summands in (6.35) cancel each other out. Hence, the terms associated with the negative path (indicated in turquoise) have no impact on the ultimate response. Nevertheless, this path may influence the *transient* system behavior.

In the following, we present graphical representations of dynamic CRNs that take into account the special structure of these networks and, as a consequence, enable to identify those paths and cycles that have not to be considered when studying steady state properties.

6.2.3.1 Labeled Interaction Graphs and R-Subgraphs

Interaction graphs are a convenient representation of dynamic CRNs if we extend the original definition given in Section 6.1.2 by incorporating the information about

which reaction causes a certain influence between two species:

Definition 6.53. We associate with the dynamic CRN $\mathcal{D} = (N, \nu)$ the *labeled* interaction graph $\mathcal{G}(N, V)$ containing nodes S_1, \dots, S_n , each representing one of the n species, and containing a directed edge e with edge weight $\omega(e) = N_{ik}V_{kj}^q$ and edge label $\text{lbl}(e) = k$ from S_j to S_i whenever $N_{ik}V_{kj}^q \neq 0$.

We shall see that subgraphs of $\mathcal{G}(N, V)$ where all outgoing edges from a node carry the same edge label and where each label is used only once are of particular importance for our investigations:

Definition 6.54. Consider a labeled interaction graph $\mathcal{G}(N, V)$ and let $I = \{i_1, \dots, i_m\}$ denote a subset of the n species indices and $K = \{k_1, \dots, k_m\}$ a subset of the q reaction indices, $m \leq n$, $m \leq q$. We call the graph $\mathcal{G}(N_{I,K}, V_{K,I}^q)$ an *R-subgraph of order m* (R-subgraph for *reduced* subgraph, see Remark 6.55) of $\mathcal{G}(N, V)$, if the matrix $V_{K,I}^q$ is derived from the matrix $V_{K,I}^q$ by keeping in each row and column exactly one entry not equal 0 and by setting all other entries to 0.

Remark 6.55. Each R-subgraph of order m of $\mathcal{G}(N, V)$ can be interpreted as a maximally reduced subgraph of $\mathcal{G}(N, V)$ containing m nodes and m edges. With $\tau(K) = (\tau_1(K), \dots, \tau_m(K))$ denoting a permutation of the reaction indices K , the matrix $V_{K,I}^\tau = (V_{k_i, i_j}^\tau)$ is defined as

$$V_{k_i, i_j}^\tau = \begin{cases} V_{k_i, i_j}^q & \text{if } k_i = \tau_j(K), \\ 0 & \text{else.} \end{cases}$$

Labeled Interaction Graphs of Weakly Monotonic CRNs. For weakly monotonic CRNs (Definition 6.39), we can, under Assumption 6.42, derive the labeled interaction graph from the stoichiometric matrix and the information whether a reaction is reversible or not. For each reaction $r_k \in \mathcal{R}$, we draw a directed edge $e = (S_j, S_i)$ with edge weight $\omega(e) = N_{ik}\alpha_{kj}$, $\alpha_{kj} \in \mathbb{R}_{>0}$, and edge label $\text{lbl}(e) = k$ whenever $N_{jk} < 0$ and $N_{ik} \neq 0$. If r_k is reversible, we draw in addition a directed edge $e = (S_j, S_i)$ with edge weight $\omega(e) = -N_{ik}\alpha_{kj}$, $\alpha_{kj} \in \mathbb{R}_{>0}$, and edge label $\text{lbl}(e) = k$ whenever $N_{jk} > 0$ and $N_{ik} \neq 0$. Thus, reaction $r_k = (y_k, y'_k)$ gives rise to

- a negative edge (S_i, S_j) between each pair of reactant species $S_i, S_j \in y_k$,
- a negative selfcycle (S_i, S_i) for each reactant species $S_i \in y_k$,

- a positive edge (S_i, S_j) from each reactant species $S_i \in y_k$ to each product species $S_j \in y'_k$.

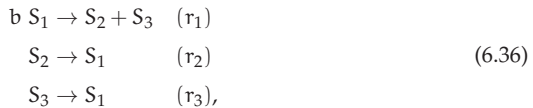
If r_k is reversible, the labeled interaction graph contains, in addition, the following edges with edge label k :

- a negative edge (S_i, S_j) between each pair of product species $S_i, S_j \in y'_k$,
- a negative selfcycle (S_i, S_i) for each product species $S_i \in y'_k$,
- a positive edge (S_i, S_j) from each product species $S_i \in y'_k$ to each reactant species $S_j \in y_k$.

Labeled Interaction Graphs of NAC CRNs. Under Assumption 6.42, we can derive from the stoichiometric matrix a labeled interaction graph that is a supergraph of all CRNs taken with NAC kinetics (Definition 6.41). For each reaction $r_k \in \mathcal{R}$, we draw a directed edge $e = (S_j, S_i)$ with edge weight $\omega(e) = -\text{sgn}(N_{jk})N_{ik}\alpha_{kj}$, $\alpha_{kj} \in \mathbb{R}_{>0}$, and edge label $\text{lbl}(e) = k$ whenever $N_{jk} \neq 0$ and $N_{ik} \neq 0$. Thus, reaction $r_k = (y_k, y'_k)$ gives rise to

- a negative edge (S_i, S_j) between each pair of reactant species $S_i, S_j \in y_k$ and between each pair of product species $S_i, S_j \in y'_k$,
- a negative selfcycle for each reactant species $S_i \in y_k$ and each product species $S_i \in y'_k$,
- a positive edge (S_i, S_j) from each reactant species $S_i \in y_k$ to each product species $S_j \in y'_k$, and
- a positive edge (S_i, S_j) from each product species $S_i \in y'_k$ to each reactant species $S_j \in y_k$.

Example 6.56. Consider the CRN of irreversible reactions



$b \in \mathbb{N}$, with stoichiometric matrix

$$N = \begin{pmatrix} -b & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}.$$

The supergraph of all CRNs of this form taken with NAC kinetics is given in Figure 6.4(b). All possible R-subgraphs of order 3 are shown in Figure 6.4(c)–(e). Assuming weakly monotonic kinetics, the associated labeled interaction graph is the one given in Figure 6.4(c), what is at the same time the only R-subgraph of order 3.

6.2.3.2 Species–Reaction Graphs and Directed Species–Reaction Graphs

Craciun and Feinberg introduced bipartite Species–Reaction graphs (SR graphs) as representations of dynamic CRNs (Craciun and Feinberg 2006b). Banaji and Craciun slightly modified the original definition and associated SR graphs with matrices (Banaji and Craciun 2010). Given any $n \times q$ matrix M , the SR graph $G(M)$ contains two sets of vertices, the S-vertices $\{S_1, \dots, S_n\}$, and the R-vertices $\{R_1, \dots, R_q\}$. For each entry $M_{ij} \neq 0$ of M , we draw an undirected edge e between S_i and R_j with edge sign $\text{sgn}(e) = \text{sgn}(M_{ij})$ and edge weight $\omega(e) = |M_{ij}|$.

Given a CRN with NAC kinetics and assuming that each species occurs only on one side of a reaction (Assumption 6.42), the SR graph $G(N)$ associated with the stoichiometric matrix N has been shown to be a valuable representation of the dynamic CRN, based on which necessary graph-theoretic criteria for multistationarity can be derived (Banaji and Craciun 2010). Inflows and outflows of single reactants are usually not included in the SR graph.

If one considers the smaller class of weakly monotonic CRNs, some undirected edges in the SR graph $G(N)$ are replaced by fixed-direction edges (Knight et al. 2015):

- An edge between R-vertex R_i and S-vertex S_j is directed from R_i to S_j if and only if (i) reaction r_i is irreversible, and (ii) species S_j is a product of reaction r_i .
- An edge between R-vertex R_i and S-vertex S_j is directed from S_j to R_i if and only if (i) reaction r_i is irreversible, and (ii) species S_j is the only reactant of reaction r_i .

Banaji and Craciun introduced the more general directed SR graphs (DSR graphs) as representations of pairs of matrix sets (Banaji and Craciun 2009). For a dynamic

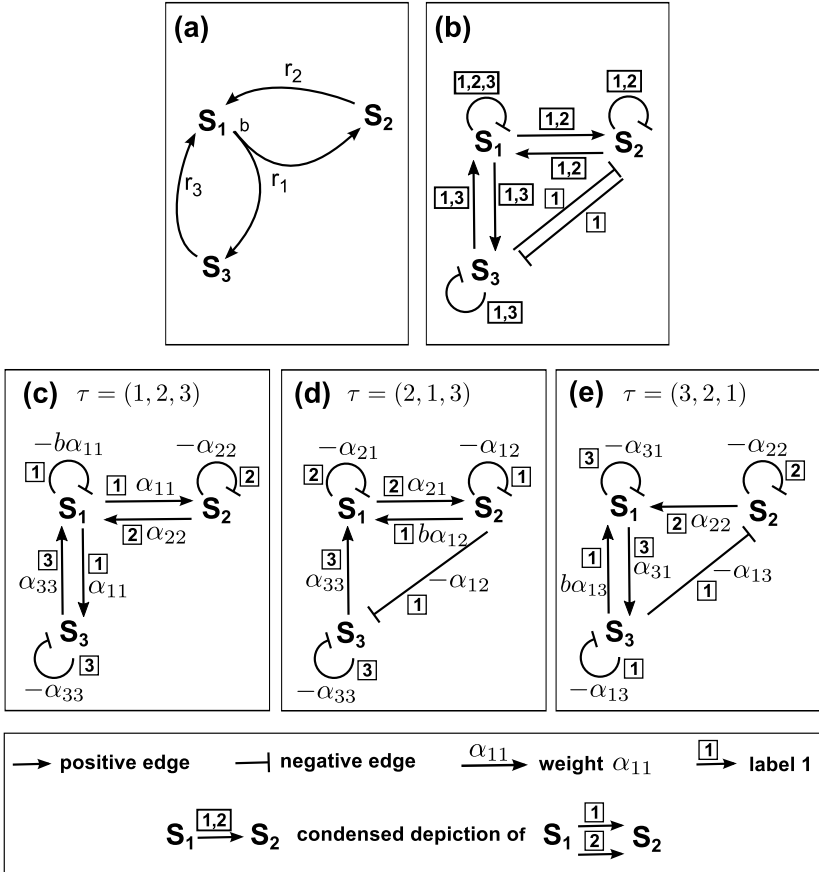


Figure 6.4: Labeled interaction graph and R-subgraphs of order 3 for the example system (6.36). (a) Stoichiometric network of the example system (6.36). Assuming NAC kinetics, any associated labeled interaction graph is a subgraph of the graph given in (b). Subfigures (c)–(e) show the R-subgraphs of order 3 of (b). For weakly monotonic kinetics, $\mathcal{G}(N, V)$ is the graph given in (c), what is at the same time the only R-subgraph of order 3 of this graph.

CRN $\mathcal{D} = (N, \nu)$, the first matrix set consists of a single matrix, the stoichiometric matrix N . The second matrix set is $\mathcal{V} = \{V(x) \mid x \in D\}$. With our monotonicity assumption on the kinetics, we consider this second set to be represented by the qualitative matrix V^q .

The DSR graph $G(N, V)$ associated with the dynamic CRN $\mathcal{D} = (N, \nu)$ with arbitrary kinetics contains two sets of vertices, the species vertices (S-vertices) $\{S_1, \dots, S_n\}$, and the reaction vertices (R-vertices) $\{R_1, \dots, R_q\}$. For each entry $N_{ij} \neq 0$ of N , we draw a directed edge e from R_j to S_i with edge sign $\text{sgn}(e) = \text{sgn}(N_{ij})$ and edge weight $\omega(e) = |N_{ij}|$. For each entry $\nu_{ji}^q \neq 0$ of the qualitative matrix V^q associated with V , we draw a directed edge from S_i to R_j with edge sign $-\nu_{ji}^q$ and edge weight ∞ . If $G(N, V)$ contains two edges of opposite direction but with the same sign between two vertices S_i and R_j , these edges are replaced by a single undirected edge with sign $\text{sgn}(N_{ij})$ and weight $|N_{ij}|$.

Remark 6.57. In contrast to SR graphs, which can only represent networks with NAC kinetics, the DSR graph can also represent influences of species that appear on both sides of a reaction. An example are enzymes, which are not consumed, but influence the reaction. The influence of an enzyme S_i in reaction R_j is represented by an edge $S_i \rightarrow R_j$ with edge weight ∞ .

Remark 6.58. The DSR graph $G(N, V)$ is an amalgamation of directed versions of two SR graphs (Banaji and Craciun 2009): The first SR graph represents the stoichiometric matrix N , and we direct all edges from R- to S-vertices. The second SR graph represents the transposed, negative qualitative matrix associated with V , $-(V^q)^T$, and we direct all edges from S- to R-vertices.

Remark 6.59. Given a stoichiometric matrix N and the information whether a respective reaction is reversible or not, the DSR graph for all systems with weakly monotonic kinetics is equal and is in fact the SR graph $G(N)$ as introduced above. The SR graph introduced for CRNs with NAC kinetics is a supergraph of all possible DSR graphs for systems with stoichiometric matrix N and arbitrary NAC kinetics. An undirected edge in the supergraph thereby corresponds to two opposing directed edges.

Definition 6.60. Given an SR or DSR graph, we make the following definitions (cf. Banaji and Craciun (2009), Shinar and Feinberg (2013)):

1. The *parity* of a walk W with even number of edges k is defined as

$$\text{par}(W) = (-1)^{k/2} \text{sgn}(W),$$

where $\text{sgn}(W)$ is the product of all edge signs of W .

- Given a cycle C containing edges e_1, e_2, \dots, e_{2r} such that e_i and $e_{(i \bmod 2r)+1}$ are adjacent for each $i = 1, \dots, 2r$, the *stoichiometry* of C is defined as

$$\text{stoich}(C) = \left| \prod_{i=1}^r \omega(e_{2i-1}) - \prod_{i=1}^r \omega(e_{2i}) \right|.$$

If any edge in C has weight ∞ , we set $\text{stoich}(C) = \infty$.

- A cycle C is an *e-cycle* if $\text{par}(C) = 1$, an *o-cycle* if $\text{par}(C) = -1$, and an *s-cycle* if $\text{stoich}(C) = 0$.
- A cycle is *oriented* if an orientation can be assigned to each undirected edge such that all edges point in the same direction (clockwise/counterclockwise).
- Given an oriented cycle $S_1 \rightarrow R_1 \rightarrow S_2 \rightarrow \dots \rightarrow R_k \rightarrow S_1$, let $f_{R \rightarrow S}$ denote the edge weight of an edge directed from an R- to an S-vertex, and let $e_{S \rightarrow R}$ denote the edge weight of an edge directed from an S- to an R-vertex. The cycle is *stoichiometrically expansive* if
 - $\frac{f_{R_1 \rightarrow S_2} f_{R_2 \rightarrow S_3} \cdots f_{R_k \rightarrow S_1}}{e_{S_1 \rightarrow R_1} e_{S_2 \rightarrow R_2} \cdots e_{S_k \rightarrow R_k}} > 1$, or if
 - any edge in the cycle has weight ∞ .
- A *critical subgraph* of an SR or DSR graph is a union of oriented e-cycles taken from the graph, where each edge that occurs in more than one cycle has the same orientation in all these cycles.
- Two oriented cycles have an *S-to-R intersection* if each component of their intersection is an *S-to-R path*, that is, a path starting at a species and ending at a reaction node. They have *R-to-S intersection* if each component of their intersection is an *R-to-S path*.

Example 6.61. Figure 6.5 shows the SR graphs for the CRN (6.36) taken with NAC kinetics (Figure 6.5(a)) and with weakly monotonic kinetics (Figure 6.5(b)). The SR graph in subfigure (a) contains two e-cycles, labelled *I* and *II*, and one o-cycle (the large outer cycle). All cycles can be oriented either clockwise or counterclockwise. Cycles *I* and *II* have an S-to-R intersection if cycle *I* is oriented counterclockwise and cycle *II* clockwise. For $\alpha = 1$, none of the cycles is stoichiometrically expansive,

whatever orientation we choose. For $b > 1$, cycle *I* is stoichiometrically expansive if it is oriented clockwise, and cycle *II* is stoichiometrically expansive if it is oriented counterclockwise.

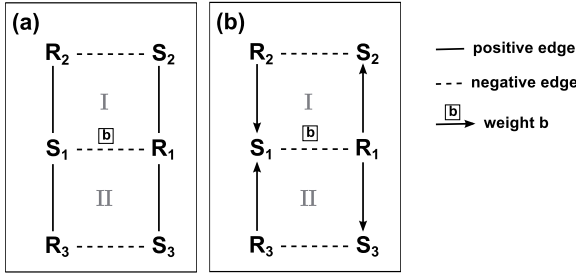


Figure 6.5: SR graphs for the CRN example (6.36). **(a)** SR graph $G(N)$ for the example system (6.36). This graph represents the dynamic CRN with NAC kinetics. **(b)** SR graph for the CRN (6.36) taken with weakly monotonic kinetics. This graph is derived from (a) by replacing some of the undirected edges by fixed-direction edges (indicated by arrowheads). For all subfigures: edges without specified edge weight have edge weight 1.

Assuming weakly monotonic kinetics, the associated SR graph (Figure 6.5(b)) only contains the two e-cycles *I* and *II* with fixed orientation.

Choosing any weakly monotonic kinetics, the associated DSR graph would be the SR graph given in Figure 6.5(b). For any NAC kinetics, the DSR graph would be a subgraph of the SR graph given in Figure 6.5(a) (see Remark 6.59).

Example 6.62. Consider a CRN given by

$$N = \begin{pmatrix} -1 & 0 & 0 \\ -1 & 0 & 0 \\ 1 & -1 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad V^q = \begin{pmatrix} \alpha_{11} & \alpha_{12} & -\alpha_{13} & 0 \\ 0 & 0 & \alpha_{23} & 0 \\ 0 & 0 & \alpha_{33} & \alpha_{34} \end{pmatrix}. \quad (6.37)$$

Assuming mass-action kinetics, the set of reactions is given as follows:



Note that Assumption 6.42 is not met, as species S_4 induces reaction 3 without

being consumed. The associated DSR graph is given in Figure 6.6. The edge weight ∞ of the edge from S_4 to R_3 indicates that S_4 influences reaction R_3 , but is not affected by the reaction. As a consequence, the o-cycle involving species S_3 and S_4 is stoichiometrically expansive.

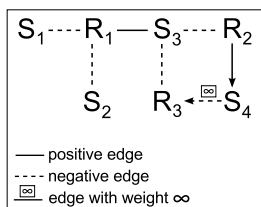


Figure 6.6: DSR graph for the CRN example (6.37).

6.2.3.3 Interrelation between Labeled Interaction Graphs and Directed Species–Reaction Graphs

Labeled interaction graphs and DSR graphs are both representations of CRNs. Here, we will discuss the interrelation between the two formalisms, in particular focusing on the DSR representation of a given R-subgraph.

Consider an R-subgraph with vertices S_1, \dots, S_k and edge labels τ_1, \dots, τ_k , where the edge from S_i to S_j has weight $N_{j\tau_i} V_{\tau_i i}^q$ and edge label τ_i . The corresponding DSR graph contains the S-vertices S_1, \dots, S_k and the R-vertices $R_{\tau_1}, \dots, R_{\tau_k}$. If $i \neq j$, an edge from S_i to S_j in the R-subgraph corresponds to the path $S_i \rightarrow R_{\tau_i} \rightarrow S_j$ in the DSR graph. This path has odd parity if it corresponds to a negative edge, and even parity if it corresponds to a positive edge in the R-subgraph. The edge between R_{τ_i} and S_j is always directed from R_{τ_i} to S_j and has weight $|N_{j\tau_i}|$. In contrast, the edge between S_i and R_{τ_i} can also be undirected, and its edge weight is depending on the selfcycle of S_i . A negative selfcycle on S_i always corresponds to an undirected edge between S_i and R_{τ_i} with weight $|N_{i\tau_i}|$. A positive selfcycle on S_i corresponds to a directed edge from S_i to R_{τ_i} with weight ∞ . In addition, the DSR graph contains in this case an edge directed from R_{τ_i} to S_i of opposite sign; thus, a positive selfcycle in the R-subgraph always corresponds to an e-cycle with weight ∞ in the DSR graph. Finally, if S_i has a zero selfcycle, the DSR graph only contains an edge directed from

S_i to R_{τ_i} with weight ∞ between the two nodes. Zero selfcycles correspond to species that are not consumed in a reaction, yet influencing the reaction kinetics, that is, act as enzymes (see also Remark 6.57. Importantly, the reaction node R_{τ_i} has the species node S_i as only predecessor, and S_i has R_{τ_i} as only successor. Thus, branching points in the DSR graph associated with an R-subgraph are either different edges pointing into an S-node, or different edges pointing out of an R-node.

6.2.4 Necessary Motifs for Inverse Response

In this section, we will identify network motifs causing a certain ultimate response in a CRN. We will restrict ourselves herein to criteria for inverse response.

It can be shown that $\det(\tilde{J})$ is the sum of all principal minors of J of order s (Wiuf and Feliu 2013, Proposition 5.3). The following lemma shows analogously that $\det(\tilde{J}_{i|-e_j}^q)$ is the sum of all principal minors of $J_{i|-e_j}^q$ of order s . Thus, we do not have to consider graphical representations of \tilde{J}^q , but can derive conclusions on the sign structure of $\det(\tilde{J}_{i|-e_j}^q)$ and thus on non-exclusion of an inverse response based on the labeled interaction graph associated with J^q , that is, based on $\mathcal{G}(N, V)$.

Lemma 6.63. *Under Assumption 6.44, we have*

$$\det(\tilde{J}_{i|-e_j}^q) = \sum_{\substack{I \subseteq \{1, \dots, n\} \\ \#I=s}} \det((J_{i|-e_j}^q)_{I,I})$$

for all $i \in \{1, \dots, n\}$ and all $j \in I^b$.

Proof. Apart from one additional aspect, the lemma is proven in the same way as Proposition 5.3 of Wiuf and Feliu (2013). For the sake of completeness, we state the complete proof in Appendix G. \square

In Section 6.1.4, we saw that each term in the determinant expansion of $J_{m|-e_i}^q$ is the product of the weight of a path from S_i to S_m in the interaction graph $\mathcal{G}(J^q)$ and of the signed principal minor corresponding to the complementary system to that path (see Equations (6.15) and (6.16)). The same is true for CRNs, but only paths in R-subgraphs of order s of the labeled interaction graph $\mathcal{G}(N, V)$ contribute to the determinant, as the following lemma shows.

Lemma 6.64. *Let $I \subseteq \{1, \dots, n\}$ denote a subset of s species indices, $K \subseteq \{1, \dots, q\}$ a subset of $s-1$ reaction indices, and τ a permutation of K . For a path P in $\mathcal{G}(N_{I,K}, V_{K,I}^{\tau})$, let $I^P \subseteq I$*

denote the set of indices of all species on the path, $K^P \subseteq K$ the set of all edge labels on the path, and let $I^{\bar{P}} = I \setminus I^P$, $K^{\bar{P}} = K \setminus K^P$. With $\mathcal{P}_{i \rightarrow m}(N_{I,K} V_{K,I}^{\tau})$ denoting the set of all paths from S_i to S_m in $\mathcal{G}(N_{I,K} V_{K,I}^{\tau})$, we have that

$$\det((J_{m|-e_i}^q)_{I,I}) = (-1)^s \sum_{\substack{K \subseteq \{1, \dots, q\} \\ \#K=s-1}} \sum_{\tau \in \text{Per}(K)} \sum_{P \in \mathcal{P}_{i \rightarrow m}(N_{I,K} V_{K,I}^{\tau})} \omega(P) \sum_{L \in \mathcal{L}(N_{I^{\bar{P}}, K^{\bar{P}}} V_{K^{\bar{P}}, I^{\bar{P}}}^{\tau})} (-1)^{c(L)} \omega(L).$$

Proof. We introduce the $n \times (q+1)$ matrix \widehat{N}^i that is derived from N by setting all entries in row i to 0 and by appending the negative i th unit vector as column $q+1$. Analogously, the $(q+1) \times n$ matrix \widehat{V}^m is derived from V^q by setting all entries in column m to 0 and by appending the transposed m th unit vector as row $q+1$. We then have

$$\det((J_{m|-e_i}^q)_{I,I}) = \det(\widehat{N}_{I, \{1, \dots, q+1\}}^i \widehat{V}_{\{1, \dots, q+1\}, I}^m), \quad \text{for all } i \in I^b, \quad (6.39)$$

and, using the Binet-Cauchy formula, that

$$\det((J_{m|-e_i}^q)_{I,I}) = \sum_{\substack{K \subseteq \{1, \dots, q\} \\ \#K=s-1}} \det(\widehat{N}_{I, K \cup \{q+1\}}^i) \det(\widehat{V}_{K \cup \{q+1\}, I}^m), \quad \text{for all } i \in I^b. \quad (6.40)$$

Writing $\det(\widehat{V}_{K \cup \{q+1\}, I}^m)$ in terms of permutations $\tau' \in \text{Per}(K \cup \{q+1\})$ (see Definition 6.8) and letting $\widehat{V}^{m\tau'} = (\widehat{V}_{ij}^{m\tau'})$ be given by

$$\widehat{V}_{ij}^{m\tau'} = \begin{cases} \widehat{V}_{ij}^m & \text{if } i = \tau'_j, \\ 0 & \text{else,} \end{cases}$$

we have that

$$\det((J_{m|-e_i}^q)_{I,I}) = \sum_{\substack{K \subseteq \{1, \dots, q\} \\ \#K=s-1}} \sum_{\tau' \in \text{Per}(K \cup \{q+1\})} \det(\widehat{N}_{I, K \cup \{q+1\}}^i \widehat{V}_{K \cup \{q+1\}, I}^{m\tau'}). \quad (6.41)$$

For $\det((J_{m|-e_i}^q)_{I,I}) \neq 0$, it must hold that $m \in I$ and $i \in I$. The interaction graph $\mathcal{G}(\widehat{N}_{I, K \cup \{q+1\}}^i \widehat{V}_{K \cup \{q+1\}, I}^{m\tau'})$ is derived from $\mathcal{G}(N_{I,K} V_{K,I}^{\tau})$, $\tau \in \text{Per}(K)$, by removing all incoming edges in node S_i and all outgoing edges of node S_m , and by adding an edge S_m to S_i with weight -1 and label $q+1$. Thus, each path from S_i to S_m gives rise to a cycle in $\mathcal{G}(\widehat{N}_{I, K \cup \{q+1\}}^i \widehat{V}_{K \cup \{q+1\}, I}^{m\tau'})$, and this cycle is the only cycle containing node S_i or

S_m . Hence, according to Theorem 6.16, we get

$$\det((J_{m|-e_i}^q)_{I,I}) = (-1)^s \sum_{\substack{K \subseteq \{1, \dots, q\} \\ \#K = s-1}} \sum_{\tau \in \text{Per}(K)} \sum_{P \in \mathcal{P}_{i \rightarrow m}(N_{I,K}, V_{K,I}^\tau)} \omega(P) \sum_{L \in \mathcal{L}(N_{I^P, K^P}, V_{K^P, I^P}^\tau)} (-1)^{c(L)} \omega(L),$$

□

From Lemma 6.64 we see that all paths whose corresponding term appears in the determinant expansion of $(J_{m|-e_i}^q)_{I,I}$ have pairwise distinct edge labels. With $\det((J_{m|-b}^q)_{I,I}) = b_i \sum_{i \in I_b} \det((J_{m|-e_i}^q)_{I,I})$, the following assumption guarantees that all paths in the determinant expansion of $(J_{m|-b}^q)_{I,I}$ are of the same sign.

Assumption 6.65. Consider the system given in (6.26) and its labeled interaction graph $\mathcal{G}(N, V)$. Let $\mathcal{P}_{i \rightarrow j}$ denote the set of all paths in $\mathcal{G}(N, V)$ from S_i to S_j with less than s nodes and pairwise distinct edges labels, and let $\text{sgn}(P)$ denote the sign of P , $P \in \mathcal{P}_{i \rightarrow j}$. Then, for S_j , the sign $\text{sgn}(b_i) \text{sgn}(P)$ is equal for all $i \in I^b$ and all $P \in \mathcal{P}_{i \rightarrow j}$.

Lemma 6.66. Consider the system given in (6.26) and its labeled interaction graph $\mathcal{G}(N, V)$. A shortest path from a node S_j to another node S_i in $\mathcal{G}(N, V)$ has pairwise distinct edge labels.

Proof. Assume a path between two nodes S_j and S_i in $\mathcal{G}(N, V)$ contains two edges with the same label k , $S_{i_1} \rightarrow S_{i_2}$ and $S_{i_3} \rightarrow S_{i_4}$. We have that

$$N_{i_2 k} V_{k i_1} \neq 0, \quad \text{and} \quad N_{i_4 k} V_{k i_3} \neq 0.$$

This implies

$$N_{i_4 k} V_{k i_1} \neq 0,$$

and, hence, that $\mathcal{G}(N, V)$ contains an edge from S_{i_1} to S_{i_4} . Thus, the considered path cannot be the shortest path between S_j and S_i . □

We can now state a condition to exclude inverse response in CRNs in analogy to Proposition 6.30.

Proposition 6.67. Consider the system given in (6.26) at steady state $x^*(0)$ and a change of the external input u from 0 to 1. Let $I \subseteq \{1, \dots, n\}$ denote a subset of s species indices, $K \subseteq \{1, \dots, q\}$ a subset of $s-1$ reaction indices, and τ a permutation of K . For a path P in the R -subgraph $\mathcal{G}(N_{I,K}, V_{K,I}^\tau)$, let $I^P \subseteq I$ denote the set of indices of all species indices on the path, $K^P \subseteq K$ the set of all edge labels on the path, and let $I^{\bar{P}} = I \setminus I^P$, $K^{\bar{P}} = K \setminus K^P$. Under

Assumptions 6.44, 6.47, and 6.22, and if S_m fulfills Assumption 6.65, an inverse response of x_m within stoichiometric compatibility classes is excluded if

$$N_{I^{\bar{P}}, K^{\bar{P}}} V_{K^{\bar{P}}, I^{\bar{P}}}^{\tau} \text{ has a signed determinant and } (-1)^{s-\#I^{\bar{P}}} \det(N_{I^{\bar{P}}, K^{\bar{P}}} V_{K^{\bar{P}}, I^{\bar{P}}}^{\tau}) \geq 0$$

holds for all τ, I, K , and for all paths P from S_i to S_m with $i \in I_b$.

Proof. If S_m fulfills Assumption 6.65, we have that $\text{sgn}(b_i \omega(P))$ is equal for all $i \in I^b$ and all $P \in \mathcal{P}_{i \rightarrow m}(N_{I, K} V_{K, I}^{\tau})$. Furthermore, it follows from Lemma 6.66 and Theorem 6.18 that $\text{sgn}(b_i \omega(P)) = s_m^I$. With

$$\sum_{L \in \mathcal{L}(N_{I^{\bar{P}}, K^{\bar{P}}} V_{K^{\bar{P}}, I^{\bar{P}}}^{\tau})} (-1)^{c(L)} \omega(L) = (-1)^{s-\#I^{\bar{P}}} \det(N_{I^{\bar{P}}, K^{\bar{P}}} V_{K^{\bar{P}}, I^{\bar{P}}}^{\tau}) \quad (6.42)$$

and Theorem 6.49, this completes the proof. \square

The previous results tell us that—as it is the case for general ODE systems—necessary for an inverse response in a CRN are paths of different sign, where, in contrast to the general case, each edge must arise from a different reaction, or a negative determinant of a subsystem that is complementary to one of the paths. In Section 6.1.5, we saw that a positive cycle is necessary for the negative sign of the determinant of the complementary system. In the following, we will discuss structural prerequisites for negative signed principal minors in CRNs.

Lemma 6.68. *Let A denote an $n \times n$ matrix and $\mathcal{G}(A)$ the associated Coates graph, and let $(-1)^n \det(A) < 0$. Then, the graph $\mathcal{G}(A)$ either contains a positive selfcycle, or there is an index set $I \subseteq \{1, \dots, n\}$, $\#I \geq 2$, such that, using the notation from Theorem 6.17, the following inequality holds for all $i \in I$:*

$$\sum_{\substack{C \in \mathcal{C}^1(A_{I, I}) \\ \omega(C) > 0}} (-1)^{\#I^{\bar{C}}} \omega(C) \det(A_{I^{\bar{C}}, I^{\bar{C}}}) > - \sum_{\substack{C \in \mathcal{C}^1(A_{I, I}) \\ \omega(C) < 0}} (-1)^{\#I^{\bar{C}}} \omega(C) \det(A_{I^{\bar{C}}, I^{\bar{C}}}). \quad (6.43)$$

We will refer to (6.43) as inequality of dominant positive cycles in the following.

Proof. According to Theorem 6.17, $(-1)^n \det(A) < 0$ is equivalent to

$$\sum_{C \in \mathcal{C}^1(A)} (-1)^{n-\#I^{\bar{C}}-1} \omega(C) \det(A_{\{1, \dots, n\} \setminus I^{\bar{C}}, \{1, \dots, n\} \setminus I^{\bar{C}}}) < 0, \quad i = 1, \dots, n. \quad (6.44)$$

For $n = 1$, $(-1)^n \det(A) < 0$ can only hold if $A_{11} > 0$, that is, there is a positive selfcycle. For $n = 2$, there must either be a positive self-cycle, or, assuming all selfcycles are nonpositive, the inequalities $A_{12}A_{21} > 0$ and $A_{12}A_{21} > A_{11}A_{22}$ must hold.

For $n > 2$, we consider two cases:

1. One of the signed principal minors in (6.44) is negative, that is, the inequalities (6.44) must hold for a submatrix $A_{I,I}$ of A with $I \subset \{1, \dots, n\}$, $\#I \geq 2$.
2. All signed principal minors in (6.44) are nonnegative, that is,

$$(-1)^{n-\#I^C} \det(A_{I^C,I^C}) \geq 0. \quad (6.45)$$

In this case, the sign of a summand in (6.44) depends on the sign of the respective cycle: if $\omega(C) > 0$, the corresponding summand is negative, and if $\omega(C) < 0$, the corresponding summand is positive.

Hence, there must be an index set $I \subseteq \{1, \dots, n\}$, $\#I \geq 2$, fulfilling the inequality of dominant positive cycles (6.43). \square

From Proposition 6.67 and Lemma 6.68, we see that a necessary condition for inverse response within stoichiometric compatibility classes is the existence of a positive selfcycle or of an R-subgraph $\mathcal{G}(N_{I,K}, V_{K,I}^T)$ of order $2 \leq k \leq n$ fulfilling, with $A = N_{I,K}V_{K,I}^T$, the inequality of dominant positive cycles (6.43). Recall that the matrix $V_{K,I}^T$ contains in each row and column exactly one non-zero entry (see Definition 6.54). Therefore, by reordering the reaction indices accordingly, we can choose $\tau_i = i$, and each element of the matrix A is then given as $A_{ij} = N_{ij}V_{jj}$ with $V_{jj} \neq 0$.

Lemma 6.69. *Consider an $n \times n$ matrix $A = (A_{ij})$, where $A_{ij} = N_{ij}V_{jj}$ for all $i, j \in I, I = \{1, \dots, n\}$, and whose signed principal minors of order k with $k \leq n - 1$ are nonnegative. If $(-1)^n \det(A) < 0$, one of the following holds:*

1. The graph $\mathcal{G}(A)$ contains a positive selfcycle,
2. The graph $\mathcal{G}(A)$ contains a positive cycle C and there is a node index $i \in I^C$ with $N_{ii} = 0$,
3. The graph $\mathcal{G}(A)$ contains a positive cycle C satisfying, with $I^C = \{c_1, \dots, c_r\}$ and

nodes S_{c_i} and $S_{c_{(i \bmod r)+1}}$ being adjacent for each $i \in \{1, \dots, r\}$,

$$\left| \prod_{i=1}^r N_{c_{(i \bmod r)+1}, c_i} \right| > \left| \prod_{i=1}^r N_{c_i, c_i} \right|.$$

4. The graph $\mathcal{G}(A)$ contains two positive cycles that share at least one node.

Proof. According to Lemma 6.68, the graph $\mathcal{G}(A)$ either contains a positive selfcycle, or the following inequality holds for all $i \in \{1, \dots, n\}$:

$$\sum_{\substack{C \in \mathcal{C}^+(A) \\ \omega(C) > 0}} (-1)^{\#\bar{C}} \omega(C) \det(A_{I_{\bar{C}}, I_{\bar{C}}}) > - \sum_{\substack{C \in \mathcal{C}^i(A) \\ \omega(C) < 0}} (-1)^{\#\bar{C}} \omega(C) \det(A_{I_{\bar{C}}, I_{\bar{C}}}) \quad (6.46)$$

Assume all positive cycles in $\mathcal{G}(A)$ are disjoint and let C_i denote the positive cycle running over node S_i . Furthermore, let $I_{C_i} = \{c_1, \dots, c_{r_i}\}$ denote the set of node indices of C_i and let nodes S_{c_j} and $S_{c_{(j \bmod r_i)+1}}$ be adjacent for all $j \in \{1, \dots, r_i\}$. For each $i \in \{1, \dots, n\}$, the left-hand side of (6.46) then contains only the summand

$$(-1)^{n-r_i} \prod_{j=1}^{r_i} N_{c_{(j \bmod r_i)+1}, c_j} V_{c_j, c_j} \det(A_{I_{\bar{C}_i}, I_{\bar{C}_i}}), \quad (6.47)$$

where $I_{\bar{C}_i} = I \setminus I_{C_i}$. The principal minor $\det(A_{I_{\bar{C}_i}, I_{\bar{C}_i}})$ must be nonzero in order to fulfill (6.46). In the following, let $N_{jj} \neq 0$ for all $j \in I_{C_i}$. The right-hand side of (6.46) then contains the summand $-(-1)^{n-1} N_{ii} V_{ii} \det(A_{I \setminus \{i\}, I \setminus \{i\}})$. We now show that $\det(A_{I \setminus \{i\}, I \setminus \{i\}})$ is nonzero. We choose a node S_k with $k \in I_{C_i}$, that is, one of the nodes forming the positive cycle over S_i and write $\det(A_{I \setminus \{i\}, I \setminus \{i\}})$ in terms of cycles running over S_k . As $\mathcal{G}(A)$ contains only disjoint positive cycles, all terms in the determinant expansion of $A_{I \setminus \{i\}, I \setminus \{i\}}$ correspond to negative cycles running over node S_k ; consequently, all nonzero terms are of the same sign. Thus, $\det(A_{I \setminus \{i\}, I \setminus \{i\}})$ can only be zero if all terms are zero. As $N_{kk} \neq 0$, the term $N_{kk} V_{kk} \det(A_{I \setminus \{i, k\}, I \setminus \{i, k\}})$ corresponding to the negative selfcycle on node S_k can only be zero if $\det(A_{I \setminus \{i, k\}, I \setminus \{i, k\}})$ is zero. Again, we can write this determinant in terms of cycles running over another node S_l , $l \in I_{C_i}$. Finally, we see that the right-hand side of (6.46) contains the nonzero term $(-1)^n \prod_{j=1}^{r_i} N_{c_j, c_j} V_{c_j, c_j} \det(A_{I_{\bar{C}_i}, I_{\bar{C}_i}})$. As $(-1)^{n-r_i} \det(A_{I_{\bar{C}_i}, I_{\bar{C}_i}}) > 0$, this implies that condition 3 of the lemma must be satisfied. \square

From Proposition 6.67 and Lemma 6.68 and 6.69, we can now derive the following

proposition about inverse response.

Proposition 6.70. *Consider the system given in (6.26) at steady state $x^*(0)$ and a change of the external input u from 0 to 1. Under Assumptions 6.44, 6.47, and 6.22, and if S_m fulfills Assumption 6.65, an inverse response of x_m within stoichiometric compatibility classes cannot be excluded if any of the labeled interaction graphs associated with complementary systems of the paths from U to S_m contains an R-subgraph with the following properties:*

- all nodes of the R-subgraph are involved in at least one positive cycle, and
- the R-Subgraph satisfies one of the four conditions given in Lemma 6.69.

Proof. From Proposition 6.67 and Lemma 6.68, we see that a necessary condition for inverse response of a species x_m is the existence of an R-subgraph $\mathcal{G}(A)$ in one of the complementary graphs of the paths from U to S_m , which fulfills the inequality of dominant positive cycles 6.43. As this inequality must hold for all nodes of the graph, each node must be involved in at least one positive cycle, what proves the first property of the Proposition. The second property follows directly from Proposition 6.67 and Lemma 6.69 by recalling that, if $\mathcal{G}(A)$ is an R-subgraph, the matrix elements A_{ij} can always be written as $N_{ij}V_{jj}$. \square

Referring to the remarks in Section 6.2.3.3, we can state Proposition 6.70 also in terms of DSR graphs. Assumption 6.65 is translated as follows:

Assumption 6.71 (Formulation of Assumption 6.65 for DSR graphs). Consider the system given in (6.26) and its DSR graph $G(N, V)$. Let $\mathcal{P}_{i \rightarrow j}$ denote the set of all paths in $G(N, V)$ from S_i to S_j with less than $2(s - 1)$ edges, and let $\text{par}(P)$ denote the parity of P , $P \in \mathcal{P}_{i \rightarrow j}$. Then, for S_j , the sign $\text{sgn}(b_i) \text{par}(P)$ is equal for all $i \in I^p$ and all $P \in \mathcal{P}_{i \rightarrow j}$.

Proposition 6.72. *Consider the system given in (6.26) at steady state $x^*(0)$ and a switch of the external input u from 0 to 1. Under Assumptions 6.44, 6.47, and 6.22, and if S_m fulfills Assumption 6.71, an inverse response of x_m within stoichiometric compatibility classes is excluded if none of the DSR graphs associated with the complementary system to any path from S_i , $i \in I^p$, to S_m contains a critical subgraph satisfying*

each S-node has exactly one outgoing edge, and each R-node has exactly one incoming edge, and

- (i) *there is a stoichiometrically expansive e-cycle, or*

(ii) two e-cycles have S-to-R intersection.

Proof. The required R-subgraph stated in Proposition 6.70 corresponds (as each R-subgraph) to a DSR graph where each S-node has exactly one outgoing edge, and each R-node has exactly one ingoing edge (see Section 6.2.3.3). As each node of the R-subgraph must be involved in at least one positive cycle, each species node in the DSR graph must be part of an e-cycle. We choose the reaction indices again in such a way that $\tau_i = i$. Recall that in DSR graphs corresponding to R-subgraphs, an edge between R_i and S_j is always directed from R_i to S_j . As S_j is the only predecessor of R_j , and R_j the only successor of S_j , the edges in any e-cycle which contains two or more S-nodes are always directed from S_j to R_j . E-cycles with one S-node only occur if S_j in the interaction graph is part of a positive selfcycle; in this case we have a stoichiometrical expansive e-cycle $S_j \xrightarrow{\infty} R_j \rightarrow S_j$. All e-cycles are thus consistently oriented, and the R-subgraph corresponds to a critical subgraph. The R-subgraph must fulfill one of the conditions given in Lemma 6.69. The structures needed in Conditions 1, 2, and 3 each correspond to a stoichiometric expansive e-cycle in the DSR graph. As branching points of a DSR graph associated with an R-subgraph are either different edges pointing into an S-node, or different edges pointing out of an R-node, two e-cycles that share at least one node in the R-subgraph correspond to two e-cycles with S-to-R intersection in the DSR graph. \square

Remark 6.73. The necessary graphical condition for inverse response stated in Proposition 6.72 is very similar to necessary DSR graph conditions for multistationarity stated in Banaji and Craciun (2009). We will discuss this in more detail in Section 6.2.5.

Example 6.74. Consider again the CRN (6.36). Assuming weakly monotonic kinetics, the qualitative Jacobian $J^q = NV^q$ is given by

$$J^q = \begin{pmatrix} -bV_{11} & V_{22} & V_{33} \\ V_{11} & -V_{22} & 0 \\ V_{11} & 0 & -V_{33} \end{pmatrix}, \quad (6.48)$$

and $\det(J^q) = (2 - a)V_{11}V_{22}V_{33}$. In this example, the sign of the determinant of the Jacobian depends on the value of b . For $b < 2$, the signed principal minor $(-1)^3 \det(J^q)$ is negative, for $b > 2$, it is positive, and for $b = 2$, it is zero. Thus, for a negative signed principal minor, an amplification of S_1 is required. However, the necessary condition stated in Lemma 6.69 is fulfilled and inverse response can

thus not be excluded, irrespective of the value of b : the labeled interaction graph (Figure 6.4(b)) contains two positive cycles sharing the node S_1 (analogously, the SR graph given in Figure 6.5(b) contains two even cycles with S-to-R intersection). In the following, we derive another necessary condition, which is more conclusive for this example.

Lemma 6.75. *Consider an $n \times n$ matrix $A = (A_{ij})$, where $A_{ij} = N_{ij}V_{jj}$ for all $i, j \in I, I = \{1, \dots, n\}$, and where $\det(N) \neq 0$. Let $I^* \subseteq I$ denote the set of indices such that, for all $j \in I^*$, the edge from S_j to S_i is contained in a positive cycle in $\mathcal{G}(A)$. If A fulfills the inequality of dominant positive cycles (6.43), we have for all $i \in I$ that the inequality*

$$\sum_{j \in I^*} N_{ij} \lambda_j > \left| \sum_{j \in I \setminus I^*} N_{ij} \lambda_j \right| \quad (6.49)$$

is satisfied for the solution $\lambda = (\lambda_1, \dots, \lambda_n)^T$ of $N\lambda = e_i$.

In particular, we have

$$\sum_{j \in I} N_{ij} \lambda_j = 1. \quad (6.50)$$

We will refer to (6.49) as amplifier inequality in the following.

Proof. We inverse each positive cycle in $\mathcal{G}(A)$ containing node S_i to a negative cycle by inverting the sign of the ingoing edge in S_i : we define $N^* = (N_{ij}^*)$ by

$$N_{ij}^* = \begin{cases} -N_{ij}, & \text{if } j \in I^*, \\ N_{ij}, & \text{if } j \in I \setminus I^*, \end{cases} \quad (6.51)$$

and $A^* = (a_{ij}^*)$, where $a_{ij}^* = N_{ij}^* V_{jj}$. Given the inequality of dominant positive cycles (6.43), $\text{sgn}(\det(A)) = -\text{sgn}(\det(A^*))$, and, as $\det(A) = \det(N) \sum_{j \in I} V_{jj}$, we have $\text{sgn}(\det(N)) = -\text{sgn}(\det(N^*))$.

Expanding $\det(N)$ along row $i, i \in I$, we receive

$$\det(N) = \sum_{j \in I} (-1)^{i+j} N_{ij} \det(N_{I \setminus \{i\}, I \setminus \{j\}}), \quad (6.52)$$

and, with $\lambda_j := (-1)^{i+j} \frac{\det(N_{I \setminus \{i\}, I \setminus \{j\}})}{\det(N)}$, we have

$$1 = \sum_{j \in I} N_{ij} \lambda_j > 0. \quad (6.53)$$

Note that according to Cramer's rule, $\lambda = (\lambda_1, \dots, \lambda_n)$ is the unique solution of $N\lambda = e_i$.

As $\text{sgn}(\det(N)) = -\text{sgn}(\det(N^*))$, we have that

$$\sum_{j \in I^*} -N_{ij}\lambda_j + \sum_{j \in I \setminus I^*} N_{ij}\lambda_j < 0. \quad (6.54)$$

An inequality of the form $-a + b < 0 < a + b$ implies $a > 0$ and $|a| > |b|$. \square

In the following, we want to relate Lemma 6.75 to *elementary modes*, a central concept in the field of stoichiometric and metabolic network analysis (Schuster et al. 1999).

Definition 6.76. Given an $n \times m$ stoichiometric matrix N , the vector $\mu = (\mu_1, \dots, \mu_m)^T$, $\mu \neq 0$, is an elementary mode if it satisfies

- (i) $N\mu = 0$,
- (ii) the vector μ is thermodynamically feasible, that is, $\mu_i \geq 0$ if the i th column of N refers to an irreversible reaction, and
- (iii) the vector μ is support-minimal, that is, there is no vector $\mu' \neq 0$ with $I_{\mu'} \subset I_{\mu}$ fulfilling $N\mu' = 0$, where $I_{\mu} = \{i : \mu_i \neq 0\}$ and $I_{\mu'} = \{i : \mu'_i \neq 0\}$.

The solution of $N\lambda = e_i$ can therefore be regarded as an elementary mode of the matrix $(N, -e_i)$, where all reactions contained in N are assumed to be reversible, and the outflow of S_i is assumed to be irreversible. As the amplifier inequality (6.49) must be fulfilled for each $i \in I$, the conditions in Lemma 6.75 show that a necessary motif for a negative signed principal minor is a substructure enabling the production of each of its species in steady state, without external input (if for some i the solution λ of $N\lambda = e_i$ does not satisfy the amplifier inequality (6.49), the corresponding species S_i is not produced, but consumed.) This suggests the following definition.

Definition 6.77. We say that a CRN contains an *amplifying motif of order k* if its associated labeled interaction graph $\mathcal{G}(N, V)$ contains an R-subgraph $\mathcal{G}(N_{I,K}V_{K,I}^r)$ of order k , $2 \leq k \leq n$, satisfying, with $A = N_{I,K}V_{K,I}^r$, the conditions (6.49) and (6.50) given in Lemma 6.75.

Theorem 6.78. Consider the system given in (6.26) at steady state $x^*(0)$ and a change of the external input u from 0 to 1. Furthermore, consider the subgraphs of the labeled

interaction graph $\mathcal{G}(N, V)$, which are complementary to a path from S_i to S_m , where S_i denotes a perturbed node, that is, $i \in I_b$. Under Assumptions 6.44, 6.47, and 6.22, and if S_m fulfills Assumption 6.65, an inverse response of x_m within stoichiometric compatibility classes is excluded if none of those complementary subgraphs to the paths from S_i to S_m contains an amplifying motif or a positive selfcycle.

Proof. From Proposition 6.67 and Lemma 6.68, we saw that under the given assumptions, a necessary condition for inverse response within stoichiometric compatibility classes is the existence of a positive selfcycle, or of an R-subgraph $\mathcal{G}(N_{l,k}, V_{k,l}^T)$ of order $2 \leq k \leq n$ fulfilling, with $A = N_{l,k}V_{k,l}^T$, the inequality of dominant positive cycles (6.43). Furthermore, the R-subgraph or the positive selfcycle must be located in a subgraph that is complementary to one of the paths from a perturbed node to S_m (see Proposition 6.67). If none of the complementary subgraphs to the paths from S_i to S_m contains an amplifying motif or a positive selfcycle, we can conclude from Lemma 6.75 that the inequality of dominant positive cycles (6.43) is not satisfied by the corresponding R-subgraphs. \square

In order to identify the amplifying motifs in a given CRN, one can compute all possible R-subgraphs and check for each R-subgraph if a λ with the required properties exists for all nodes. As the computation of all R-subgraphs is computationally demanding, it is beneficial to include information about the structure of an amplifying motif in order to decrease the number of subgraphs to be analyzed. We do not want to examine this in more detail here. A related work is from Joshi and Shiu (2012), who also analyzed subnetworks to determine whether all terms in the determinant expansion of the Jacobian are of the same sign. They presented an algorithm which reduces a CRN according to certain criteria, thus reducing the number of subnetworks to be analyzed. Similar techniques could be applied here. We will discuss relations to this work in more detail in Section 6.2.5. Given an R-subgraph, the λ can be computed using algorithms for the computation of elementary modes.

6.2.5 Relation to Multistationarity

In this section, we will present some previous results on multistationarity of CRNs and show the close connection to our results on inverse response.

Definition 6.79. (Wiuf and Feliu 2013) A dynamic CRN with $n \times q$ stoichiometric matrix N has the *capacity for multiple positive steady states* over v^q if there exist distinct stoichiometrically compatible vectors $c, c' \in \mathbb{R}_{>0}^n$ such that $Nv(c) = Nv(c') = 0$ for

all kinetics $v \in v^q$. A steady state $x^* \in \mathbb{R}_{>0}^n$ is *degenerate* if $\ker(NV(x^*)) \cap \text{im}(N) \neq 0$, otherwise the steady state is *nondegenerate*.

A large class of methods that enable to decide on a system's capacity for multiple steady states is based on injectivity, as only non-injective systems can have multiple steady states. An example is the work of Wiuf and Feliu (2013), who derived criteria for the preclusion of multiple steady states for general kinetics.

Definition 6.80. (Wiuf and Feliu 2013) A dynamic CRN with kinetics $v \in v^q$ and $n \times m$ stoichiometric matrix N is *injective* over v^q if for any pair of distinct stoichiometrically compatible vectors $c, c' \in \mathbb{R}_{>0}^m$ we have $Nv(c) \neq Nv(c')$ for all $v \in v^q$.

Proposition 6.81. (Feliu and Wiuf 2013) Consider a dynamic CRN with kinetics $v \in v^q$ and $n \times q$ stoichiometric matrix N of rank s . The system is injective over v^q if and only if the matrix \tilde{J}^q is SNS. If $\det(\tilde{J}^q)$ is identically zero, all steady states are degenerate.

Proposition 6.81 implies the following criterion for the preclusion of multiple steady states.

Corollary 6.82. Consider a CRN with kinetics $v \in v^q$ and $n \times q$ stoichiometric matrix N of rank s . Let $I \subseteq \{1, \dots, n\}$ denote a subset of s species indices, $K \subseteq \{1, \dots, q\}$ a subset of s reaction indices, and τ a permutation of K . If $N_{I,K} V_{K,I}^\tau$ has nonnegative signed principal minors for all I, K , and τ , the system does not have the capacity for multiple nondegenerate steady states over v^q .

Proof. This follows as

$$\det(\tilde{J}^q) = \sum_{\substack{I \subseteq \{1, \dots, n\} \\ \#I=s}} \det(J_{I,I}^q)$$

and as

$$\det(J_{I,I}^q) = \sum_{\substack{K \subseteq \{1, \dots, q\} \\ \#K=\#I=s}} \sum_{\tau \in \text{Per } K} \det(N_{I,K} V_{K,I}^\tau).$$

□

Corollary 6.82 immediately shows the connection to our results about inverse response: just as an amplifying motif is necessary for an inverse response that is not caused by oppositely signed feedforward paths or a positive selfcycle, it is also necessary for multiple nondegenerate steady states if all selfcycles are non-positive. While for inverse response, the amplification must be found in a subsystem complementary to any path from the disturbed to the observed node, the amplifying

motif inducing multistationarity can be located in any R-subgraph of order k , where $2 \leq k \leq \text{rank}(N)$.

Proposition 6.83. *Consider a CRN with kinetics $v \in v^q$ and $n \times q$ stoichiometric matrix N of rank s . If all selfcycles in $\mathcal{G}(N, V)$ are nonpositive, and if $\mathcal{G}(N, V)$ does not contain an amplifying motif of order k with $2 \leq k \leq s$, the system does not have the capacity for multiple nondegenerate steady states over v^q .*

From Corollary 6.82, we can also derive the two following propositions for multistationarity, which relate to Propositions 6.70 and 6.72 for inverse response.

Proposition 6.84. *Consider a CRN with kinetics $v \in v^q$ and $n \times q$ stoichiometric matrix N of rank s . If there is no R-subgraph of $\mathcal{G}(N, V)$ of order k with $2 \leq k \leq s$ for which we have*

- *all nodes of the R-subgraph are involved in a positive cycle, and*
- *any of the four conditions given in Lemma 6.69 is satisfied,*

the system does not have the capacity for multiple nondegenerate steady states over v^q .

Proposition 6.85. *Consider a CRN with kinetics $v \in v^q$ and $n \times q$ stoichiometric matrix N of rank s . If the DSR graph $G(N, V)$ does not contain a critical subgraph satisfying*

each S-node has exactly one outgoing edge, and each R-node has exactly one incoming edge, and

- (i) *there is a stoichiometrically expansive e-cycle, or*
- (ii) *two e-cycles have S-to-R intersection,*

the system does not have the capacity for multiple nondegenerate steady states over v^q .

Banaji and Craciun (2009) stated the following DSR graph result about multistationarity:

Proposition 6.86. *A dynamic CRN does not have the capacity for multiple positive nondegenerate steady states if all e-cycles in the DSR graph are s-cycles and if no two e-cycles have S-to-R or R-to-S intersection.*

For NAC kinetics, Propositions 6.85 and 6.86 are equivalent, as the DSR graph of an NAC system only contains undirected edges. However, for weakly monotonic kinetics Proposition 6.85 is sharper.

Example 6.87. Consider the CRN of irreversible reactions



with inflows and outflows in each species. With r_4 denoting the outflow in S_1 , r_5 denoting the outflow in S_2 , and r_6 denoting the outflow in S_3 , and assuming weakly monotonic kinetics, the qualitative Jacobian is given by

$$J^q = \begin{pmatrix} -\alpha_{11} - \alpha_{12} - \alpha_{14} & \alpha_{23} & \alpha_{33} \\ \alpha_{11} & -\alpha_{23} - \alpha_{25} & 0 \\ \alpha_{12} & 0 & -\alpha_{33} - \alpha_{36} \end{pmatrix}, \tag{6.56}$$

with $\alpha_{ij} \in \mathbb{R}^+$, and the determinant expansion of J^q is

$$\begin{aligned} \det(J^q) &= -\alpha_{33}\alpha_{11}\alpha_{25} - \alpha_{33}\alpha_{14}\alpha_{23} - \alpha_{33}\alpha_{14}\alpha_{25} - \alpha_{36}\alpha_{11}\alpha_{25} \\ &\quad - \alpha_{36}\alpha_{12}\alpha_{23} - \alpha_{36}\alpha_{12}\alpha_{25} - \alpha_{36}\alpha_{14}\alpha_{23} - \alpha_{36}\alpha_{14}\alpha_{25} \end{aligned}$$

Hence, multistationarity can be excluded for all weakly monotonic kinetics. According to Proposition 6.86, multistationarity cannot be excluded, while Proposition 6.85 tells us that multiple steady states for the system taken with weakly monotonic kinetics are not possible.

Remark 6.88. Proposition 6.85 is consistent with SR graph results derived by Knight et al. (2015). These results were derived in the context of concordance (Shinar and Feinberg 2012). The connections between SR graphs and concordance were worked out by introducing a sign-causality graph, from which a central inequality containing stoichiometric coefficients and signed coefficients for each reaction was derived (Shinar and Feinberg 2013). This inequality shows similarities with our results presented in Lemma 6.75. A central difference, however, is that concordance results are derived considering the overall network, whereas the identification of an amplifying motif requires the analysis of subnetworks. A closer investigation how our results relate to the theory of concordance is subject for future research.

Example 6.89. Consider again the CRN discussed in Example 6.56, but with inflows and outflows on every node (fully open network). First, we assume NAC kinetics for all reactions. Both Proposition 6.85 and Proposition 6.86 tell us that multistationarity

cannot be excluded, independently of the choice of b . Indeed, the system taken with NAC kinetics has the capacity for multiple positive steady states for all choices of b , the determinant expansion has summands of opposite sign for all choices of b (not shown explicitly here).

For $b = 1$, the R-subgraph with $I = \{1, 2, 3\}$, $K = \{1, 2, 3\}$, and $\tau = (1, 2, 3)$ (see Figure 6.7(a)) is an amplifying motif: with λ^i denoting the solution of $N_{I,K}\lambda^i = e^i$, the inequalities given in Lemma 6.75 are satisfied by $\lambda^1 = (1, 1, 1)^T$, $\lambda^2 = (1, 0, 1)^T$, $\lambda^3 = (1, 1, 0)$. The corresponding DSR graph (see Figure 6.5(b)) has two e-cycles with S-to-R intersection. For $b = 2$, the amplifying motif is the R-subgraph with $I = \{1, 2\}$, $K = \{1, 2\}$, $\tau = (2, 1)$, and $\lambda^1 = (-1, -1)^T$, $\lambda^2 = (-1, -2)^T$ (see Figure 6.7(b)). The corresponding DSR graph has the stoichiometrically expansive e-cycle $S_1 \xrightarrow{3} R_1 \rightarrow S_2 \rightarrow R_2 \rightarrow S_1$ (see Figure 6.5(b)). For $b = 3$, the amplifying motif is again the R-subgraph with $I = \{1, 2\}$, $K = \{1, 2\}$, $\tau = (2, 1)$ (see Figure 6.7(c)), now with $\lambda^1 = (-0.5, -0.5)^T$, $\lambda^2 = (-0.5, -1.5)^T$.

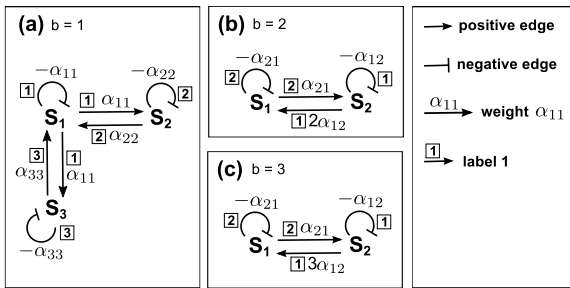


Figure 6.7: R-subgraphs corresponding to amplifying motifs for the CRN example (6.36) taken with NAC kinetics.

The situation changes if we assume weakly monotonic kinetics. Again, both Proposition 6.85 and Proposition 6.86 tell us that multistationarity cannot be excluded, independently of the choice of b . However, for weakly monotonic kinetics, the system only has the capacity for multiple positive steady states if $b < 2$, what

can be seen from the determinant expansion:

$$J^q = \begin{pmatrix} -bV_{11} - V_{14} & V_{22} & V_{33} \\ V_{11} & -V_{22} - V_{25} & 0 \\ V_{11} & 0 & -V_{33} - V_{36} \end{pmatrix}$$

$$\det(J^q) = (2-b)V_{11}V_{22}V_{33} + (1-b)V_{11}V_{22}V_{36} + (1-b)V_{11}V_{25}V_{33} - bV_{11}V_{25}V_{36} \\ - V_{14}V_{22}V_{33} - V_{14}V_{22}V_{36} - V_{14}V_{25}V_{33} - V_{14}V_{25}V_{36}$$

Also based on Proposition 6.83 multiple steady states for $b \geq 2$ can be excluded, as in this case no amplifying motifs exists:

The system has three possible R-subgraphs, one of order 3 ($I = K = \{1, 2, 3\}$), and two of order 2 ($I = K = \{1, 2\}$, and $I = K = \{1, 3\}$; see Figure 6.8). For the first R-subgraph (Figure 6.8(a)), the equation of $N\lambda^1 = (1, 0, 0)^T$ has no solution for $b = 2$, and has the solution $\lambda^1 = (1/(2-b), 1/(2-b), 1/(2-b))^T$ for $b \neq 2$. Given the notation from Lemma 6.75, the index set of all nodes preceding S_1 in a positive cycle is $I^* = \{2, 3\}$, and the amplifier inequality (6.49) reads

$$\frac{1}{2-b} + \frac{1}{2-b} > \left| \frac{-b}{2-b} \right|.$$

The inequality is only satisfied for $0 < b < 2$. Thus, for $b \geq 2$, the given R-subgraph is not an amplifying motif.

For the second R-subgraph (Figure 6.8(b)), the solution of $N\lambda^1 = (1, 0)^T$ is $\lambda^1 = (1/(1-b), 1/(1-b))$ for $b \neq 1$. The amplifier inequality (6.49) reads, with $I^* = \{2\}$,

$$\frac{1}{1-b} > \left| \frac{-b}{1-b} \right|.$$

which is not satisfied for $b > 1$, and thus particularly not for $b \geq 2$.

Analogously, it can be shown that the R-subgraph given in (Figure 6.8(c)) is no amplifying motif for $b \geq 2$.

6.3 Discussion

In this chapter, we provided conditions to predict possible responses to parameter or species concentration changes in steady state when only the network structure is given. In previous work dealing with this subject, general ODE systems were considered whose structure was supposed to be given by the sign structure of the

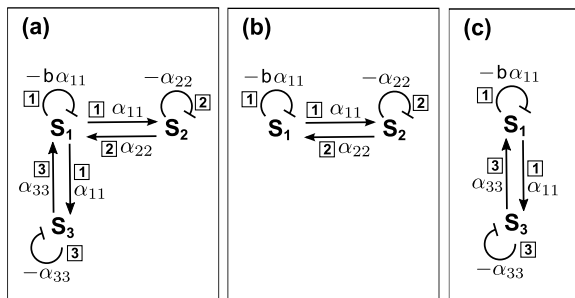


Figure 6.8: R-subgraphs for the CRN example (6.36) taken with weakly monotonic kinetics.

Jacobian matrix, and determinant conditions as well as interaction graph results have been obtained (Maurya et al. 2003; Oyeleye and Kramer 1988). While these were local results derived using linearizations, we used homotopy paths to come up with global conclusions and thus generalized results.

A central issue of the presented work was to transfer the results of perturbation analysis to the class of chemical reaction networks (CRNs). CRNs are characterized by the fact that their Jacobian is a product of the stoichiometric matrix N and the matrix V containing the partial derivatives of the reaction rates. Assuming weakly monotonic or non-autocatalytic (NAC) reaction rates, the qualitative Jacobian matrix of a CRN is given solely by its stoichiometric matrix. Furthermore, CRNs usually contain conservation relations. Given these characteristics, the results for perturbation analysis derived for general ODE systems are of very poor predictive power for CRNs. Thus, we adapted the previous methods, thereby taking into account the special structure of CRNs. As a main result, we came up with a determinant criterion to exclude an inverse response in CRNs (Theorem 6.49).

In addition, we identified network motifs that can cause an inverse response. To this end, we introduced labeled interaction graphs as suited graphical representations of CRNs. Besides from oppositely signed feedforward paths or positive selfcycles, an inverse response can arise from a positive cycle or a combination of cycles fulfilling certain additional structural constraints. In particular, we could show that an amplifying motif, which is a substructure enabling the production of each of its species in steady state, is needed for an inverse response that is not caused by different path signs or positive selfcycles (Theorem 6.78).

A common graphical representation of CRNs are species reaction (SR) graphs and directed species reaction (DSR) graphs. The consideration of labeled interaction graphs enabled us to show how interaction graph results and SR/DSR graph results relate to each other.

A subject that has been widely studied in the context of CRNs is multistationarity (e.g., Banaji and Craciun 2010; Craciun and Feinberg 2005; Shinar and Feinberg 2012; Wiuf and Feliu 2013). Strikingly, we could show that the same network motifs are necessary both for multistationarity and for inverse response. Our presented graphical criteria are stronger than the ones derived by Banaji and Craciun (2009) if the DSR graph contains directed edges. By showing that an amplifying motif is necessary for multistationarity, we were able to exclude multistationarity in systems where purely graph theoretic criteria are not conclusive.

In future research, the relation between our results concerning the amplifying motif and the theory of concordant CRNs (Shinar and Feinberg 2012, 2013) has to be further examined. The concordance property of a network is necessary *and* sufficient for injectivity with regard to certain kinetics. Whether or not a network is concordant can be checked in a convenient way using the Chemical Reaction Network Toolbox [4]. An additional benefit of the amplifying motif, however, is that it is quite intuitive and that it can in some cases be directly checked by network inspection. Nonetheless, efficient computational methods to identify amplifying motifs are still to be developed.

7 | Conclusions

In this thesis, we presented new methods for qualitative modeling and structural analysis of signaling networks and for data-driven network interrogation. With applications to two central signaling pathways for liver regeneration, we showed how these methods can contribute to unravel structure and functioning of cellular signaling systems.

The high complexity of signaling networks demands for scalable computational methods. Furthermore, enhancements of experimental techniques, which enable the collection of large amounts of experimental data, require automated data-driven approaches. Methods that are based on mechanistic modeling using ordinary differential equations are at present not suited for modeling large-scale systems.

In this thesis, we presented, on the one hand, purely qualitative modeling approaches. We developed methods which enable an automated data-driven network interrogation based on interaction graphs (Chapter 3). Using integer linear programming, the methods were already applicable to systems of considerable size. Enhanced power of the approach could be achieved by recent extensions of our presented work based on answer set programming (Thiele et al. 2015). Within the framework of static logical models as presented in Klamt et al. (2006), we derived algorithmic improvements for the computation of minimal intervention sets, thus raising the scalability of this approach (Chapter 4). Furthermore, we showed in that chapter how logical models can be applied to compare the qualitative input–output behavior of a given signaling network structure to experimental high-throughput data in a systematic way. Beyond the methods presented in this thesis, an automated approach for training signaling network structures to experimental data based on predictions of the input–output behavior in logical models has been developed (Saez-Rodriguez et al. 2007).

Besides purely qualitative modeling approaches, we showed the potential of combining qualitative and quantitative mathematical modeling. In Chapter 5, we presented a hybrid modeling approach: Based on interaction graphs, minimal model structures were selected from a vast search space of potential model candidates. In a subsequent step, these models were translated into ODE models for integrating

quantitative details. In Chapter 6, we developed algebraic and graph-theoretic criteria that enable to exclude certain dynamic behavior of variables in an ODE model independently of concrete parameter values. These methods can also be used to preselect model structures and, thus, to narrow down the number of ODE models to be discriminated by parameter estimations.

With this work, we showed the potential of qualitative modeling for the analysis of signaling networks. In particular, we demonstrated how different modeling approaches complement each other, and how results obtained by qualitative modeling methods can serve as a base for more detailed quantitative modeling. Combined modeling approaches are, in our view, indispensable in the future in order to get a holistic understanding of and to accommodate ever increasing networks and mechanistic details in cellular biology. One possibility is to provide formalisms which enable to derive models of different complexity from a common base (e.g., Kolczyk et al. 2012). Additionally, there are tools providing modeling pipelines by featuring different modeling formalisms with increasing complexity within one framework (e.g., Terfve et al. 2012). Future developments of the work presented in this thesis should also consider automated combinations of the shown modeling techniques to come up with integrated modeling pipelines.

A | Integer Linear Programming Formulations for Interaction Graph Training

A.1 Implementation.

The four optimization problems presented in Section 3.2 were implemented by Steffen Klamt and Ioannis Melas in the new software SigNetTrainer. The toolbox is available in two versions. The first is written in C and uses routines from the ILP solver GUROBI [5], whereas the second version is implemented in MATLAB and uses the IBM ILOG CPLEX Optimizer (for which free academic versions can be obtained [7]) as ILP solver. Thus, SigNetTrainer benefits from state-of-the-art-solvers for ILP problems, which use a number of methodologies to deal with large-scale problems. For a more general introduction to ILP algorithms we refer to Bertsimas and Tsitsiklis (1997). Source code and manual of both versions of SigNetTrainer are available for download [3].

A.2 ILP Formulation of Sign Consistency

In Section 3.2.2.1, we stated Constraints 1, 2, and 3, which must be satisfied by a sign-consistent labeling. Constraint 1 is equivalent to

$$\begin{aligned} u_{i,k}^+ &= \max(0, \sigma_i \cdot x_{i_t,k}) \\ u_{i,k}^- &= \max(0, -\sigma_i \cdot x_{i_t,k}). \end{aligned} \tag{A.1}$$

As the max operator is not linear (required for an ILP), we introduce the binary variables $d1_{i,k}, \dots, d4_{i,k}$ to linearize (A.1) in the following way:

$$\begin{aligned}
 u_{i,k}^+ &\geq 0 \\
 u_{i,k}^+ &\geq \sigma_i \cdot x_{i_t,k} \\
 u_{i,k}^+ + 2 d1_{i,k} &\leq 2 \\
 u_{i,k}^+ - \sigma_i \cdot x_{i_t,k} + 2 d2_{i,k} &\leq 2 \\
 d1_{i,k} + d2_{i,k} &= 1 \\
 u_{i,k}^- &\geq 0 \\
 u_{i,k}^- &\geq -\sigma_i \cdot x_{i_t,k} \\
 u_{i,k}^- + 2 d3_{i,k} &\leq 2 \\
 u_{i,k}^- + \sigma_i \cdot x_{i_t,k} + 2 d4_{i,k} &\leq 2 \\
 d3_{i,k} + d4_{i,k} &= 1.
 \end{aligned} \tag{A.2}$$

Let $I_{\rightarrow v_j}$ denote the index set of all edges with head v_j , that is $I_{\rightarrow v_j} = \{i \in I_E : a_i = (v_{i_t}, v_j)\}$. Constraint 2 can be written in terms of linear inequalities as follows:

$$\begin{aligned}
 x_{j,k}^+ &\geq u_{i,k}^+ \quad \forall i \in I_{\rightarrow v_j}, \\
 x_{j,k}^- &\geq u_{i,k}^- \quad \forall i \in I_{\rightarrow v_j}, \\
 x_{j,k}^+ &\leq \sum_{i \in I_{\rightarrow v_j}} u_{i,k}^+ \\
 x_{j,k}^- &\leq \sum_{i \in I_{\rightarrow v_j}} u_{i,k}^-.
 \end{aligned} \tag{A.3}$$

Constraint 3 is equivalent to the following inequalities:

$$\begin{aligned}
 x_{j,k} &\leq x_{j,k}^+ \\
 x_{j,k} &\geq -x_{j,k}^- \\
 x_{j,k} &\leq 2 x_{j,k}^+ - x_{j,k}^- \\
 x_{j,k} &\geq -2 x_{j,k}^- + x_{j,k}^+.
 \end{aligned} \tag{A.4}$$

A.3 ILP Formulation of SCEN_FIT

Computing a single SCEN_FIT solution. Eq. (3.2) in Section 3.2.2.2 describes the optimization problem to find a sign-consistent labeling $x_k = (x_{1,k}, \dots, x_{n_v,k})$ that minimizes the measurement-prediction-mismatch. We introduce $abs_{j,k} = |m_{j,k} - x_{j,k}|$, $abs_{j,k} \in \{0, 1, 2\}$, to reformulate the absolute value in the objective function of Eq. (3.2). Given an experimental scenario k , a SCEN_FIT solution solves the optimization problem

$$\begin{aligned}
 & \underset{x_k}{\text{minimize}} && \sum_{\substack{j \in I_V \\ m_{j,k} \neq \text{NaN}}} abs_{j,k} \\
 & \text{subject to} && abs_{j,k} \geq m_{j,k} - x_{j,k} \quad \forall j \in I_V \text{ with } m_{j,k} \neq \text{NaN}, \\
 & && abs_{j,k} \geq x_{j,k} - m_{j,k} \quad \forall j \in I_V \text{ with } m_{j,k} \neq \text{NaN}, \\
 & && x_{j,k} = p_{j,k} \quad \forall j \in I_V \text{ with } p_{j,k} \neq \text{NaN}, \\
 & && \text{Inequalities (A.4)} \quad \forall j \in I_V \text{ with } p_{j,k} = \text{NaN}, \\
 & && \text{Inequalities (A.3)} \quad \forall j \in I_V \text{ with } p_{j,k} = \text{NaN}, \\
 & && \text{Inequalities (A.2)}.
 \end{aligned} \tag{A.5}$$

Enumeration of SCEN_FIT solutions. To enumerate SCEN_FIT solutions, problem (A.5) is solved repeatedly. After the first iteration, the objective function is replaced by

$$\sum_{\substack{j \in I_V \\ m_{j,k} \neq \text{NaN}}} abs_{j,k} = \text{objval}, \tag{A.6}$$

where objval is the minimum value of the objective function found in the first run of the algorithm. In addition, for each previous solution s , the following constraints are added to (A.5) in order to exclude this solution (reformulation of the nonlinear Eq. (3.3)):

$$\begin{aligned}
 & \sum_{j \in I_V} dx_{j,k,s} \geq 1 \\
 & -x_{j,k} + dx_{j,k,s} - 4 dx_{2j,k,s} \leq x_{j,k,s} \\
 & x_{j,k} + dx_{j,k,s} - 4 dx_{1j,k,s} \leq -x_{j,k,s} \\
 & dx_{1j,k,s} + dx_{2j,k,s} = 1,
 \end{aligned} \tag{A.7}$$

The variables $dx_{j,k,s}$ (integer), $dx_{1j,k,s}$ (binary), and $dx_{2j,k,s}$ (binary) are auxiliary variables.

A.4 ILP Formulation of MCoS

Computing a single MCoS. To compute an MCoS for a sign-inconsistent scenario k , we introduce the binary variables $B_{j,k}^+$ and $B_{j,k}^-$ representing artificial perturbations. Inequalities (A.4) are modified as follows (cf. Constraint 4, Section 3.2.2.3):

$$\begin{aligned}
 x_{j,k} &\leq x_{j,k}^+ + B_{j,k}^+ \\
 x_{j,k} &\geq -x_{j,k}^- - B_{j,k}^- \\
 x_{j,k} &\leq 2x_{j,k}^+ - x_{j,k}^- + 2B_{j,k}^+ \\
 x_{j,k} &\geq -2x_{j,k}^- + x_{j,k}^+ - 2B_{j,k}^-.
 \end{aligned} \tag{A.8}$$

An MCoS for a given experimental scenario k solves the optimization problem

$$\begin{aligned}
 &\underset{B_k^-, B_k^+}{\text{minimize}} && \sum_{j \in I_V} (B_{j,k}^+ + B_{j,k}^-) \\
 &\text{subject to} && \sum_{\substack{j \in I_V \\ m_{j,k} \neq \text{NaN}}} \text{abs}_{j,k} = 0, \\
 &&& \text{abs}_{j,k} \geq m_{j,k} - x_{j,k} \quad \forall j \in I_V \text{ with } m_{j,k} \neq \text{NaN}, \\
 &&& \text{abs}_{j,k} \geq x_{j,k} - m_{j,k} \quad \forall j \in I_V \text{ with } m_{j,k} \neq \text{NaN}, \\
 &&& x_{j,k} = p_{j,k} \quad \forall j \in I_V \text{ with } p_{j,k} \neq \text{NaN}, \\
 &&& \text{Inequalities (A.8)} \quad \forall j \in I_V \text{ with } p_{j,k} = \text{NaN}, \\
 &&& \text{Inequalities (A.3)} \quad \forall j \in I_V \text{ with } p_{j,k} = \text{NaN}, \\
 &&& \text{Inequalities (A.2)}.
 \end{aligned} \tag{A.9}$$

Enumeration of MCoSs. To enumerate MCoSs, problem (A.9) is solved repeatedly. After the first iteration, the objective function is replaced by

$$\sum_{j \in I_V} (B_{j,k}^+ + B_{j,k}^-) = \text{objval}, \tag{A.10}$$

where objval is the minimum value of the objective function found in the first run of the algorithm. In addition, for each previous solution s , the following constraints are added to (A.9) in order to exclude this solution (reformulation of the nonlinear Eq. (3.7)):

$$\sum_{\substack{j \in I_V \\ B_{j,k,s}^+ = 0}} B_{j,k}^+ + \sum_{\substack{j \in I_V \\ B_{j,k,s}^- = 0}} B_{j,k}^- - \sum_{\substack{j \in I_V \\ B_{j,k,s}^+ = 1}} (B_{j,k}^+ - 1) - \sum_{\substack{j \in I_V \\ B_{j,k,s}^- = 1}} (B_{j,k}^- - 1) \geq 1. \quad (\text{A.11})$$

A.5 ILP Formulation of OPT_SUBGRAPH

Computing a single optimal subgraph. The removal of edges is implemented using binary values y_i and modifying inequalities (A.2) as follows:

$$\begin{aligned} u_{i,k}^+ &\geq 0 \\ u_{i,k}^+ &\geq \sigma_i \cdot x_{i_t,k} - y_i \\ u_{i,k}^+ + 3 d1_{i,k} &\leq 3 \\ u_{i,k}^+ + y_i - \sigma_i \cdot x_{i_t,k} + 3 d2_{i,k} &\leq 3 \\ d1_{i,k} + d2_{i,k} &= 1 \\ u_{i,k}^- &\geq 0 \\ u_{i,k}^- &\geq -\sigma_i \cdot x_{i_t,k} - y_i \\ u_{i,k}^- + 3 d3_{i,k} &\leq 3 \\ u_{i,k}^- + y_i + \sigma_i \cdot x_{i_t,k} + 3 d4_{i,k} &\leq 3 \\ d3_{i,k} + d4_{i,k} &= 1. \end{aligned} \quad (\text{A.12})$$

An optimal subgraph of the given interaction graph solves the optimization problem

$$\begin{aligned} &\underset{x_1, \dots, x_{n_S}}{\text{minimize}} && \sum_{\substack{(j,k) \in I_V \times I_S \\ m_{j,k} \neq \text{NaN}}} \text{abs}_{j,k} \\ &\text{subject to} && \text{abs}_{j,k} \geq m_{j,k} - x_{j,k} \quad \forall (j,k) \in I_V \times I_S \text{ with } m_{j,k} \neq \text{NaN}, \\ & && \text{abs}_{j,k} \geq x_{j,k} - m_{j,k} \quad \forall (j,k) \in I_V \times I_S \text{ with } m_{j,k} \neq \text{NaN}, \\ & && x_{j,k} = p_{j,k} \quad \forall (j,k) \in I_V \times I_S \text{ with } p_{j,k} \neq \text{NaN}, \\ & && \text{Inequalities (A.4)} \quad \forall (j,k) \in I_V \times I_S \text{ with } p_{j,k} = \text{NaN}, \\ & && \text{Inequalities (A.3)} \quad \forall (j,k) \in I_V \times I_S \text{ with } p_{j,k} = \text{NaN}, \\ & && \text{Inequalities (A.12)}. \end{aligned} \quad (\text{A.13})$$

Enumeration of optimal subgraphs. To enumerate optimal subgraphs, problem (A.13) is solved repeatedly. After the first iteration, the objective function is replaced by

$$\sum_{\substack{(j,k) \in I_V \times I_S \\ m_{j,k} \neq \text{NaN}}} abs_{j,k} = \text{objval}, \quad (\text{A.14})$$

where objval is the minimum value of the objective function found in the first run of the algorithm. In addition, for each previous solution s , the following constraints are added to (A.13) in order to exclude this solution (reformulation of the nonlinear Eq. (3.10)):

$$\sum_{\substack{i \in I_E \\ y_{i,s} = 0}} y_i - \sum_{\substack{i \in I_E \\ y_{i,s} = 1}} (y_i - 1) \geq 1. \quad (\text{A.15})$$

B | Model Compression for Interaction Graphs

Rule 1 (removal of non-controllable and non-observable nodes, cf. Saez-Rodriguez et al. (2009)): *Non-controllable* nodes (which cannot be affected by any of the perturbed nodes in any scenario) and *non-observable* nodes (which do not influence any measured (readout) node in any scenario) define non-identifiable parts of the network. Therefore, these nodes as well as all edges they are connected to can be removed. Non-observable and non-controllable nodes can easily be identified by shortest path algorithms.

Rule 2 (removal of parallel edges): If there are two parallel edges of the same sign, we may safely remove one of them (Figure B.1(a)).

Rule 3 (absorbing a node with a single input edge): If a latent node (neither measured nor perturbed in any of the experimental scenarios) has only one single incoming edge, then we can remove this node together with the incoming edge and reconnect all the outgoing edges of this node to its only predecessor node (under consideration of edge signs; Figure B.1(b)).

Rule 4 (absorbing a node with a single output edge): If a latent node has only one single outgoing edge, then we can remove this node together with the outgoing edge and reconnect all its incoming edges to its only successor node (under consideration of edge signs; Figure B.1(c)).

Rule 1 is performed once at the beginning, whereas Rules 2–4 are iteratively used until no further rule can be applied (note that new parallel edges may arise after applying Rules 3 or 4).

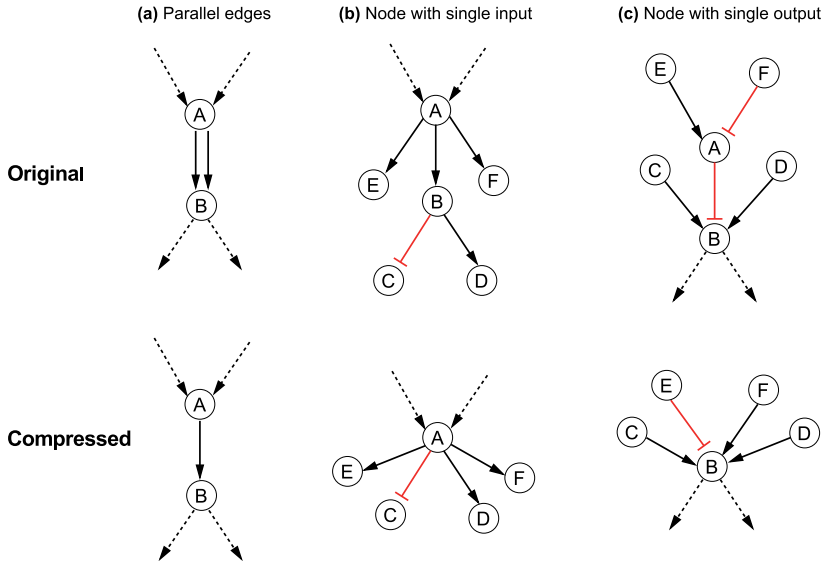


Figure B.1: Basic network compression rules for interaction graphs. (a) Parallel edges (Rule 2). **(b)** Nodes with single input (Rule 3). **(c)** Nodes with single output (Rule 4). Figure adapted from Melas, Samaga et al. (2013).

C | Documentation of ERBB Models

C.1 List of Species

The following list shows the species of the logical ERBB model (M1). It contains two types of auxiliary species that have no biological correspondents, *dummy* species and *reservoirs*.

Dummy species are necessary if one wants to assign different time variables to interactions forming one hyperarc. A typical example is $A \cdot !B \rightarrow C$, describing that both the presence of A and absence of B are necessary to activate C. In order to reflect a time delay of the inhibitory activity of B, we introduce the species B_dummy and describe the original interaction with two interactions $B \rightarrow B_dummy$ (time 2) and $A \cdot !B_dummy$ (time 1).

A reservoir is introduced whenever a molecule causes different downstream events depending on how it is activated. Here, we have to use more than one compound to describe the molecule in the model. However, as all these compounds represent the same biological species, we associate them with a reservoir, pointing out that they share the same pool. Inactivation of the reservoir will then affect the activation of all correspondents of this species.

Dummy species and reservoirs are only needed for the logical modeling formalism. However, as we derived the interaction graph model from the logical model, the interaction graph contains the same set of species. Removing dummy nodes and reservoirs from the interaction graph and compressing the respective interactions would lead to the same results for the interaction graph analysis.

Model variant M2 of the logical model contains the same set of species as model M1, except for some dummy species (see Appendix C.4).

C Documentation of ERBB Models

No.	Model Name	Full Name	Comment
1	actin_reorg	actin reorganization	symbolizes actin reorganizing effects initiated by LIMK1
2	akt	RAC-alpha serine/threonine-protein kinase	also known as protein kinase B (PKB)
3	aktd		dummy species
4	ap1	Activator protein 1	heterodimer composed of c-Jun and c-Fos
5	pro_apoptotic		symbolizes the pro-apoptotic effect of BAD
6	ar	Amphiregulin	ligand that binds specifically ERBB1
7	bad	Bcl2-associated agonist of cell death	phosphorylated (the unphosphorylated form promotes apoptosis)
8	bir	Biregulin	synthetic Neuregulin/EGF chimera
9	btc	Betacellulin	ligand with dual specificity, binds ERBB1 and ERBB4
10	ca	Calcium	cytosolic Ca ²⁺ -ions
11	ccbl	E3 ubiquitin-protein ligase CBL	
12	cfos	Proto-oncogene c-Fos	
13	cjun	Proto-oncogene c-Jun	
14	cmyc	Myc proto-oncogene protein	
15	creb	Cyclic AMP-responsive element-binding protein	
16	csrc	Neuronal proto-oncogene tyrosine-protein kinase Src	
17	dag	Diacylglycerol	
18	egf	Epidermal growth factor	ligand that binds specifically ERBB1
19	elk1	ETS domain-containing protein Elk-1	
20	endocyt_degrad		symbolizes endocytosis/degradation of the receptors

Continued on next page

No.	Model Name	Full Name	Comment
21	epr	Epiregulin	ligand with dual specificity, binds specifically ERBB1 and ERBB4
22	eps8r	Epidermal growth factor receptor kinase substrate 8	reservoir of EPS8
23	erbb1	Epidermal growth factor receptor	also known as EGFR
24	erbb11		homodimer composed of two ERBB1 receptors
25	erbb12		heterodimer composed of ERBB1 and ERBB2
26	erbb13		heterodimer composed of ERBB1 and ERBB3
27	erbb14		heterodimer composed of ERBB1 and ERBB4
28	erbb2	Receptor tyrosine-protein kinase erbB-2	no ligand, preferred heterodimerization partner of the other ERBBs; also known as Proto-oncogene Neu
29	erbb23		heterodimer composed of ERBB2 and ERBB3
30	erbb24		heterodimer composed of ERBB2 and ERBB4
31	erbb3	Receptor tyrosine-protein kinase erbB-3	kinase-defective
32	erbb34		heterodimer composed of ERBB3 and ERBB4
33	erbb4	Receptor tyrosine-protein kinase erbB-4	
34	erbb44		homodimer composed of two ERBB4 receptors
35	erk12	Extracellular signal-regulated kinase 1 (ERK1) or 2 (ERK2)	since both ERK1 and ERK2 catalyze the same reactions, we do not distinguish between them; also known as MAPK1/3
36	gab1	GRB2-associated-binding protein 1	

Continued on next page

C Documentation of ERBB Models

No.	Model Name	Full Name	Comment
37	grb2	Growth factor receptor-bound protein 2	
38	gsk3	Glycogen synthase kinase-3 beta	
39	hbegf	Heparin-binding EGF-like growth factor	ligand with dual specificity, bind ERBB1 and ERBB4
40	hsp27	Heat shock protein beta-1	
41	ip3	Inositol 1,4,5-trisphosphate	
42	jnk	c-Jun N-terminal kinase 1	also known as MAPK8
43	limk1	LIM domain kinase 1	
44	mek12	MAPK/ERK kinase 1 or 2	since both MEK1 and MEK2 catalyze the same reactions, we do not distinguish between them; also known as MAPKK1/MAPKK2
45	mekk1	MAPK/ERK kinase kinase 1	also known as MAPKKK1
46	mekk4	MAPK/ERK kinase kinase 4	also known as MAPKKK4
47	mk2	MAP kinase-activated protein kinase 2	
48	mkk3	Dual specificity mitogen-activated protein kinase kinase 3	also known as MAPKK3
49	mkk4	Dual specificity mitogen-activated protein kinase kinase 4	also known as MAPKK4
50	mkk6	Dual specificity mitogen-activated protein kinase kinase 6	also known as MAPKK6
51	mkk7	Dual specificity mitogen-activated protein kinase kinase 7	also known as MAPKK7
52	mkp	Dual specificity protein phosphatase	
53	mlk3	Mixed lineage kinase 3	

Continued on next page

No.	Model Name	Full Name	Comment
54	mtorr	Mammalian target of rapamycin	reservoir of mTOR
55	mtor_rap		mTOR-raptor complex
56	mtor_ric		mTOR-rictor complex
57	nck	Cytoplasmic protein NCK	
58	nrg1a	Neuregulin-1 alpha	ligand with dual specificity, binds ERBB3 and ERBB4
59	nrg1b	Neuregulin-1 beta	ligand with dual specificity, binds ERBB3 and ERBB4
60	nrg2a	Neuregulin-2 alpha	ligand with dual specificity, binds ERBB3 and ERBB4
61	nrg2b	Neuregulin-2 beta	ligand with dual specificity, binds ERBB3 and ERBB4
62	nrg3	Neuregulin-3	ligand that binds specifically ERBB4
63	nrg4	Neuregulin-4	ligand that binds specifically ERBB4
64	nucerk12		phosphorylated dimer of ERK1 or ERK2, located in the nucleus
65	p38	Mitogen-activated protein kinase 14	
66	p70s6k_1	70 kDa ribosomal protein S6 kinase 1	p70S6K phosphorylated at autoinhibitory sites
67	p70s6k_2		p70S6K phosphorylated at the catalytic sites T389 and S229
68	p90rsk	90 kDa ribosomal protein S6 kinase	also known as S6K-alpha, p90S6K
69	p90rskerk12d		dummy species
70	pak1	p21-activated kinase 1	
71	pdk1	3-phosphoinositide-dependent protein kinase 1	
72	pi34p2	Phosphatidylinositol (3,4)-bisphosphate (PtdIns(3,4)P ₂)	

Continued on next page

C Documentation of ERBB Models

No.	Model Name	Full Name	Comment
73	pi3k	Phosphatidylinositol-4,5-bisphosphate 3-kinase	
74	pi3kr		reservoir of PI3K
75	pip3	Phosphatidylinositol (3,4,5)-trisphosphate	
76	pkc	Protein kinase C	
77	plcg	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma	
78	pp2a	Serine/threonine-protein phosphatase 2A	
79	pp2b	Serine/threonine-protein phosphatase 2B	also known as Calcineurin
80	pten	Phosphatase and tensin homolog	
81	ptend		dummy species
82	rab5a	Ras-related protein Rab-5A	
83	rac_cdc42	Ras-related C3 botulinum toxin substrate 1/ Cell division control protein 42 homolog	Rac1 or CDC42
84	raf1	RAF proto-oncogene serine/threonine-protein kinase	also known as cRaf
85	ras	GTPase Ras	
86	rasgap	Ras GTPase-activating protein 1	
87	rheb	GTP-binding protein Rheb	
88	rin1	Ras and Rab interactor 1	
89	rntre	Related to the N-terminus of tre	
90	shc	SHC-transforming protein 1	

Continued on next page

C.2 List of Interactions of the Logical ERBB Model (M1)

No.	Model Name	Full Name	Comment
91	ship2	SH2 domain-containing inositol 5'-phosphatase 2	
92	ship2d		dummy species
93	shp1	Tyrosine-protein phosphatase non-receptor type 6	also known as PTP-1C
94	shp1d		dummy species
95	shp2	Tyrosine-protein phosphatase non-receptor type 11	also known as PTP-1D
96	sos1	Son of sevenless homolog 1	
97	sos1r		reservoir of SOS1
98	sos1_eps8_e3b1		complex of SOS1, EPS8, and E3B1/ABI1 (Abl interactor 1)
99	stat1	Signal transducer and activator of transcription 1	
100	stat3	Signal transducer and activator of transcription 3	
101	stat5	Signal transducer and activator of transcription 5	
102	tgfa	Transforming growth factor alpha	ligand that binds specifically ERBB1
103	tsc1_tsc2	Tuberous sclerosis 1/2 protein	complex of TSC1 (also known as Hamartin) and TSC2 (also known as Tuberin)
104	vav2	Guanine nucleotide exchange factor VAV2	

C.2 List of Interactions of the Logical ERBB Model (M1)

In the following, we list the interactions included in the logical ERBB Model (M1). The interactions of the interaction graph model can be derived from these interactions by splitting the hyperarcs and subsequent removal of parallel edges, as de-

scribed in Section 2.4.2.1.

Notation

- $\rightarrow A$ species A is an input to the model
- $A \rightarrow$ species A is an output of the model
- $A \rightarrow B$ species A activates species B
- $A \cdot B \rightarrow C$ species A AND B activate C (and both A and B are necessary for activation)
- $A \cdot !B \rightarrow C$ species C is activated when A AND NOT B are present

Inputs

No.	Interaction	Value	Documentation
1	\rightarrow erbb1	1	The four members of the ErbB receptor family.
2	\rightarrow erbb2	1	
3	\rightarrow erbb3	1	
4	\rightarrow erbb4	1	
5	\rightarrow ar	0	Ligands of the EGF ligand family.
6	\rightarrow bir	0	
7	\rightarrow btc	0	
8	\rightarrow egf	0	
9	\rightarrow epr	0	
10	\rightarrow hbegef	0	
11	\rightarrow nrg1a	0	
12	\rightarrow nrg1b	0	
13	\rightarrow nrg2a	0	
14	\rightarrow nrg2b	0	
15	\rightarrow nrg3	0	
16	\rightarrow nrg4	0	
17	\rightarrow tgfa	0	
18	\rightarrow csrc	1	It is not clear how c-Src is activated. Whereas in Sato et al. (2002), activation of c-Src by SHC (in response to EGF) is reported that leads to phosphorylation of STAT, in Olayioye et al. (1999), the activation of STAT after EGF stimulation was more rapid than activation of c-Src in response to EGF.
19	\rightarrow pten	0	PTEN and SHIP2 both downregulate PIP ₃ synthesis. As we could not find any information on PTEN and SHIP2 regulation, we included them as inputs in the model.
20	\rightarrow ship2	0	
21	\rightarrow pdk1	1	PDK1 appears to be constitutively active (Newton 2003).

Continued on next page

No.	Interaction	Value	Documentation
22	→ pp2a	0	
23	→ pp2b	0	Regulation of these phosphatases is not considered in the model.
24	→ mkp	0	
25	→ sos1r	1	We consider two different pools of SOS1: activated through GRB2, leading to Ras-GEF activity (interaction 112), and in a complex with EPS8 and E3B1 (interaction 118), leading to Rac-GEF activity.
26	→ eps8r	1	EPS8 can form a complex with SOS1/E3B1 (interaction 118) or RN-tre (interaction 183).
27	→ pi3kr	1	This reservoir represents the inactive form of PI3K that participates in the activation of SOS1_EPS8_E3B1 (see interaction 118).
28	→ mtorr	1	mTOR can complex with rictor (see interaction 134) or raptor (see interaction 139). Therefore, a reservoir of mTOR is included in the model.

Activation of ErbB dimers

The four ERBB receptors form different homo- and heterodimers. Binding of a ligand leads to autophosphorylation of tyrosine residues that provide docking sites for proteins with SH2 or PTB domains. As ERBB2 does not bind to ligands of the EGF family (Citri and Yarden 2006), ERBB2 can only be activated in heterodimers. However, ERBB2 is the preferred heterodimerization partner of the other ERBBs; thus, the other heterodimers 13, 14, and 34 are only formed in absence of ERBB2. An exception are AR and HB-EGF activated dimers: AR activates ERBB3, but not ERBB2 (Beerli and Hynes 1996), so we assume that AR activates ERBB13 dimers also in presence of ERBB2. ERBB3 is kinase-defective, thus ERBB3-homodimers are inactive (Citri and Yarden 2006; Olayioye et al. 2000). According to Landau and Ben-Tal (2008), only one of the receptors in a dimer is phosphorylated. In ERB1/ERBB2 dimers, only ERBB2 becomes phosphorylated (Landau and Ben-Tal 2008). As ERBB3 is kinase-defective, ERBB3 is the phosphorylated partner in ERBB3 heterodimers (Landau and Ben-Tal 2008). We could not find any information on the phosphorylation of ERBB1/ERBB4 and ERBB2/ERBB4 dimers. Therefore, we assumed that proteins that can only bind to ERBB1 are activated through ERBB1 homodimers; proteins that can bind to ERBB2 are activated through ERBB1/ERBB2 dimers; proteins that can bind to ERBB3 are activated through all possible ERBB3 dimers (13, 23, 34), and proteins that can bind to ERBB4 are activated through ERBB4 homodimers.

C Documentation of ERBB Models

No.	Interaction	Time	Documentation
29	erbb11 → shp1	1	SHP1 binds to ERBB1 at phosphorylated Y1173 (Keilhack et al. 1998).
30	shp1 → shp1d	2	SHP1 dephosphorylates the ERBB1 dimers (negative feedback; see, e.g., interaction 31). We assume that this dephosphorylation is a late event and, thus, included a dummy species.
31	ar · !shp1d · erbb1 → erbb11	1	AR binds ERBB1 homodimers and ERBB13.
32	ar · !shp1d · erbb1 · erbb3 → erbb13	1	However, the affinity of AR towards ERBB1 is significantly lower than the affinity of EGF (Beerli and Hynes 1996). Interaction 32 is not included in Oda et al. (2005).
33	bir · !shp1d · erbb1 → erbb11	1	BIR activates the following ERBB dimers: 11, 12, 23, 24, 44 (Jones et al. 1999). Since in Jones et al. (1999)
34	bir · !shp1d · erbb1 · erbb2 → erbb12	1	the dimers 13, 14, and 34 are not analyzed, and this is the only source about binding affinities for BIR we
35	bir · erbb2 · erbb3 → erbb23	1	found, we cannot rule out the possibility that 13, 14,
36	bir · erbb2 · erbb4 → erbb24	1	and 34 are also activated. As BIR is an artificial
37	bir · erbb4 → erbb44	1	ligand, one could think about not considering it in the model.
38	btc · !shp1d · erbb1 → erbb11	1	BTC activates the following ERBB dimers: 11, 12, 24, 44 (Jones et al. 1999). Additionally, it activates 13
39	btc · !shp1d · erbb1 · erbb2 → erbb12	1	(Alroy and Yarden 1997) (not part of Oda et al. (2005)). In Alroy and Yarden (1997), activation of 14
40	btc · !shp1d · erbb1 · !erbb2 · erbb3 → erbb13	1	was detected, whereas this is not reported in Graus-Porta et al. (1997). Therefore, we decided not
41	btc · erbb2 · erbb3 → erbb23	1	to include activation of 14. In Wang et al. (1998) and
42	btc · erbb2 · erbb4 → erbb24	1	Graus-Porta et al. (1997), activation of 23 is reported,
43	btc · erbb4 → erbb44	1	what is contradictory to Jones et al. (1999) and not mentioned in Alroy and Yarden (1997). However, it is in accordance with the findings of Beerli and Hynes (1996), which state that BTC activates ERBB3 when all ERBB receptors are present; thus, we included it in the model.

Continued on next page

C.2 List of Interactions of the Logical ERBB Model (M1)

No.	Interaction	Time	Documentation
44	egf · !shp1d · erbb1 → erbb11	1	EGF activates the following ERBB dimers: 11, 12, 24 (Jones et al. 1999). In absence of ERBB2, also 13 and
45	egf · !shp1d · erbb1 · erbb2 → erbb12	1	14 can be activated (Graus-Porta et al. 1997; Olayioye et al. 1998). Activation of 13 and 14 is not included
46	egf · !shp1d · erbb1 · !erbb2 · erbb3 → erbb13	1	in Oda et al. (2005). Furthermore, in Wang et al. (1998) and Graus-Porta et al. (1997), activation of 23
47	egf · !shp1d · erbb1 · !erbb2 · erbb4 → erbb14	1	is mentioned. We decided not to consider this in the model for two reasons: First, in Jones et al. (1999), 23
48	egf · erbb2 · erbb4 → erbb24	1	dimers were studied, but no measurable binding of EGF was detected. Second, in Shelly et al. (1998) it is stated that this activation occurs only at very high ligand concentrations.
49	epr · !shp1d · erbb1 → erbb11	1	EPR activates the following ERBB dimers: 11, 12, 23, 24 (Jones et al. 1999). Additionally, activation of 13,
50	epr · !shp1d · erbb1 · erbb2 → erbb12	1	14 (high), and 34, 44 (low) is mentioned in Shelly et al. (1998) (none of these part of the map of Oda
51	epr · !shp1d · erbb1 · erbb3 · !erbb2 → erbb13	1	et al. (2005)). We included activation of 13 and 14 in the model, because these heterodimers are not part
52	epr · !shp1d · erbb1 · erbb4 · !erbb2 → erbb14	1	of the analysis in Jones et al. (1999) and, thus, the results in Jones et al. (1999) and Shelly et al. (1998)
53	epr · erbb2 · erbb3 → erbb23	1	are not contradictory. Activation of 44 was not considered, because this is contradictory to Jones
54	epr · erbb2 · erbb4 → erbb24	1	et al. (1999) and, furthermore, was mentioned as low in Shelly et al. (1998). The interaction with 34 was also not included in the model as it was mentioned as low.
55	hbegf · !shp1d · erbb1 → erbb11	1	HB-EGF activates the following ERBB dimers: 11, 12, 24 (Jones et al. 1999). In Beerli and Hynes (1996),
56	hbegf · !shp1d · erbb1 · erbb2 → erbb12	1	activation of ERBB3 in response to HB-EGF is stated—in presence of ERBB2. One possibility is that
57	hbegf · erbb2 · erbb4 → erbb24	1	HB-EGF activates ERBB13 dimers even in presence of ERBB2. Alternatively, HB-EGF might activate ERBB23 dimers, in contradiction to the findings in Jones et al. (1999).

Continued on next page

C Documentation of ERBB Models

No.	Interaction	Time	Documentation
58	nrg1a · !shp1d · erbb1 · erbb3 · !erbb2 → erbb13	1	
59	nrg1a · !shp1d · erbb1 · erbb4 · !erbb2 → erbb14	1	NRG1 α activates the following ERBB dimers: 23, 24, 44. Furthermore (not mentioned in Oda et al. (2005)),
60	nrg1a · erbb2 · erbb3 → erbb23	1	it activates 13 (Alroy and Yarden 1997; Graus-Porta et al. 1997; Olayioye et al. 1998), 14 (Alroy and
61	nrg1a · erbb2 · erbb4 → erbb24	1	Yarden 1997; Graus-Porta et al. 1997; Olayioye et al. 1998; Pinkas-Kramarski et al. 1998) and 34 (Alroy
62	nrg1a · erbb3 · erbb4 · !erbb2 → erbb34	1	and Yarden 1997; Pinkas-Kramarski et al. 1998).
63	nrg1a · erbb4 → erbb44	1	
64	nrg1b · !shp1d · erbb1 · erbb3 · !erbb2 → erbb13	1	
65	nrg1b · !shp1d · erbb1 · erbb4 · !erbb2 → erbb14	1	NRG1 β activates the following ERBB dimers: 23, 24,
66	nrg1b · erbb2 · erbb3 → erbb23	1	44 (Jones et al. 1999). In accordance with Pinkas-Kramarski et al. (1998) and Alroy and Yarden
67	nrg1b · erbb2 · erbb4 → erbb24	1	(1997), also 13, 14, and 34 can be activated (not depicted in Oda et al. (2005)).
68	nrg1b · erbb3 · erbb4 · !erbb2 → erbb34	1	
69	nrg1b · erbb4 → erbb44	1	
70	nrg2a · !shp1d · erbb1 · erbb3 · !erbb2 → erbb13	1	NRG2 α activates the ERBB dimer 24 (Jones et al. 1999). Additionally, 13, 14 and 34 are activated
71	nrg2a · !shp1d · erbb1 · erbb4 · !erbb2 → erbb14	1	(Pinkas-Kramarski et al. 1998) (not included in Oda et al. (2005)). In Pinkas-Kramarski et al. (1998),
72	nrg2a · erbb2 · erbb4 → erbb24	1	activation of 11, 12, 23, and 44 is also reported. We decided not to include this in the model, since these
73	nrg2a · erbb3 · erbb4 · !erbb2 → erbb34	1	interactions are in contradiction to Jones et al. (1999).
74	nrg2b · !shp1d · erbb1 · erbb4 · !erbb2 → erbb14	1	
75	nrg2b · erbb2 · erbb3 → erbb23	1	In accordance with Jones et al. (1999), NRG2 β activates the following ERBB dimers: 23, 24, 44.
76	nrg2b · erbb2 · erbb4 → erbb24	1	Furthermore, activation of 14 and 34 is reported in Pinkas-Kramarski et al. (1998) (not depicted in Oda
77	nrg2b · erbb3 · erbb4 · !erbb2 → erbb34	1	et al. (2005)).
78	nrg2b · erbb4 → erbb44	1	

Continued on next page

C.2 List of Interactions of the Logical ERBB Model (M1)

No.	Interaction	Time	Documentation
79	$\text{nr}g3 \cdot \text{er}bb2 \cdot \text{er}bb4 \rightarrow \text{er}bb24$	1	NRG3 activates the following ERBB dimers: 24, 44 (Jones et al. 1999). However, in Jones et al. (1999) 13, 14, and 34 are not analyzed. Therefore, and since we were not able to find another reference dealing with binding specificities of NRG3, we cannot be sure that there is no interaction between NRG3 and these dimers.
80	$\text{nr}g3 \cdot \text{er}bb4 \rightarrow \text{er}bb44$	1	
81	$\text{nr}g4 \cdot \text{!shp}1d \cdot \text{er}bb1 \cdot \text{er}bb4 \cdot \text{!er}bb2 \rightarrow \text{er}bb14$	1	NRG4 activates the following ERBB dimers: 14, 24, 44 (Harari et al. 1999). In Oda et al. (2005) also activation of 34 is depicted. However, we could not find a source where this is described; thus, we decided not to include it in the model.
82	$\text{nr}g4 \cdot \text{er}bb2 \cdot \text{er}bb4 \rightarrow \text{er}bb24$	1	
83	$\text{nr}g4 \cdot \text{er}bb4 \rightarrow \text{er}bb44$	1	
84	$\text{tgfa} \cdot \text{!shp}1d \cdot \text{er}bb1 \rightarrow \text{er}bb11$	1	TGF α activates the following ERBB dimers: 11, 12, 24 (Jones et al. 1999). In Alroy and Yarden (1997), activation of 13 and 14 (not part of the map of Oda et al. (2005)) is mentioned.
85	$\text{tgfa} \cdot \text{!shp}1d \cdot \text{er}bb1 \cdot \text{er}bb2 \rightarrow \text{er}bb12$	1	
86	$\text{tgfa} \cdot \text{!shp}1d \cdot \text{er}bb1 \cdot \text{!er}bb2 \cdot \text{er}bb3 \rightarrow \text{er}bb13$	1	
87	$\text{tgfa} \cdot \text{!shp}1d \cdot \text{er}bb1 \cdot \text{!er}bb2 \cdot \text{er}bb4 \rightarrow \text{er}bb14$	1	
88	$\text{tgfa} \cdot \text{er}bb2 \cdot \text{er}bb4 \rightarrow \text{er}bb24$	1	

Activation of adapter proteins

No.	Interaction	Time	Documentation
89	$\text{er}bb11 \rightarrow \text{shc}$	1	SHC binds to all types of ERBB receptors. On ERBB1, the binding sites are pY1148 (via PTB domain) and pY1173 (via PTB and SH2 domain) (Olayioye et al. 2000).
90	$\text{er}bb12 \rightarrow \text{shc}$	1	
91	$\text{er}bb13 \rightarrow \text{shc}$	1	
92	$\text{er}bb14 \rightarrow \text{shc}$	1	
93	$\text{er}bb23 \rightarrow \text{shc}$	1	
94	$\text{er}bb24 \rightarrow \text{shc}$	1	
95	$\text{er}bb34 \rightarrow \text{shc}$	1	
96	$\text{er}bb44 \rightarrow \text{shc}$	1	

Continued on next page

No.	Interaction	Time	Documentation
97	shc → grb2	1	
98	erbb11 → grb2	1	
99	erbb12 → grb2	1	GRB2 can bind to ERBB dimers via SHC (Okabayashi et al. 1994), or directly via its SH2 domain. As there are binding sites for GRB2 on all ERBB receptors (Schulze et al. 2005), we assumed that GRB2 directly interacts with all possible ERBB dimers.
100	erbb13 → grb2	1	
101	erbb14 → grb2	1	
102	erbb23 → grb2	1	
103	erbb24 → grb2	1	
104	erbb34 → grb2	1	
105	erbb44 → grb2	1	
106	erbb11 → gab1	1	GAB1 can bind directly to ERBB1 receptors or via GRB2 and is phosphorylated on Y-residues by the receptor kinase (Rodrigues et al. 2000).
107	grb2 → gab1	1	
108	erbb11 → nck	1	NCK binds to EGFR (Li et al. 2001). In Schulze et al. (2005), a NCK binding site on ERBB4 is reported.
109	erbb14 → nck	1	
110	erbb44 → nck	1	

Activation of the G-proteins Ras and Rac

No.	Interaction	Time	Documentation
111	p90rsk · erk12 → p90rskerk12d	2	Interaction that was introduced for modeling the time delay p90RSK and ERK1/2 phosphorylate and thus inhibit SOS1 with (see interaction 102).
112	sos1r · grb2 · !p90rskerk12d → sos1	1	SOS1 is bound to GRB2. Binding of GRB2 to EGFR, either directly or indirectly through SHC, leads to activation of SOS1 (Buday and Downward 1993). Serine/threonine phosphorylation of SOS1 by p90RSK or ERK1/2 causes dissociation of GRB2-SOS1 from SHC or from the phosphorylated receptor (Douville and Downward 1997).
113	gab1 → shp2	1	Phosphorylated GAB1 recruits and activates SHP2 (Montagner et al. 2005).
114	gab1 · !shp2 → rasgap	1	RasGAP can bind tyrosine phosphorylation sites on GAB1, GAB1-bound SHP2 dephosphorylates these sites (Montagner et al. 2005) . Note that this reaction is not part of the map of Oda et al. (2005).

Continued on next page

C.2 List of Interactions of the Logical ERBB Model (M1)

No.	Interaction	Time	Documentation
115	$\text{sos1} \cdot \text{lrasgap} \rightarrow \text{ras}$	1	Interaction with SOS1 increases the rate of GDP/GTP exchange of Ras (Li et al. 1993). The GTPase activating protein RasGAP leads to hydrolysis of Ras-bound GTP to GDP (Cox and Der 2003).
116	$\text{erbb11} \cdot \text{pip3} \rightarrow \text{vav2}$	1	The SH2 domain of VAV2 binds to pY992 or pY1148 on ERBB1. The receptor phosphorylates VAV2 on Y142, Y159, and Y172. Both PIP ₃ and PtdIns(3,4)P ₂ mediate the nucleotide exchange activity of VAV2 towards Rac and CDC42 (Tamás et al. 2003). In Oda et al. (2005), PI3K instead of PIP ₃ /PtdIns(3,4)P ₂ is depicted to influence the Rac-GEF activity of VAV2.
117	$\text{erbb11} \cdot \text{pi34p2} \rightarrow \text{vav2}$	1	
118	$\text{sos1r} \cdot \text{eps8r} \cdot \text{pi3kr} \cdot \text{pip3} \rightarrow \text{sos1_eps8_e3b1}$	1	SOS1, EPS8, and E3B1/ABI1 form a complex, which is necessary for the Rac-GEF activity of SOS1. Furthermore, binding of the p85 subunit of PI3K to the complex is required for a basal Rac-GEF activity, which is increased by PIP ₃ (Innocenti et al. 2003). Note that this reaction differs from the description in the map of Oda et al. (2005): There, only Ras-activated PI3K influences the GEF-activity, whereas we assume the inactivated form to mediate this effect. Additionally, the influence of PIP ₃ is not considered in the map of Oda et al. (2005).
119	$\text{vav2} \rightarrow \text{rac_cdc42}$	1	Both VAV2 and SOS1 (the latter complexed with EPS8 and E3B1) act as GEF for Rac and CDC42 (Innocenti et al. 2003; Tamás et al. 2003). As we did not find any indication in the literature for VAV2 and SOS1 activating Rac/CDC42 cooperatively, we assume that both proteins can catalyze the GDP/GTP exchange independently of each other.
120	$\text{sos1_eps8_e3b1} \rightarrow \text{rac_cdc42}$	1	

Activation of STATs

No.	Interaction	Time	Documentation
121	erbb11 · csrc → stat1	1	STAT1, 3, and 5 can be activated through ERBB1
122	erbb11 · csrc → stat3	1	homodimers, whereas ERBB1 heterodimers do not seem to contribute to STAT activation. Neuregulins, which cannot
123	erbb11 · csrc → stat5	1	activate ERBB1 dimers, induce activation of STAT5
124	erbb24 · csrc → stat5	1	through ERBB24. After ligand binding, c-Src is recruited to the activated receptor and phosphorylates receptor bound STATs on the consensus C-terminal Y-residue (Olayioye et al. 1999).

PI3K signaling

No.	Interaction	Time	Documentation
125	erbb13 · pi3kr → pi3k	1	In Oda et al. (2005), activation of PI3K through all
126	erbb34 · pi3kr → pi3k	1	ERBB3 and ERBB4 dimers is considered. Indeed,
127	erbb23 · pi3kr → pi3k	1	there are binding sites for the p85 subunit of PI3K on both ERBB3 and ERBB4. However, there are six sites on ERBB3 and only one site on ERBB4, suggesting that ERBB3 is the main activator of PI3K (Olayioye et al. 2000). Furthermore, there are naturally occurring ERBB4 isoforms that do not contain the binding site for PI3K (Elenius et al. 1999). Thus, we decided to include PI3K interaction only with ERBB3 dimers.
128	ras · pi3kr → pi3k	1	GTP-bound Ras activates PI3K (Downward 1998b).
129	pi3kr · gab1 → pi3k	1	Phosphorylated GAB1 recruits and activates PI3K (Montagner et al. 2005). GAB1-bound SHP2 dephosphorylates the PI3K binding site of GAB1. However, we decided not to include the negative influence of SHP2 on PI3K in the model, because it seems as if SHP2 indeed downregulates PI3K, but does not completely inhibit PI3K activation through GAB1 (Montagner et al. 2005; Zhang et al. 2002).
130	pten → ptend	2	We assume that downregulation of PI3K signaling is
131	ship2 → ship2d	2	time delayed. Therefore, we introduced dummy species, which are activated at time 2.

Continued on next page

C.2 List of Interactions of the Logical ERBB Model (M1)

No.	Interaction	Time	Documentation
132	ship2d · !ptend · pi3k → pi34p2	1	PI3K phosphorylates PtdIns(4,5)P ₂ at the D3 position and thus generates PtdIns(3,4,5)P ₃ (PIP ₃) (Vanhaesebroeck et al. 1997). Since PtdIns(4,5)P ₂ is one of the major phosphorylated forms of PtdIns (Tolias and Cantley 1999), we assumed that it is always present in the cell and did not consider its regulation. PTEN dephosphorylates PIP ₃ at the D3 position, SHIP2 catalyzes the dephosphorylation at D5 (Scheid and Woodgett 2003).
133	!ship2d · !ptend · pi3k → pip3	1	
134	mtorr → mtor_ric	1	As the molecular mechanisms of the regulation of the mTOR-ricor complex is unknown (Sarbasov et al. 2005b), we assumed that it is only activated by its reservoir (comparable to an external input of the model).
135	!pp2a · pip3 · pdk1 · mtor_ric → akt	1	PIP ₃ or PtdIns(3,4)P ₂ recruit AKT and PDK1 to the plasma membrane. At the membrane, the HM region of AKT is phosphorylated at S473, probably by the mTOR-ricor complex (Sarbasov et al. 2005a). The phosphorylated HM region of AKT stabilizes PDK1 so that PDK1 can phosphorylate T308 of AKT (Scheid and Woodgett 2003). PP2A dephosphorylates AKT (Andjelkovic et al. 1996).
136	!pp2a · pi34p2 · pdk1 · mtor_ric → akt	1	
137	!akt → tsc1_tsc2	1	AKT phosphorylates TSC2 and thus inhibits the Rheb-GAP activity of the TSC1/TSC2 complex. GTP-bound Rheb activates the mTOR-raptor complex (Hay and Sonenberg 2004).
138	!tsc1_tsc2 → rheb	1	
139	rheb · mtorr → mtor_rap	1	
140	erk12 → p70s6k_1	1	Phosphorylation of several S/T residues (S404, S411, S418, S424, T421) in the C-terminal autoinhibitory domain of p70S6K leads to a conformational change that enables the phosphorylation of the catalytic sites T389 and S229 (Berven and Crouch 2000). Both JNK and ERK1/2 are able to phosphorylate the autoinhibitory sites (Mukhopadhyay et al. 1992). However, the mechanism of activation of these sites is not well understood and additional kinases are probably involved in this step (Berven and Crouch 2000). We refer to p70S6K phosphorylated at the autoinhibitory sites as p70s6k_1.
141	jnk → p70s6k_1	1	

Continued on next page

No.	Interaction	Time	Documentation
142	pdk1 · p70s6k_1 · mtorrap → p70s6k_2	1	mTOR phosphorylates p70S6K on T389 (Hou et al. 2007), and T229 is phosphorylated by PDK1 (Berven and Crouch 2000; Downward 1998a).
143	!pak1 · !akt → bad	1	PAK1 phosphorylates BAD on S112 and S136, independently of PI3K (Schürmann et al. 2000). AKT phosphorylates BAD on S136. In some cell types (e.g., cerebellar granule cells), this suffices for inhibiting apoptosis. However, in other cell types (e.g., Il-3 dependent hematopoietic cells), BAD must be phosphorylated on S136 and S112 (Datta et al. 1997).
144	bad → pro_apoptotic	1	Phosphorylation of BAD avoids its proapoptotic function (Schürmann et al. 2000).

Activation of PKC

No.	Interaction	Time	Documentation
145	erbb11 → plcg	1	PLC γ is phosphorylated by ERBB1 at Y1254, Y783, Y771, and Y472 (Kim et al. 1990).
146	plcg → dag	1	PLC γ hydrolyzes PtdIns(4,5)P ₂ to generate DAG and
147	plcg → ip3	1	IP ₃ (Kim et al. 2000). We did not consider PLC β (as depicted in Oda et al. (2005)), since this is part of the G-coupled receptor signaling.
148	ip3 → ca	1	Binding of IP ₃ to its receptor at the endoplasmatic reticulum leads to Ca ²⁺ release into the cytosol (Alberts et al. 2004; Kim et al. 2000).
149	pdk1 · dag · ca → pkc	1	PKC is phosphorylated at its activation loop by PDK1. This leads to autophosphorylation and the release of PKC into the cytoplasm. A pseudosubstrate is bound to the substrate-binding cavity, which is released after binding to the second messengers Ca ²⁺ and DAG (Newton 2003). Note that the influence of calcium ions on this reaction is not part of the map of Oda et al. (2005).

MAPK cascades

No.	Interaction	Time	Documentation
150	akt → aktD	2	Interaction that is only included for modeling the time delay AKT deactivates Raf1 with.
151	ras · csrc · !aktD → raf1	1	Ras recruits Raf1 to the plasma membrane, where it is phosphorylated at various sites. c-Src phosphorylates Raf1 on Y341, PAK1 phosphorylates S338, whereas it seems as if phosphorylation of either S338 or Y341 is sufficient for Raf1 activation. However, both kinases lead to different activation levels of Raf1 (the highest to be achieved in combination) that might stimulate different biological outcomes (King et al. 2001).
152	ras · pak1 · !aktD → raf1	1	
153	rac_cdc42 → mlk3	1	MLK3 binds GTP-bound CDC42/Rac and is thus activated (Vacratsis et al. 2002).
154	rac_cdc42 → mekk4	1	MEKK4 contains a CDC42/Rac interactive binding motif and is activated after binding to CDC42/Rac. This binding is independent of the nucleotide bound to CDC42/Rac, thus MEKK4 can be activated by the GDP- and GTP-bound protein (Schlesinger et al. 1998). Perhaps a different activation for CDC42/Rac is necessary for binding MEKK4. We only included binding to the GTP-bound state in the model.
155	rac_cdc42 → mekk1	1	MEKK1 binds CDC43/Rac and is thus activated (Schlesinger et al. 1998).
156	mekk1 → mek12	1	MEKK1 and Raf1 both phosphorylate MEK1 and MEK2 (Chen et al. 2001; Schlesinger et al. 1998).
157	raf1 → mek12	1	
158	mek12 → erk12	1	MEK1 and MEK2 phosphorylate ERK1/2 (Robinson and Cobb 1997).
159	mekk1 → mkk7	1	MEKK1 phosphorylates and thus activates MKK7 (Lu et al. 1997).
160	mekk4 → mkk4	1	MKK4 is phosphorylated by MLK3, MEKK1 (Tibbles et al. 1996), and MEKK4 (Gerwins et al. 1997).
161	mlk3 → mkk4	1	
162	mekk1 → mkk4	1	
163	mkk7 · mkk4 → jnk	1	MKK4 and MKK7 cooperate to activate JNK. MKK4 phosphorylates Y185, MKK7 phosphorylates T183 (Kishimoto et al. 2003).

Continued on next page

No.	Interaction	Time	Documentation
164	mlk3 → mkk3	1	MLK3 phosphorylates and thus activates MKK3 and MKK6 (Tibbles et al. 1996).
165	mlk3 → mkk6	1	
166	mkk3 → p38	1	MKK3, MKK4, and MKK6 phosphorylate p38 on threonine and tyrosine residues (Raigneaud et al. 1996).
167	mkk4 → p38	1	
168	mkk6 → p38	1	

Activation downstream of MAPK (mainly activation of transcription factors)

No.	Interaction	Time	Documentation
169	!mkp · erk12 → nucerk12	1	MKP dephosphorylates ERK1/2 and thus inhibits phosphorylation of transcription factors like Elk1 in the nucleus (Sun et al. 1993).
170	!pp2b · nucerk12 → elk1	1	ERK1/2 phosphorylates Elk1 at S383 and S389 (Cavigelli et al. 1995). PP2B dephosphorylates Elk1 (Tian and Karin 1999). In Tian and Karin (1999), activation of Elk1 through other MAPKs is also mentioned. However, for phosphorylating transcription factors translocation to the nucleus is necessary, which is stimulated in the case of JNK by UV-irradiation (Cavigelli et al. 1995).
171	p38 → mk2	1	MK2 binds to and is phosphorylated by p38 (Gaestel 2006).
172	mk2 → hsp27	1	MK2 phosphorylates HSP27 on S15, S78, and S82 (Stokoe et al. 1992).
173	pdk1 · erk12 → p90rsk	1	ERK1/2 and PDK1 activate p90RSK by phosphorylation. ERK1/2 activates the C-terminal domain, PDK1 the N-terminal domain (Ser227), whereas the first is necessary for the latter (Frödin et al. 2000).
174	p90rsk → creb	1	p90RSK phosphorylates CREB on S133 and thus activates it (De Cesare et al. 1998).
175	mk2 → creb	1	MK2 phosphorylates and thus activates CREB at S133 (Tan et al. 1996).

Continued on next page

C.2 List of Interactions of the Logical ERBB Model (M1)

No.	Interaction	Time	Documentation
176	!p90rsk · !akt → gsk3	1	p90RSK phosphorylates GSK3 on S9 and thus deactivates it in response to EGF. AKT also phosphorylates S9—however, we are not sure whether this occurs only in response to insulin. Other kinases, like p70S6K and PKC, are also known to deactivate GSK3; however, not in all cell types and not in response to EGF, so their influence has to be further studied before including it in the model (Grimes and Jope 2001).
177	nucerk12 · !gsk3 → cmyc	1	ERK1/2 phosphorylates c-Myc at S62 and thus stabilizes it. The phosphorylation of S62 is necessary for the phosphorylation of T58 by GSK3 beta, which leads to ubiquitin dependent degradation of c-Myc (Sears et al. 2000). GSK3 beta is not included in Oda et al. (2005).
178	jnk → cjun	1	JNK phosphorylates c-Jun. Unphosphorylated c-Jun is ubiquitinated and degraded. Phosphorylation by JNK also increases the transcriptional activity of c-Jun (Karin et al. 1997). Regulation of transcription of c-Jun was not considered here; this could be included in a model regarding multi-level activation.
179	!pp2a · jnk → cfos	1	JNK phosphorylates c-Fos and thus prevents it from degradation (Coronella-Wood et al. 2004). Note that this reaction is independent of ERK (Coronella-Wood et al. 2004). ERK1/2 and p90RSK have been shown to coordinately phosphorylate c-Fos (Murphy et al. 2002). The influence of p90RSK is not depicted in Oda et al. (2005). PP2A dephosphorylates c-Fos, whereas PP2B does not (Coronella-Wood et al. 2004), in contradiction to the depiction in Oda et al. (2005). Regulation of transcription of c-Fos was not considered here; this could be included in a model regarding multi-level activation.
180	!pp2a · p90rsk · erk12 → cfos	1	
181	cjun · cfos → ap1	1	c-Jun and c-Fos heterodimerize and form the transcription factor AP1 (Karin et al. 1997).

Endocytosis

No.	Interaction	Time	Documentation
182	erbb11 → ccbl	1	c-Cbl binds ERBB1 at pY1045, leading to degradation of the receptor in the lysosome (Citri and Yarden 2006).
183	erbb11 · eps8r → rntr	1	RN-tre binds to the adapter protein EPS8 and is phosphorylated in response to EGF stimulation (Lanzetti et al. 2000).
184	ras → rin1	1	Rab5-GEF activity of RIN1 is potentiated by activated Ras (Tall et al. 2001).
185	rntr · rin1 → rab5a	1	RIN1 mediates GDP/GTP exchange for Rab5A, thus activating it (Tall et al. 2001). The GTPase activating protein RN-tre acts on Rab5A and inhibits internalization of EGFR (Lanzetti et al. 2000).
186	rab5a · ccbl → endocyt_degrad	1	c-Cbl and Rab5A are both involved in the endocytic trafficking of ERBB receptors. c-Cbl is necessary for degradation of the receptors, while Rab5A controls the formation and fusion of endocytic vesicles (Citri and Yarden 2006).

Actin reorganization

No.	Interaction	Time	Documentation
187	rac_cdc42 · grb2 → pak1	1	PAK1 is recruited to the plasma membrane via GRB2 (Puto et al. 2003) or NCK (Li et al. 2001), where it is activated through GTP-bound Rac/CDC42 (Edwards et al. 1999). Note that the influence of NCK is not included in Oda et al. (2005).
188	rac_cdc42 · nck → pak1	1	
189	pak1 → limk1	1	PAK1 (activated through Rac/CDC42) phosphorylates LIMK1 at T508 (Edwards et al. 1999).
190	limk1 → actin_reorg	1	LIMK1 phosphorylates cofilin, thereby leading to accumulation of actin filaments and aggregates (Edwards et al. 1999)

Outputs

No.	Interaction
191	pkc →
192	p70s6k_2 →
193	hsp27 →
194	elk1 →
195	creb →
196	cmyc →
197	pro_apoptotic →
198	ap1 →
199	actin_reorg →
200	stat1 →
201	stat3 →
202	stat5 →

C.3 Interactions that are only included in ERBB interaction graph model

In addition to the interactions given in Appendix C.2, the interaction graph model contains the following interactions.

Interaction	Documentation
pip3 → gab1	PIP ₃ recruits GAB1 molecules to the EGFR and thus enhances the activity of GAB1 (Rodrigues et al, 2000). As we do not consider multilevel activation, we decided to exclude this positive feedback loop from the logical model.
!endocyt_degrad → erbb11	ERBB1 homodimers activated through EGF are endocytosed and subsequently degraded—in contrast to TGF α -bound receptors (Lenferink et al. 1998) and the other ERBB dimers. As the detailed endocytosis mechanism is not part of the model, we did not distinguish between EGF-bound and TGF α -bound receptors. Internalized receptors are still capable of activating signaling pathways (Citri and Yarden 2006), so we decided to exclude this reaction in the logical model.

C.4 Description of Model M2

Additional dummy species in model M2

In addition to the species given in C.1, the following dummy species are introduced to enable the description with logical operators with incomplete truth tables (ITT gates): pak1_csrc_dummy, jnk_erk_p90rsk_dummy, p90rsk_erk12_dummy, dag_ca_dummy.

Incomplete truth tables in model M2

All gates from model M1 (C.2) that point into nodes given in the first column (Nodes) are replaced by the interactions given in the second column. ITT gates are denoted with *, AND gates with .

Node	Interactions
endocyt_degrad	ccb1 * rab5a → endocyt_degrad
rntre	erbb11 * eps8r → rntre
rac_cdc42	sos1_eps8_e3b1 * vav2 → rac_cdc42
raf1	pak1 * csrc → pak1_csrc_dummy !akt d · pak1_csrc_dummy · ras → raf1
p70s6k_1	jnk * erk → p70s6k_1
creb	p90rsk * mk2 → creb
cfos	jnk * erk * p90rsk → jnk_erk_p90rsk_dummy jnk_erk_p90rsk_dummy · !pp2a → cfos
sos1	p90rsk * erk12 → p90rsk_erk12_dummy !p90rskerkerk12d · grb2 · sos1r → sos1
mkk4	mekk1 * mekk4 * mlk3 → mkk4
p38	mkk4 * mkk3 * mkk6 → p38
gsk3	!akt * !p90rsk → gsk3
bad	!akt * !pak1 → bad
pkc	dag * ca → dag_ca_dummy dag_ca_dummy · pdk1 → pkc
mek12	raf1 * mekk1 → mek12

D | Sensitivities of Binarization of HepG2 Data to Chosen Parameters

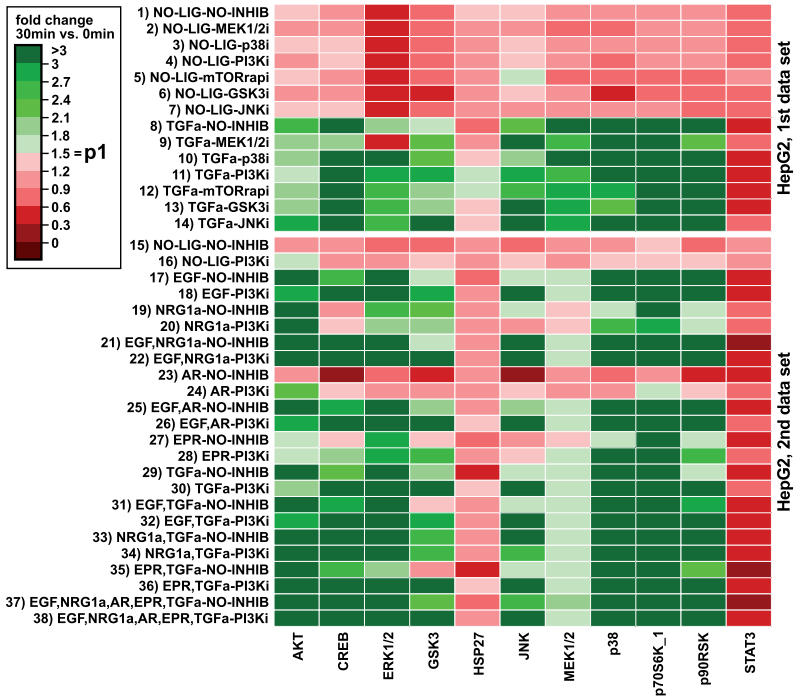


Figure D.1: Sensitivity to Parameter for Relative Significance. Ratio between 30 and 0 minutes time point lies above (green) or below (red) chosen threshold for relative significance ($p_1 = 1.5$). The darker a field is colored, the larger the distance to the chosen threshold is, that is, the binarization is less sensitive to the parameter.

D Sensitivities of Binarization of HepG2 Data to Chosen Parameters

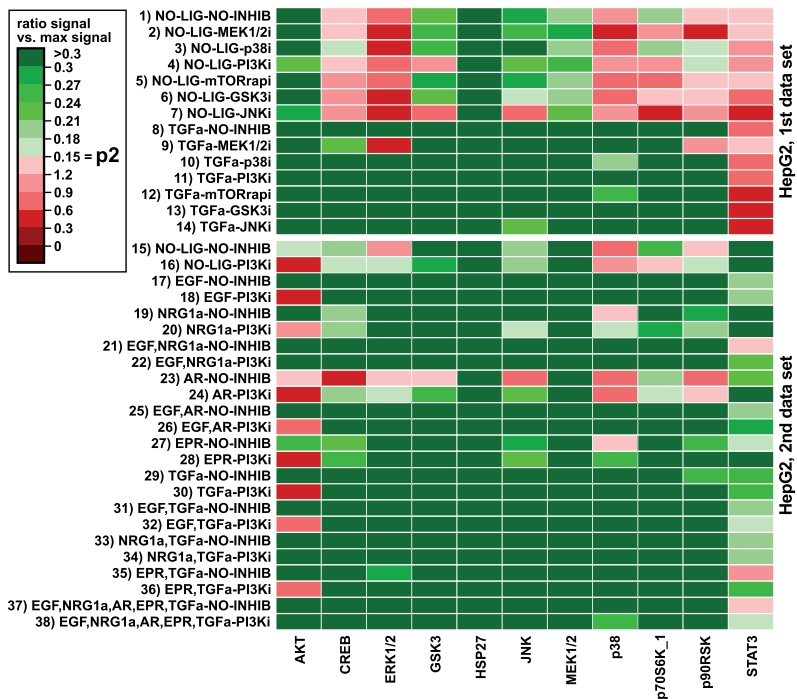


Figure D.2: Sensitivity to Parameter for Absolute Significance. Ratio between signal and maximum value for this signal over all conditions lies above (green) or below (red) chosen threshold for absolute significance ($p_2 = 0.15$). The darker a field is colored, the larger the distance to the chosen threshold is, that is, the binarization is less sensitive to the parameter.

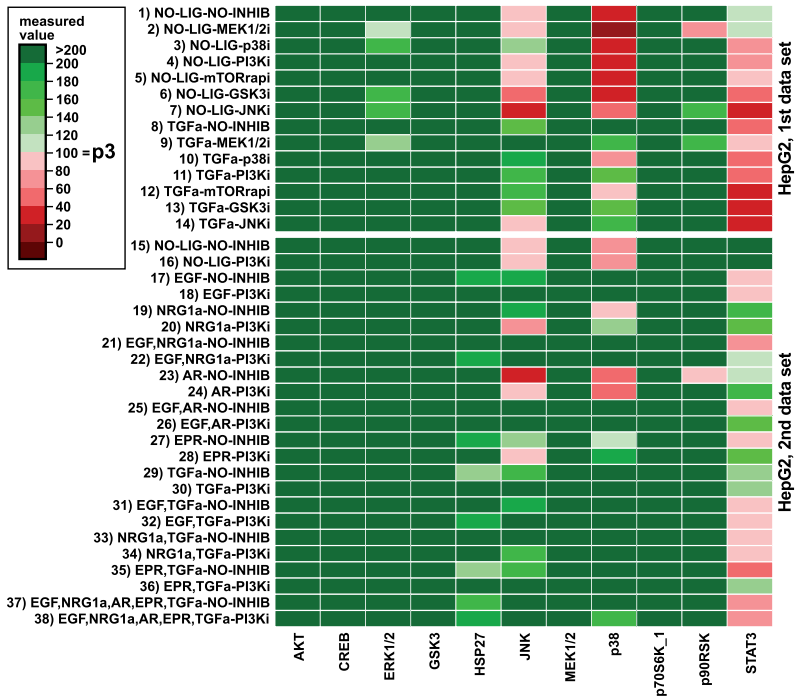


Figure D.3: Sensitivity to Parameter for Experimental Noise. Measured signal lies above (green) or below (red) the chosen threshold for experimental noise ($p_3 = 100$). The darker a field is colored, the larger the distance to the chosen threshold is, that is, the binarization is less sensitive to the parameter.

All three figures adapted from Samaga et al. (2009).

E | Documentation of the HGF Interaction Graph Master Model

E.1 List of Species

Here, we describe the species of the HGF interaction graph model presented in Chapter 5. The species are considered in their active form.

No.	Model Name	Full Name	Comment
1	akt	RAC-alpha serine/threonine-protein kinase	also known as protein kinase B (PKB)
2	braf	Serine/threonine-protein kinase B-raf	
3	crk_crkl	Adapter molecule crk/ Crk-like protein	represents Crk or CRKL
4	csrc	Neuronal proto-oncogene tyrosine-protein kinase Src	
5	c3g	Rap guanine nucleotide exchange factor 1	
6	dock180	Dedicator of cytokinesis	also known as DOCK1
7	erk	Extracellular signal-regulated kinase 1 (ERK1) or 2 (ERK2)	since both ERK1 and ERK2 catalyze the same reactions, we do not distinguish between them; also known as MAPK1/3
8	gab1	GRB2-associated-binding protein 1	
9	grb2	Growth factor receptor-bound protein 2	
10	hgf	Hepatocyte growth factor	also known as Scatter factor
11	mek		MEK1/2 phosphorylated at S217/S221
12	mek_s298	MAPK/ERK kinase 1 or 2	MEK1/2 phosphorylated at S298
13	mek_t292		MEK1/2 phosphorylated at T292

Continued on next page

E Documentation of the HGF Interaction Graph Master Model

No.	Model Name	Full Name	Comment
14	Met	Hepatocyte growth factor receptor	
15	pak	p21-activated kinase	
16	pdk1	3-phosphoinositide-dependent protein kinase 1	
17	pip3	Phosphatidylinositol (3,4,5)-trisphosphate	
18	pi3k	Phosphatidylinositol-4,5-bisphosphate 3-kinase	
19	rac	Ras-related C3 botulinum toxin substrate 1	
20	raf1	RAF proto-oncogene serine/threonine-protein kinase	also known as cRaf
21	rap1	Ras-related protein Rap-1	
22	ras	GTPase Ras	
23	rasgap	Ras GTPase-activating protein 1	
24	rkip	Raf-1 kinase inhibitor protein	
25	rsk_s		p90RSK phosphorylated on a single serine residue
26	rsk_d	90 kDa ribosomal protein S6 kinase	p90RSK phosphorylated on two serine residues, considered as active p90RSK
27	shc	SHC-transforming protein 1	
28	shp2	Tyrosine-protein phosphatase non-receptor type 11	also known as PTP-1D
29	sos1	Son of sevenless homolog 1	
30	sos1_eps8_e3b1		Complex of SOS1, Eps8 (EGF receptor kinase substrate 8), and E3B1/ABI1 (Abl interactor 1)

E.2 List of Interactions of the HGF Core Model

In the following table, we describe the list of reactions of the HGF interaction graph (core model) presented in Chapter 5. Table adapted from D'Alessandro, Samaga, Maiwald et al. (2015).

Activation of Met and adapter proteins

No.	Interaction	Documentation
1	hgf → met	Binding of HGF to Met induces autophosphorylation of the receptor at tyrosines 1334 and 1335. Subsequently, Met phosphorylation occurs at tyrosines 1349 and 1356 (Furge et al. 2000).
2	met → shc	The PTB of SHC associates with the Met receptor via Y1349 and Y1356 (Furge et al. 2000).
3	met → grb2	GRB2 is recruited to Met directly (Y1356) and indirectly via SHC (Furge et al. 2000).
4	shc → grb2	
5	met → csrc	The SH2 domain of c-Src can bind both Y1349 and Y1356 of the Met receptor (Ponzetto et al. 1994).
6	met → gab1	GAB1 binds Met both directly (Y1349) and indirectly via GRB2 binding at Y1356 (Furge et al. 2000).
7	grb2 → gab1	
8	csrc → gab1	c-Src binds to and phosphorylates GAB1 in response to HGF. Inhibiting c-Src partially inhibited HGF-induced GAB1 phosphorylation by 40 to 60% (Chan et al. 2003). Additional phosphorylation through Met is required for GAB1 activation. One hypothesis is that c-Src and Met predominantly phosphorylate GAB1 at different tyrosine residues (Chan et al. 2003).
9	gab1 → crk_crkl	Crk/CRKL binds to GAB1 after HGF stimulation (Garcia-Guzman et al. 1999; Sakkab et al. 2000).

Activation of G-proteins Ras and Rac

No.	Interaction	Documentation
10	grb2 → sos1	SOS1 is bound to GRB2. Recruitment of the GRB2-SOS1 complex to the activated Met receptor (Y1349/Y1356) (Ponzetto et al. 1994) leads to activation of SOS1.
11	sos1 → ras	Interaction of Ras with SOS1 increases the rate of GDP/GTP exchange of Ras (Li et al. 1993).

Continued on next page

No.	Interaction	Documentation
12	crk_crkl → dock180	Coexpression of Crk with DOCK180 results in activation of the JNK pathway, indicating that DOCK180 connects Crk to the Rac/JNK pathway. However, the possibility that Crk acts through SOS1 cannot be ruled out (Dolfi et al. 1998).
13	dock180 → rac	DOCK180 is a Rac-specific GEF (Rossman et al. 2005).

Activation of the PI3K pathway

No.	Interaction	Documentation
14	met → pi3k	PI3K can bind to Met both directly (Ponzetto et al. 1993) and indirectly via
15	gab1 → pi3k	Gab1 (Furge et al. 2000).
16	pi3k → pip3	PI3K phosphorylates PtdIns(4,5)P ₂ at the D3 position and thus generates PIP ₃ (Vanhaesebroeck et al. 1997). Since PtdIns(4,5)P ₂ is one of the major phosphorylated forms of PtdIns (Tolias and Cantley 1999), we assume that it is always present in the cell and do not consider its regulation.
17	pip3 → akt	PIP3 recruits AKT and PDK1 to the plasma membrane. Subsequently, AKT
18	pdk1 → akt	is phosphorylated on S473 and T308 (Scheid and Woodgett 2003).

Activation of the MAPK pathway

No.	Interaction	Documentation
19	ras → raf1	Ras recruits Raf1 to the plasma membrane, where it is phosphorylated at various sites (King et al. 2001).
20	raf1 → mek	MEK1/2 can be activated by Raf1 (Chen et al. 2001).
21	mek → erk	MEK1 and MEK2 phosphorylate ERK1/2 (Robinson and Cobb 1997).
22	erk → rsk_s	ERK1/2 and PDK1 phosphorylate p90RSK. ERK1/2 activates the
23	pdk1 → rsk_d	C-terminal domain, PDK1 the N-terminal domain, whereas the first is
24	rsk_s → rsk_d	necessary for the latter (Frödin et al. 2000).
25	grb2 → pak	PAK is recruited to the plasma membrane via GRB2 (Puto et al. 2003),
26	rac → pak	where it is activated through GTP-bound Rac (Edwards et al. 1999).

E.3 List of Candidate Mechanisms

The following table describes the list of reactions defined as candidate mechanisms in the HGF interaction graph master model presented in Chapter 5. Each number corresponds to a candidate mechanism. The letters indicate the interactions for the candidate mechanisms composed by more than one interaction (edge). Edges marked with an asterisk are included in more than one candidate mechanism. Table adapted from D'Alessandro, Samaga, Maiwald et al. (2015).

No.	Interaction	Documentation
1a	gab1 → rasgap	RasGAP binds phosphorylated GAB1 (Montagner et al. 2005).
1b*	rasgap ⊣ ras	The GTPase activating protein RasGAP induces the hydrolysis of Ras-bound GTP to GDP (Cox and Der 2003).
2a	gab1 → shp2	SHP2 binds to phosphorylated GAB1 in response to HGF stimulation (Rosário and Birchmeier 2003).
2b	shp2 ⊣ rasgap	SHP2 dephosphorylates the RasGAP binding site on GAB1 (Dance et al. 2008).
2c*	rasgap ⊣ ras	The GTPase activating protein RasGAP induces the hydrolysis of Ras-bound GTP to GDP (Cox and Der 2003).
3a	crk_crkl → c3g	Crk and CRKL affect the Rap1 activator C3G. C3G-dependent Rap1 activation and CRKL recruitment to GAB1 have been shown in response to HGF (Sakkab et al. 2000).
3b	c3g → rap1	
3c	rap1 → braf	Rap1 activates BRAf in different cell types and for different stimuli (Stork 2003). In human melanoma cell lines, Rap1-dependent activation of ERK in response to HGF was shown (Gao et al. 2006).
3d	braf → mek	MEK1/2 can be activated by Raf1 or BRAf (Chen et al. 2001).
4	erk ⊣ sos1	Serine/threonine phosphorylation of SOS1 by ERK1/2 and/or p90RSK causes dissociation of GRB2-SOS1 (Douville and Downward 1997).
5	rsk_d ⊣ sos1	
6a	erk ⊣ rkip	ERK-mediated phosphorylation of RKIP triggers RKIP dissociation from Raf1 and thus counteracts the inhibitory function of RKIP (Shin et al. 2009).
6b	rkip ⊣ raf1	RKIP association with Raf1 disrupts the interaction between MEK and Raf1, which is required for MEK phosphorylation (Yeung et al. 1999).

Continued on next page

No.	Interaction	Documentation
7	erk ⊣ raf1	ERK phosphorylates Raf1 at five inhibitory serine sites (Dougherty et al. 2005).
8	pdk1 → mek	PDK1 phosphorylates MEK1 on S222 and MEK2 on S226 (Sato et al. 2004).
9a	pip3 → sos1_eps8_e3b1	SOS1, EPS8, and E3B1/ABI1 form a complex which is necessary for the Rac-GEF activity of SOS1. The basal Rac-GEF activity is increased by PIP ₃ (Innocenti et al. 2002, 2003).
9b	sos1_eps8_e3b1 → rac	
10	pak → raf1	PAK phosphorylates Raf1 on S338 (King et al. 2001).
11a	pak → mek_s298	PAK1 phosphorylates MEK1, increasing the ability of Raf1 to activate MEK1 (Eblen et al. 2004).
11b*	mek_s298 → mek	
12a	erk → mek_t292	ERK phosphorylates MEK1 on T292, preventing further phosphorylation of MEK1 by PAK on S298 (Eblen et al. 2004).
12b	mek_t292 ⊣ mek_s298	
12c*	mek_s298 → mek	
13	pak → akt	PAK promotes AKT recruitment to the plasma membrane and functions as a scaffold to facilitate complex formation between PDK1 and AKT (Higuchi et al. 2008).
14	erk → pi3k	ERK phosphorylates GAB1 in response to HGF and potentiates GAB1-PI3K association and subsequent PI3K activation (Yu et al. 2001).
15	ras → pi3k	GTP-bound Ras activates PI3K (Downward 1998b).
16	pip3 → gab1	Plasma membrane localization of GAB1 requires PIP ₃ (Maroun et al. 1999; Rosário and Birchmeier 2003).
17	akt ⊣ raf1	AKT negatively regulates Raf1 by phosphorylation of S259 (Mabuchi et al. 2002; Zimmermann and Moelling 1999).

F | Documentation of HGF ODE Models

F.1 Reactions in the ODE Models

Here, we describe the list of reactions of the HGF ODE models discussed in Chapter 5. Section F.1.1 describes the reactions of the core model. Section F.1.2 describes the reactions of the candidate mechanisms. In the second column of the respective tables, each reaction is shown in a schematic representation; in the third column, the respective kinetic rate law is shown. Parameters Met_inh , PDK_inh and MEK_inh represent binary values dependent on the respective experimental condition. Tables adapted from D'Alessandro, Samaga, Maiwald et al. (2015).

F.1.1 Reactions in the Core Model

No.	Reactions	Kinetic rate law
1	Met receptor \rightarrow Phosphorylated Met receptor	$Met_activation * Met * HGF / (1 + Met_act_inh * Met_inh)$
2	Phosphorylated Met receptor \rightarrow Met receptor	$pMet_dephosphorylation * pMet$
3	$\emptyset \rightarrow$ Met receptor	$Met_prod_deg_ss$
4	Met receptor $\rightarrow \emptyset$	$Met * Met_prod_deg_ss$
5	Phosphorylated Met receptor $\rightarrow \emptyset$	$pMet_degradation * pMet$
6	Gab1 + Phosphorylated Met receptor \rightarrow Gab1_Met receptor complex	$Gab1_pMet_binding * Gab1 * pMet$
7	Gab1_Met receptor complex \rightarrow Gab1 + Phosphorylated Met receptor	$Gab1_pMet_diss * Gab1_pMet$
8	Rac \rightarrow Active Rac	$Rac_activation * Rac * Gab1_pMet$
9	Active Rac \rightarrow Rac	$Rac_deactivation * Rac_active$
10	PAK \rightarrow Phosphorylated PAK	$PAK_phosphorylation * PAK1 * Rac_active$
11	Phosphorylated Pak \rightarrow Pak	$PAK_dephosphorylation * pPAK1$

Continued on next page

No.	Reactions	Kinetic rate law
12	PI3K → Active PI3K	PI3K_activation_by_Gab1 * PI3K * Gab1_pMet
13	Active PI3K → PI3K	PI3K_inactivation * PI3K_active
14	Akt → Phosphorylated Akt	Akt_activation_supported_by_PDK1 * Akt * PI3K_active / (1 + PDK_act_inh*PDK_inh)
15	Phosphorylated Akt → Akt	pAkt_deactivation * pAkt
16	SOS → SOS_Met receptor complex	SOS_recruitment_by_pMet * SOS * pMet
17	SOS_Met receptor complex → SOS	mSOS_release_from_membrane * mSOS_pMet
18	Ras → Active Ras	Ras_activation_by_mSOS * Ras * mSOS_pMet
19	Active Ras → Ras	Ras_deactivation * Ras_active
20	Raf → Phosphorylated Raf	Raf_activation * Raf * Ras_active
21	Phosphorylated Raf → Raf	Raf_inactivation * pRaf
22	MEK → Phosphorylated MEK	MEK_phosphorylation_by_pRaf * MEK * pRaf
23	Phosphorylated MEK → MEK	MEK_dephosphorylation * pMEK
24	ERK → Phosphorylated ERK	ERK_phosphorylation_by_pMEK * ERK * pMEK / (1 + MEK_act_inh*MEK_inh)
25	Phosphorylated ERK → ERK	ERK_dephosphorylation * pERK
26	RSK → Single phosphorylated RSK	RSK_phosphorylation_by_pERK * pERK * RSK
27	Single phosphorylated RSK → Double phosphorylated RSK	RSK_phosphorylation_by_PDK1 * single_pRSK / (1 + PDK_act_inh * PDK_inh)
28	Single phosphorylated RSK → RSK	RSK_dephosphorylation_single * single_pRSK
29	Double phosphorylated RSK → RSK	RSK_dephosphorylation_double * double_pRSK

F.1.2 Candidate mechanisms

No.	Reactions	Kinetic rate law
30	Gab1 + Phosphorylated Met receptor → Gab1_Met receptor complex	Gab1_pMet_binding_by_PI3K_active * Gab1 * pMet * PI3K_active
31	Rac → Active Rac	Rac_activation_by_PI3K_active * Rac * PI3K_active * Gab1_pMet
32	SOS_Met receptor complex → SOS	mSOS_release_by_pERK * mSOS_pMet * pERK
33	SOS_Met receptor complex → SOS	mSOS_release_by_pRSK * mSOS_pMet * double_pRSK
34	Ras → Active Ras	Ras_activation_by_Gab1 * Ras * Gab1_pMet * mSOS_pMet
35	Raf → Phosphorylated Raf	Raf_activation_by_PAK * Raf * pPAK1 * Ras_active
36	Raf → Phosphorylated Raf	Raf_activation_by_pERK * Raf * pERK * Ras_active
37	Phosphorylated Raf → Raf	pRaf_dephosphorylation_by_Akt * pRaf * pAkt
38	MEK → Phosphorylated MEK	MEK_phosphorylation_by_PDK1 * MEK * pRaf / (1 + PDK_act_inh * PDK_inh)
39	MEK → Phosphorylated MEK	MEK_phosphorylation_by_Gab1 * MEK * Gab1_pMet
40	MEK → Phosphorylated MEK	MEK_phosphorylation_by_pPAK * MEK * pPAK1 * pRaf
41	PI3K → Active PI3K	PI3K_activation_by_Ras_active * PI3K * Ras_active
42	PI3K → Active PI3K	PI3K_activation_by_pERK * PI3K * pERK * Gab1_pMet

F.2 Parameter Names and Values of the Final Model Structure

Here, we describe the list of parameter names and values of the final ODE model structure presented in Chapter 5, model 4_8_12. The model was calibrated on 2200 data points and 25 experimental conditions. The second column shows the \log_{10} value of each kinetic parameter involved in the model reactions shown in Sections F.1.1 and F.1.2. The x indicates that the parameter is not present in model 4_8_12. Table adapted from D'Alessandro, Samaga, Maiwald et al. (2015).

Parameter Name	Estimated Parameter Value (\log_{10})
Akt_activation_supported_by_PDK1	-0.27659584
ERK_dephosphorylation	-0.384157706
ERK_down	1.194738824
ERK_phosphorylation_by_pMEK	2.982496902
Gab1_pMet_binding	-4.601296578
Gab1_pMet_binding_by_PI3K_active	-0.409307235
Gab1_pMet_diss	-1.488874084
MEK_dephosphorylation	2.79250161
MEK_phosphorylation_by_PDK1	2.999999786
MEK_phosphorylation_by_pRaf	-1.999156196
MEK_act_inh	1.666481986
Met_act_inh	2.54026603
Met_activation	-1.934829524
Met_prod_deg_ss	-2.89724024
PAK_dephosphorylation	-4.944639884
PAK_phosphorylation	1.778987643
PI3K_activation_by_Gab1	-4.945663945
PI3K_activation_by_Ras_active	1.887411595
PI3K_activation_by_pERK	0.445509919
PI3K_inactivation	0.159466716
PDK_act_inh	0.196825209
RSK_dephosphorylation_double	0.787276371

Continued on next page

F.2 Parameter Names and Values of the Final Model Structure

Parameter Name	Estimated Parameter Value (log₁₀)
RSK_dephosphorylation_single	-0.7166527
RSK_phosphorylation_by_PDK1	-0.568280334
RSK_phosphorylation_by_pERK	-0.559320897
Rac_activation	-4.963891394
Rac_deactivation	-4.999447793
Raf_activation	-0.834642816
Raf_activation_by_PAK	0.512776283
Raf_activation_by_pERK	0.223987013
Raf_inactivation	0.687062866
Ras_activation_by_mSOS	2.552300571
Ras_deactivation	-1.159750752
SOS_recruitment_by_pMet	-3.42126982
mSOS_release_by_pRSK	3
mSOS_release_from_membrane	-4.999999297
pAkt_deactivation	0.348215544
pMet_degradation	-0.86658565
pMet_dephosphorylation	-0.302018341
pRaf_dephosphorylation_by_Akt	3
MEK_phosphorylation_by_Gab1	x
MEK_phosphorylation_by_pPAK	x
Rac_activation_by_PI3K_active	x
Ras_activation_by_Gab1	x
mSOS_release_by_pERK	x

G | Proofs from Chapter 6

Proof of Theorem 6.18. In the following, we show that the graph-theoretic criterion for initial response (Theorem 6.18) is valid for the non-linearized system, and not only for its linearization as shown in Oyeleye and Kramer (1988). The proof is due to D. Flockerzi and given here only for the sake of completeness.

Consider the smooth system

$$\dot{x} = f(x) + b, \quad x(0) = \xi \text{ with } f(\xi) = 0.$$

We consider the deviation $y(t) := x(t) - \xi$, where $y(0) = 0$. Of course, the deviation $y(t)$ is also dependent on b . With J_0 we denote the Jacobian at state ξ , that is, $J_0 = f_x(\xi)$.

According to the mean value theorem, we have

$$\begin{aligned} y(\alpha t) &= y(\rho\alpha t) \Big|_{\rho=0}^{\rho=1} \\ &= \int_0^1 \dot{y}(\rho\alpha t) \alpha t d\rho \\ &= \int_0^1 [b + f(\xi + y(\rho\alpha t))] \alpha t d\rho \\ &= b\alpha t + \int_0^1 f(\xi + y(\rho\alpha t)) \alpha t d\rho \end{aligned} \tag{G.1}$$

Furthermore, we have

$$\begin{aligned} f(\xi + y(\rho\alpha t)) &= f(\xi + \sigma y(\rho\alpha t)) \Big|_{\sigma=0}^{\sigma=1} \\ &= \int_0^1 f_x(\xi + \sigma y(\rho\alpha t)) y(\rho\alpha t) d\sigma \end{aligned} \tag{G.2}$$

(G.1) and (G.2) together give

$$y(\alpha t) = b\alpha t + \int_0^1 \left[\int_0^1 f_x(\xi + \sigma y(\rho\alpha t)) d\sigma \right] y(\rho\alpha t) \alpha t d\rho \tag{G.3}$$

With $A(y(\rho\alpha t)) := \int_0^1 f_x(\xi + \sigma y(\rho\alpha t)) d\sigma$, and with $\alpha = 1$, $\rho = s_1$, we have

$$\frac{y(t)}{t} = b + \int_0^1 A(y(s_1 t)) y(s_1 t) ds_1.$$

According to (G.3), and with $\alpha = s_1$, $\rho = s_2$, this gives

$$\begin{aligned} \frac{y(t)}{t} &= b + \int_0^1 A(y(s_1 t)) \left(b s_1 t + \int_0^1 A(y(s_2 s_1 t)) y(s_2 s_1 t) s_1 t ds_2 \right) ds_1 \\ &= b + t \int_0^1 A(y(s_1 t)) s_1 ds_1 b + t \int_0^1 A(y(s_1 t)) s_1 \int_0^1 A(y(s_2 s_1 t)) y(s_2 s_1 t) ds_2 ds_1. \end{aligned}$$

The term $\int_0^1 A(y(s_1 t)) s_1 ds_1$ is of the form $M_1(t) = \frac{1}{2} J_0 + \mathcal{O}(|t|)$, where J_0 is independent of b , but the term $\mathcal{O}(|t|)$ depends on b . Replacing again $y(s_2 s_1 t)$ according to (G.3), we have

$$\frac{y(t)}{t} = b + t M_1(t) b + t^2 \left(\int_0^1 A(y(s_1 t)) s_1 \int_0^1 A(y(s_2 s_1 t)) s_2 s_1 ds_2 ds_1 b + \dots \right).$$

The term $\int_0^1 A(y(s_1 t)) s_1 \int_0^1 A(y(s_2 s_1 t)) s_2 s_1 ds_2 ds_1$ is of the form $M_2(t) = \frac{1}{3!} J_0^2 + \mathcal{O}(|t|)$.

This can be interpreted as follows: Let J_0^q denote the sign structure of J_0 . If $(J_0^q b)_i$ is positive/negative, then the initial response of x_i is positive/negative. If $(J_0^q b)_i = 0$, then we move on to the next term $(t^2 M_2(t) b)_i = (t^2 \frac{1}{3!} J_0^2 + \mathcal{O}(|t|^3))_i$ and proceed analogously.

The (i, j) -th element of $(J_0^q)^l$ is

$$(J_0^q)^l_{ij} = \sum_{k_1=1}^n \sum_{k_2=1}^n \cdots \sum_{k_{l-1}=1}^n (J_0^q)_{ik_1} (J_0^q)_{k_1 k_2} \cdots (J_0^q)_{k_{l-1} j} \quad (G.4)$$

and thus represents the sum of all directed paths of length l from node S_j to node S_i in the interaction graph $\mathcal{G}(J_0^q)$. Accordingly, the i -th element of $(J_0^q)^l b$ represents the sum of all directed paths of length $l + 1$ from the external input u to S_i in $\mathcal{G}(J_0^q, b)$.

Proof of Lemma 6.63. The sum of all principal minors of $J_{i|-e_j}^q$ of order s is $(-1)^d$ times the coefficient of λ^d in the characteristic polynomial of $J_{i|-e_j}^q$ given as $\det(J^q - \lambda E_n)$, where E_n denotes the identity matrix of size n . We introduce the matrix P as follows:

$$P = \begin{pmatrix} W^T \\ (E_n)_{\{d+1, \dots, n\}, \{1, \dots, n\}} \end{pmatrix}. \quad (G.5)$$

As $W^T = (E_d, -Z^T)$, the matrix P is an upper triangular matrix; thus, $\det(P) = 1$ and

$$\det(J_{i|e_j}^q - \lambda E_n) = \det(P J_{i|e_j}^q - \lambda P). \quad (G.6)$$

For all $x \in D$, the rows $(w^1)^T, \dots, (w^d)^T$ are orthogonal to the columns of $J(x)$; this implies that the rows $(w^1)^T, \dots, (w^d)^T$ are also orthogonal to the columns of J^q . By Assumption 6.44, $w_j^i = 0$ for all $j \in I^b$, $i \in \{1, \dots, d\}$. Thus, the top d rows of matrix $P J_{i|e_j}^q$ are zero for all $j \in I^b$. The bottom s rows are equal to the bottom s rows of $J_{i|e_j}^q$. It follows that

$$\begin{aligned} \det(P J_{i|e_j}^q - \lambda P) &= (-1)^d \lambda^d \det \begin{pmatrix} W^T & & \\ (J_{i|e_j}^q)_{\{d+1, \dots, n\}, \{1, \dots, d\}} & (J_{i|e_j}^q)_{\{d+1, \dots, n\}, \{d+1, \dots, n\}} & -\lambda E_s \end{pmatrix} \\ &= (-1)^d \lambda^d \det \begin{pmatrix} W^T & & \\ (J_{i|e_j}^q)_{d+1, \{1, \dots, n\}} & & \\ (J_{i|e_j}^q)_{\{d+2, \dots, n\}, \{1, \dots, d+1\}} & (J_{i|e_j}^q)_{\{d+2, \dots, n\}, \{d+2, \dots, n\}} & -\lambda E_{s-1} \end{pmatrix} \\ &\quad + \det \begin{pmatrix} W^T & & \\ -\lambda(e_{d+1})^T & & \\ (J_{i|e_j}^q)_{\{d+2, \dots, n\}, \{1, \dots, d+1\}} & (J_{i|e_j}^q)_{\{d+2, \dots, n\}, \{d+2, \dots, n\}} & -\lambda E_{s-1} \end{pmatrix} \\ &= (-1)^d \lambda^d \left(\det(\tilde{J}_{i|e_j}^q) \right. \\ &\quad \left. - \lambda \sum_{k=d+1}^n \det \begin{pmatrix} W^T & & \\ (J_{i|e_j}^q)_{\{d+1, \dots, n\}, \{1, \dots, d\}} & (J_{i|e_j}^q)_{\{d+1, \dots, n\}, \{d+1, \dots, n\}} & -\lambda E_s \end{pmatrix}_{\{1, \dots, n\} \setminus \{k\}, \{1, \dots, n\} \setminus \{k\}} \right) \end{aligned}$$

Thus, the coefficient of $(-1)^d \lambda^d$ of the characteristic polynomial of $J_{i|e_j}^q$ is given as the determinant of $\tilde{J}_{i|e_j}^q$.

Bibliography

Own Publications

- *D'Alessandro LA, *Samaga R, *Maiwald T, Rho SH, Bonefas S, Raue A, Iwamoto N, Kienast A, Waldow K, Meyer R, Schilling M, Timmer J, Klamt S, and Klingmüller U (2015) Disentangling the complexity of HGF signaling by combining qualitative and quantitative modeling. *PLoS Comput Biol* (4), e1004192.
- Kolczyk K, Samaga R, Conzelmann H, Mirschel S, and Conradi C (2012) The process-interaction-model: a common representation of rule-based and logical models allows studying signal transduction on different levels of detail. *BMC Bioinformatics* **13**, 251.
- *Melas IN, *Samaga R, Alexopoulos LG, and Klamt S (2013) Detecting and removing inconsistencies between experimental data and signaling network topologies using integer linear programming on interaction graphs. *PLoS Comput Biol* **9**, e1003204.
- Ryll A, Samaga R, Schaper F, Alexopoulos LG, and Klamt S (2011) Large-scale network models of IL-1 and IL-6 signalling and their hepatocellular specification. *Mol Biosyst* **7**, 3253–3270.
- Saez-Rodriguez J, Alexopoulos LG, Epperlein J, Samaga R, Lauffenburger DA, Klamt S, and Sorger PK (2009) Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol Syst Biol* **5**, 331.
- Samaga R and Klamt S (2013) Modeling approaches for qualitative and semi-quantitative analysis of cellular signaling networks. *Cell Commun Signal* **11**, 43.
- Samaga R, Saez-Rodriguez J, Alexopoulos LG, Sorger PK, and Klamt S (2009) The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. *PLoS Comput Biol* **5**, e1000438.
- Samaga R, von Kamp A, and Klamt S (2010) Computing combinatorial intervention strategies and failure modes in signaling networks. *J Comput Biol* **17**, 39–53.

* These authors contributed equally.

References

- Abramovici M, Breuer M, and Friedman A (1990) *Digital systems testing and testable design*. New York: Computer Science Press.
- Aittokallio T and Schwikowski B (2006) Graph-based methods for analysing networks in cell biology. *Brief Bioinform* **7**, 243–255.
- Albert R (2005) Scale-free networks in cell biology. *J Cell Sci* **118**, 4947–4957.
- Albert R and Othmer HG (2003) The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. *J Theor Biol* **223**, 1–18.
- Alberts B, Johnson A, Lewis J, Raff M, Roberts K, and Walter P (2004) *Molekularbiologie der Zelle*. 4th ed. Weinheim: Wiley-VCH.
- Aldridge BB, Burke JM, Lauffenburger DA, and Sorger PK (2006) Physicochemical modelling of cell signalling pathways. *Nat Cell Biol* **8**, 1195–1203.
- Aldridge BB, Saez-Rodriguez J, Muhlich JL, Sorger PK, and Lauffenburger DA (2009) Fuzzy logic analysis of kinase pathway crosstalk in TNF/EGF/insulin-induced signaling. *PLoS Comput Biol* **5**, e1000340.
- Alexopoulos LG, Saez-Rodriguez J, Cosgrove BD, Lauffenburger DA, and Sorger PK (2010) Networks inferred from biochemical data reveal profound differences in toll-like receptor and inflammatory signaling between normal and transformed hepatocytes. *Mol Cell Proteomics* **9**, 1849–1865.
- Alroy I and Yarden Y (1997) The ErbB signaling network in embryogenesis and oncogenesis: signal diversification through combinatorial ligand-receptor interactions. *FEBS Lett* **410**, 83–6.
- Andjelic M, Jakubowicz T, Cron P, Ming XF, Han JW, and Hemmings BA (1996) Activation and phosphorylation of a pleckstrin homology domain containing protein kinase (rac-pk/pkb) promoted by serum and protein phosphatase inhibitors. *Proc Natl Acad Sci U S A* **93**, 5699–5704.
- Angeli D and Sontag ED (2003) Monotone control systems. *IEEE Transactions On Automatic Control* **48**, 1684–1698.
- Angeli D, Hirsch MW, and Sontag ED (2009) Attractors in coherent systems of differential equations. *Journal Of Differential Equations* **246**, 3058–3076.
- Banaji M and Craciun G (2010) Graph-theoretic criteria for injectivity and unique equilibria in general chemical reaction systems. *Adv Appl Math* **44**, 168–184.
- Banaji M, Donnell P, and Baigent S (2007) P matrix properties, injectivity, and stability in chemical reaction systems. *SIAM J Appl Math* **67**, 1523–1547.

- Banaji M and Craciun G (2009) Graph-theoretic approaches to injectivity and multiple equilibria in systems of interacting elements. *Commun Math Sci* **7**, 867–900.
- Barabási AL and Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5**, 101–113.
- Bassett L, Maybee J, and Quirk J (1968) Qualitative economics and the scope of the correspondence principle. *Econometrica* **36**, 544–563.
- Becker V, Schilling M, Bachmann J, Baumann U, Raue A, Maiwald T, Timmer J, and Klingmüller U (2010) Covering a broad dynamic range: information processing at the erythropoietin receptor. *Science* **328**, 1404–1408.
- Beerli RR and Hynes NE (1996) Epidermal growth factor-related peptides activate distinct subsets of ErbB receptors and differ in their biological activities. *J Biol Chem* **271**, 6071–6076.
- Bertsimas D and Tsitsiklis JN (1997) *Introduction to linear optimization*. Belmont: Athena Scientific.
- Berven LA and Crouch MF (2000) Cellular function of p70S6K: a role in regulating cell motility. *Immunol Cell Biol* **78**, 447–451.
- Birchmeier C, Birchmeier W, Gherardi E, and Woude GFV (2003) Met, metastasis, motility and more. *Nat Rev Mol Cell Biol* **4**, 915–925.
- Birtwistle MR, Hatakeyama M, Yumoto N, Ogunnaike BA, Hoek JB, and Kholodenko BN (2007) Ligand-dependent responses of the ErbB signaling network: experimental and modeling analyses. *Mol Syst Biol* **3**, 144.
- Borisov N, Aksamitiene E, Kiyatkin A, Legewie S, Berkhout J, Maiwald T, Kaimachnikov NP, Timmer J, Hoek JB, and Kholodenko BN (2009) Systems-level interactions between insulin-EGF networks amplify mitogenic signaling. *Mol Syst Biol* **5**, 256.
- Borowiak M, Garratt AN, Wüstefeld T, Strehle M, Trautwein C, and Birchmeier C (2004) Met provides essential signals for liver regeneration. *Proc Natl Acad Sci U S A* **101**, 10608–10613.
- Bos JL (1989) Ras oncogenes in human cancer: a review. *Cancer Res* **49**, 4682–4689.
- Brandman O and Meyer T (2008) Feedback loops shape cellular signals in space and time. *Science* **322**, 390–395.
- Brualdi RA and Shader BL (1995) *Matrices of sign-solvable linear systems*. Cambridge Tracts in Mathematics. Cambridge Univ Press.
- Brualdi RA and Cvetković D (2009) *A combinatorial approach to matrix theory and its applications*. Discrete Mathematics and its Applications. Ed. by KH Rosen. Vol. 52. Boca Raton: Chapman & Hall/CRC.

- Buday L and Downward J (1993) Epidermal growth factor regulates p21ras through the formation of a complex of receptor, Grb2 adapter protein, and Sos nucleotide exchange factor. *Cell* **73**, 611–620.
- Burgard AP, Nikolaev EV, Schilling CH, and Maranas CD (2004) Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res* **14**, 301–312.
- Casamayor A, Morrice NA, and Alessi DR (1999) Phosphorylation of Ser-241 is essential for the activity of 3-phosphoinositide-dependent protein kinase-1: identification of five sites of phosphorylation in vivo. *Biochem J* **342**, 287–292.
- Cavigelli M, Dolfi F, Claret FX, and Karin M (1995) Induction of c-fos expression through JNK-mediated TCF/Elk-1 phosphorylation. *EMBO J* **14**, 5957–5964.
- Chan PC, Chen YL, Cheng CH, Yu KC, Cary LA, Shu KH, Ho WL, and Chen HC (2003) Src phosphorylates Grb2-associated binder 1 upon hepatocyte growth factor stimulation. *J Biol Chem* **278**, 44075–44082.
- Chaouiya C (2007) Petri net modelling of biological networks. *Brief Bioinform* **8**, 210–219.
- Chen WW, Schoeberl B, Jasper PJ, Niepel M, Nielsen UB, Lauffenburger DA, and Sorger PK (2009) Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. *Mol Syst Biol* **5**, 239.
- Chen Z, Gibson TB, Robinson F, Silvestro L, Pearson G, Xu B, Wright A, Vanderbilt C, and Cobb MH (2001) Map kinases. *Chem Rev* **101**, 2449–2476.
- Chevalier T, Schreiber T, and Ross J (1993) Toward a systematic determination of complex reaction mechanisms. *J Phys Chem* **97**, 6776–6787.
- Christensen TS, Oliveira AP, and Nielsen J (2009) Reconstruction and logical modeling of glucose repression signaling pathways in *Saccharomyces cerevisiae*. *BMC Syst Biol* **3**, 7.
- Cinquin O and Demongeot J (2002) Positive and negative feedback: striking a balance between necessary antagonists. *J Theor Biol* **216**, 229–241.
- Cirit M and Haugh JM (2012) Data-driven modelling of receptor tyrosine kinase signalling networks quantifies receptor-specific potencies of PI3K- and Ras-dependent ERK activation. *Biochem J* **441**, 77–85.
- Citri A and Yarden Y (2006) EGF-ERBB signalling: towards the systems level. *Nat Rev Mol Cell Biol* **7**, 505–16.
- Coates C (1959) Flow-graph solutions of linear algebraic equations. *IRE Trans Circuit Theory* **6**, 170–187.

- Conradi C, Saez-Rodriguez J, Gilles ED, and Raisch J (2005) Using chemical reaction network theory to discard a kinetic mechanism hypothesis. *Syst Biol (Stevenage)* **152**, 243–248.
- Conradi C and Flockerzi D (2012) Multistationarity in mass action networks with applications to ERK activation. *J Math Biol* **65**, 107–156.
- Conradi C, Flockerzi D, Raisch J, and Stelling J (2007) Subnetwork analysis reveals dynamic features of complex (bio)chemical networks. *Proc Natl Acad Sci U S A* **104**, 19175–19180.
- Coronella-Wood J, Terrand J, Sun H, and Chen QM (2004) C-Fos phosphorylation induced by H₂O₂ prevents proteasomal degradation of c-Fos in cardiomyocytes. *J Biol Chem* **279**, 33567–33574.
- Cox AD and Der CJ (2003) The dark side of Ras: regulation of apoptosis. *Oncogene* **22**, 8999–9006.
- Craciun G and Feinberg M (2005) Multiple equilibria in complex chemical reaction networks: I. The injectivity property. *Siam Journal On Applied Mathematics* **65**, 1526–1546.
- Craciun G and Feinberg M (2006a) Multiple equilibria in complex chemical reaction networks: extensions to entrapped species models. *Syst Biol (Stevenage)* **153**, 179–186.
- Craciun G and Feinberg M (2010) Multiple equilibria in complex chemical reaction networks: semiopen mass action systems. *SIAM J Appl Math* **70**, 1859–1877.
- Craciun G and Feinberg M (2006b) Multiple equilibria in complex chemical reaction networks: II. The species-reaction graph. *SIAM J Appl Math* **66**, 1321–1338.
- *D’Alessandro LA, *Samaga R, *Maiwald T, Rho SH, Bonefas S, Raue A, Iwamoto N, Kienast A, Waldow K, Meyer R, Schilling M, Timmer J, Klamt S, and Klingmüller U (2015) Disentangling the complexity of HGF signaling by combining qualitative and quantitative modeling. *PLoS Comput Biol* (4), e1004192.
- Dance M, Montagner A, Salles JP, Yart A, and Raynal P (2008) The molecular functions of Shp2 in the Ras/Mitogen-activated protein kinase (ERK1/2) pathway. *Cell Signal* **20**, 453–459.
- Datta SR, Dudek H, Tao X, Masters S, Fu H, Gotoh Y, and Greenberg ME (1997) Akt phosphorylation of BAD couples survival signals to the cell-intrinsic death machinery. *Cell* **91**, 231–241.
- De Cesare D, Jacquot S, Hanauer A, and Sassone-Corsi P (1998) Rsk-2 activity is necessary for epidermal growth factor-induced phosphorylation of CREB protein and transcription of c-fos gene. *Proc Natl Acad Sci U S A* **95**, 12202–7.

- de Jong H (2002) Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol* **9**, 67–103.
- De Jong H, Gouzé JL, Hernandez C, Page M, Sari T, and Geiselmann J (2004) Qualitative simulation of genetic regulatory networks using piecewise-linear models. *Bull Math Biol* **66**, 301–340.
- Dolfi F, Garcia-Guzman M, Ojaniemi M, Nakamura H, Matsuda M, and Vuori K (1998) The adaptor protein Crk connects multiple cellular stimuli to the JNK signaling pathway. *Proc Natl Acad Sci U S A* **95**, 15394–15399.
- Dougherty MK, Müller J, Ritt DA, Zhou M, Zhou XZ, Copeland TD, Conrads TP, Veenstra TD, Lu KP, and Morrison DK (2005) Regulation of Raf-1 by direct feedback phosphorylation. *Mol Cell* **17**, 215–224.
- Douville E and Downward J (1997) EGF induced SOS phosphorylation in PC12 cells involves P90 RSK-2. *Oncogene* **15**, 373–383.
- Downward J (1998a) Lipid-regulated kinases: some common themes at last. *Science* **279**, 673–674.
- Downward J (1998b) Ras signalling and apoptosis. *Curr Opin Genet Dev* **8**, 49–54.
- Eblen ST, Slack-Davis JK, Tarcsafalvi A, Parsons JT, Weber MJ, and Catling AD (2004) Mitogen-activated protein kinase feedback phosphorylation regulates MEK1 complex formation and activation during cellular adhesion. *Mol Cell Biol* **24**, 2308–2317.
- Edwards DC, Sanders LC, Bokoch GM, and Gill GN (1999) Activation of LIM-kinase by Pak1 couples Rac/Cdc42 GTPase signalling to actin cytoskeletal dynamics. *Nat Cell Biol* **1**, 253–259.
- Elenius K, Choi CJ, Paul S, Santiestevan E, Nishi E, and Klagsbrun M (1999) Characterization of a naturally occurring ErbB4 isoform that does not bind or activate phosphatidylinositol 3-kinase. *Oncogene* **18**, 2607–2615.
- Ellison P and Feinberg M (2000) How catalytic mechanisms reveal themselves in multiple steady-state data: I. Basic principles. *J Molec Catal A: Chem* **154**, 155–167.
- Fausto N, Campbell JS, and Riehle KJ (2006) Liver regeneration. *Hepatology* **43**, S45–S53.
- Feinberg M (1987) Chemical reaction network structure and the stability of complex isothermal reactors—I. The deficiency-zero and deficiency-one theorems. *Chemical Engineering Science* **42**, 2229–2268.
- Feliu E and Wiuf C (2013) A computational method to preclude multistationarity in networks of interacting species. *Bioinformatics* **29**, 2327–2334.

- Ferrell Jr J and Machleder EM (1998) The biochemical basis of an all-or-none cell fate switch in *Xenopus* oocytes. *Science* **280**, 895–898.
- Festa P, Pardalos P, and Resende MGC (2009) Feedback set problems. In: *Encyclopedia of optimization*. Ed. by CA Floudas and PM Pardalos. 2nd ed. New York: Springer, 1005–1016.
- Franke R, Müller M, Wundrack N, Gilles ED, Klamt S, Kähne T, and Naumann M (2008) Host-pathogen systems biology: logical modelling of hepatocyte growth factor and *Helicobacter pylori* induced c-Met signal transduction. *BMC Syst Biol* **2**, 4.
- Frödin M, Jensen CJ, Merienne K, and Gammeltoft S (2000) A phosphoserine-regulated docking site in the protein kinase RSK2 that recruits and activates PDK1. *EMBO J* **19**, 2924–2934.
- Furge KA, Zhang YW, and Woude GFV (2000) Met receptor tyrosine kinase: enhanced signaling through adapter proteins. *Oncogene* **19**, 5582–5589.
- Gaestel M (2006) MAPKAP kinases - MKs - two's company, three's a crowd. *Nat Rev Mol Cell Biol* **7**, 120–130.
- Gagneur J and Klamt S (2004) Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics* **5**, 175.
- Gao L, Feng Y, Bowers R, Becker-Hapak M, Gardner J, Council L, Linette G, Zhao H, and Cornelius LA (2006) Ras-associated protein-1 regulates extracellular signal-regulated kinase activation and migration in melanoma cells: two processes important to melanoma tumorigenesis and metastasis. *Cancer Res* **66**, 7880–7888.
- Garcia CB and Zangwill WI (1979) An approach to homotopy and degree theory. *Math Oper Res* **4**, 390–405.
- Garcia-Guzman M, Dolfi F, Zeh K, and Vuori K (1999) Met-induced JNK activation is mediated by the adapter protein Crk and correlates with the Gab1 - Crk signaling complex formation. *Oncogene* **18**, 7775–7786.
- Gebser M, Schaub T, Thiele S, and Veber P (2011) Detecting inconsistencies in large biological networks with answer set programming. *Theory and Practice of Logic Programming* **11**, 323–360.
- Gentile A, Trusolino L, and Comoglio PM (2008) The Met tyrosine kinase receptor in development and cancer. *Cancer Metastasis Rev* **27**, 85–94.
- Gerwins P, Blank JL, and Johnson GL (1997) Cloning of a novel mitogen-activated protein kinase kinase kinase, MEKK4, that selectively regulates the c-Jun amino terminal kinase pathway. *J Biol Chem* **272**, 8288–8295.

- Giacomantonio CE and Goodhill GJ (2010) A Boolean model of the gene regulatory network underlying Mammalian cortical area development. *PLoS Comput Biol* **6**, e10000936.
- Glass L and Kauffman S (1973) The logical analysis of continuous, non-linear biochemical control networks. *J Theor Biol* **39**, 103–129.
- Gouzé JL (1998) Positive and negative circuits in dynamical systems. *J Biol Syst* **6**, 11–15.
- Graus-Porta D, Beerli RR, Daly JM, and Hynes NE (1997) ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. *EMBO J* **16**, 1647–55.
- Grimes CA and Jope RS (2001) The multifaceted roles of glycogen synthase kinase 3beta in cellular signaling. *Prog Neurobiol* **65**, 391–426.
- Guziolowski C, Videla S, Eduati F, Thiele S, Cokelear T, Siegel A, and Saez-Rodriguez J (2013) Exhaustively characterizing feasible logic models of a signaling network using Answer Set Programming. *Bioinformatics* **29** (18), 2320–2326.
- Guziolowski C, Bourdé A, Moreews F, and Siegel A (2009) BioQuali Cytoscape plugin: analysing the global consistency of regulatory networks. *BMC Genomics* **10**, 244.
- Hädicke O, Grammel H, and Klamt S (2011) Metabolic network modeling of redox balancing and biohydrogen production in purple nonsulfur bacteria. *BMC Syst Biol* **5**, 150.
- Harari D, Tzahar E, Romano J, Shelly M, Pierce JH, Andrews GC, and Yarden Y (1999) Neuregulin-4: a novel growth factor that acts through the ErbB-4 receptor tyrosine kinase. *Oncogene* **18**, 2681–2689.
- Harary F (1962) The determinant of the adjacency matrix of a graph. *SIAM Rev* **4**, 202–210.
- Hardy S and Robillard PN (2004) Modeling and simulation of molecular biology systems using Petri nets: modeling goals of various approaches. *J Bioinform Comput Biol* **2**, 595–613.
- Hatakeyama M, Kimura S, Naka T, Kawasaki T, Yumoto N, Ichikawa M, Kim JH, Saito K, Saeki M, Shirouzu M, Yokoyama S, and Konagaya A (2003) A computational model on the modulation of mitogen-activated protein kinase (MAPK) and Akt pathways in heregulin-induced ErbB signalling. *Biochem J* **373**, 451–463.
- Hay N and Sonenberg N (2004) Upstream and downstream of mTOR. *Genes Dev* **18**, 1926–1945.

- Hendriks BS, Cook J, Burke JM, Beusmans JM, Lauffenburger DA, and Graaf D de (2006) Computational modelling of ErbB family phosphorylation dynamics in response to transforming growth factor alpha and heregulin indicates spatial compartmentation of phosphatase activity. *Syst Biol (Stevenage)* **153**, 22–33.
- Higuchi M, Onishi K, Kikuchi C, and Gotoh Y (2008) Scaffolding function of PAK in the PDK1-Akt pathway. *Nat Cell Biol* **10**, 1356–1364.
- Hoffmann A, Levchenko A, Scott ML, and Baltimore D (2002) The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science* **298**, 1241–1245.
- Horn F and Jackson R (1972) General mass action kinetics. *Arch Rational Mech Anal* **49**, 81–116.
- Hou Z, He L, and Qi RZ (2007) Regulation of S6 kinase 1 activation by phosphorylation at Ser-411. *J Biol Chem* **282**, 6922–6928.
- Huang CY and Ferrell Jr J (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc Natl Acad Sci U S A* **93**, 10078–10083.
- Huard J, Mueller S, Gilles ED, Klingmüller U, and Klamt S (2012) An integrative model links multiple inputs and signaling pathways to the onset of DNA synthesis in hepatocytes. *FEBS J* **279**, 3290–3313.
- Innocenti M, Tenca P, Frittoli E, Faretta M, Tocchetti A, Fiore PPD, and Scita G (2002) Mechanisms through which Sos-1 coordinates the activation of Ras and Rac. *J Cell Biol* **156**, 125–136.
- Innocenti M, Frittoli E, Ponzanelli I, Falck JR, Brachmann SM, Fiore PPD, and Scita G (2003) Phosphoinositide 3-kinase activates Rac by entering in a complex with Eps8, Abi1, and Sos-1. *J Cell Biol* **160**, 17–23.
- Iri M, Aoki K, O'Shima E, and Matsuyama H (1979) An algorithm for diagnosis of system failures in the chemical process. *Comput. Chem. Eng.* **3**, 489–493.
- Jeong H, Mason SP, Barabási AL, and Oltvai ZN (2001) Lethality and centrality in protein networks. *Nature* **411**, 41–42.
- Jones JT, Akita RW, and Sliwkowski MX (1999) Binding specificities and affinities of egf domains for ErbB receptors. *FEBS Lett* **447**, 227–231.
- Joshi B and Shiu A (2012) Simplifying the jacobian criterion for precluding multistationarity in chemical reaction networks. *SIAM J Appl Math* **72**, 857–876.
- Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, Jassal B, Gopinath GR, Wu GR, Matthews L, Lewis S, Birney E, and Stein L (2005) Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res* **33**, D428–D432.

- Karin M, Liu Z, and Zandi E (1997) AP-1 function and regulation. *Curr Opin Cell Biol* **9**, 240–246.
- Kauffman SA (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol* **22**, 437–467.
- Keilhack H, Tenev T, Nyakatura E, Godovac-Zimmermann J, Nielsen L, Seedorf K, and Böhmer FD (1998) Phosphotyrosine 1173 mediates binding of the protein-tyrosine phosphatase SHP-1 to the epidermal growth factor receptor and attenuation of receptor signaling. *J Biol Chem* **273**, 24839–24846.
- Kelder T, van Iersel MP, Hanspers K, Kutmon M, Conklin BR, Evelo CT, and Pico AR (2012) WikiPathways: building research communities on biological pathways. *Nucleic Acids Res* **40**, D1301–D1307.
- Kestler HA, Wawra C, Kracher B, and Köhl M (2008) Network modeling of signal transduction: establishing the global view. *Bioessays* **30**, 1110–1125.
- Kholodenko BN (2000) Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur J Biochem* **267**, 1583–1588.
- Kholodenko BN, Demin OV, Moehren G, and Hoek JB (1999) Quantification of short term signaling by the epidermal growth factor receptor. *J Biol Chem* **274**, 30169–30181.
- Kholodenko BN, Kiyatkin A, Bruggeman FJ, Sontag E, Westerhoff HV, and Hoek JB (2002) Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. *Proc Natl Acad Sci U S A* **99**, 12841–12846.
- Kim JW, Sim SS, Kim UH, Nishibe S, Wahl MI, Carpenter G, and Rhee SG (1990) Tyrosine residues in bovine phospholipase C-gamma phosphorylated by the epidermal growth factor receptor in vitro. *J Biol Chem* **265**, 3940–3943.
- Kim MJ, Kim E, Ryu SH, and Suh PG (2000) The mechanism of phospholipase C-gamma1 regulation. *Exp Mol Med* **32**, 101–109.
- King AJ, Wireman RS, Hamilton M, and Marshall MS (2001) Phosphorylation site specificity of the Pak-mediated regulation of Raf-1 and cooperativity with Src. *FEBS Lett* **497**, 6–14.
- Kishimoto H, Nakagawa K, Watanabe T, Kitagawa D, Momose H, Seo J, Nishitai G, Shimizu N, Ohata S, Tanemura S, Asaka S, Goto T, Fukushi H, Yoshida H, Suzuki A, Sasaki T, Wada T, Penninger JM, Nishina H, and Katada T (2003) Different properties of SEK1 and MKK7 in dual phosphorylation of stress-induced activated protein kinase SAPK/JNK in embryonic stem cells. *J Biol Chem* **278**, 16595–16601.

- Klamt S (2006) Generalized concept of minimal cut sets in biochemical networks. *Biosystems* **83**, 233–247.
- Klamt S and von Kamp A (2009) Computing paths and cycles in biological interaction graphs. *BMC Bioinformatics* **10**, 181.
- Klamt S and von Kamp A (2011) An application programming interface for CellNetAnalyzer. *Biosystems* **105**, 162–168.
- Klamt S, Saez-Rodriguez J, Lindquist JA, Simeoni L, and Gilles ED (2006) A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics* **7**, 56.
- Klamt S, Saez-Rodriguez J, and Gilles ED (2007) Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Syst Biol* **1**, 2.
- Klamt S, Haus UU, and Theis F (2009) Hypergraphs and cellular networks. *PLoS Comput Biol* **5**, e1000385.
- Klinger B, Sieber A, Fritsche-Guenther R, Witzel F, Berry L, Schumacher D, Yan Y, Durek P, Merchant M, Schäfer R, Sers C, and Blüthgen N (2013) Network quantification of EGFR signaling unveils potential for targeted combination therapy. *Mol Syst Biol* **9**, 673.
- Knight D, Shinar G, and Feinberg M (2015) Sharper graph-theoretical conditions for the stabilization of complex reaction networks. *Math Biosci* **262**, 10–27.
- Kolczyk K, Samaga R, Conzelmann H, Mirschel S, and Conradi C (2012) The process-interaction-model: a common representation of rule-based and logical models allows studying signal transduction on different levels of detail. *BMC Bioinformatics* **13**, 251.
- Krauss G (2003) *Biochemistry of signal transduction and regulation*. 3rd ed. Weinheim: Wiley-VCH.
- Kuepfer L, Peter M, Sauer U, and Stelling J (2007) Ensemble modeling for analysis of cell signaling dynamics. *Nat Biotechnol* **25**, 1001–1006.
- Kuipers B (1994) *Qualitative reasoning: modeling and simulation with incomplete knowledge*. Cambridge, MA: MIT Press.
- Kunze H and Siegel D (1994) A graph-theoretical approach to monotonicity with respect to initial conditions. In: *Comparison methods and stability theory*. Ed. by X Liu and D Siegel. Vol. 162. New York: Marcel Dekker, 207–216.
- Landau M and Ben-Tal N (2008) Dynamic equilibrium between multiple active and inactive conformations explains regulation and oncogenic mutations in ErbB receptors. *Biochim Biophys Acta* **1785**, 12–31.

- Lanzetti L, Rybin V, Malabarba MG, Christoforidis S, Scita G, Zerial M, and Fiore PPD (2000) The Eps8 protein coordinates EGF receptor signalling through Rac and trafficking through Rab5. *Nature* **408**, 374–377.
- Lenferink AE, Pinkas-Kramarski R, van de Poll ML, van Vugt MJ, Klapper LN, Tzahar E, Waterman H, Sela M, van Zoelen EJ, and Yarden Y (1998) Differential endocytic routing of homo- and hetero-dimeric ErbB tyrosine kinases confers signaling superiority to receptor heterodimers. *17*, 3385–3397.
- Li F, Long T, Lu Y, Ouyang Q, and Tang C (2004) The yeast cell-cycle network is robustly designed. *Proc Natl Acad Sci U S A* **101**, 4781–4786.
- Li N, Batzer A, Daly R, Yajnik V, Skolnik E, Chardin P, Bar-Sagi D, Margolis B, and Schlessinger J (1993) Guanine-nucleotide-releasing factor hSos1 binds to Grb2 and links receptor tyrosine kinases to Ras signalling. *Nature* **363**, 85–88.
- Li W, Fan J, and Woodley DT (2001) Nck/Dock: an adapter between cell surface receptors and the actin cytoskeleton. *Oncogene* **20**, 6403–6417.
- Lu X, Nemoto S, and Lin A (1997) Identification of c-Jun NH2-terminal protein kinase (jnk)-activating kinase 2 as an activator of JNK but not p38. *J Biol Chem* **272**, 24751–24754.
- Mabuchi S, Ohmichi M, Kimura A, Hisamoto K, Hayakawa J, Nishio Y, Adachi K, Takahashi K, Arimoto-Ishida E, Nakatsuji Y, Tasaka K, and Murata Y (2002) Inhibition of phosphorylation of BAD and Raf-1 by Akt sensitizes human ovarian cancer cells to paclitaxel. *J Biol Chem* **277**, 33490–33500.
- MacNamara A, Terfve C, Henriques D, Bernabé BP, and Saez-Rodriguez J (2012) State-time spectrum of signal transduction logic models. *Phys Biol* **9**, 045003.
- Maroun CR, Holgado-Madruga M, Royal I, Naujokas MA, Fournier TM, Wong AJ, and Park M (1999) The Gab1 PH domain is required for localization of Gab1 at sites of cell-cell contact and epithelial morphogenesis downstream from the met receptor tyrosine kinase. *Mol Cell Biol* **19**, 1784–1799.
- Maurya MR, Rengaswamy R, and Venkatasubramanian V (2003) A systematic framework for the development and analysis of signed digraphs for chemical processes. 1. Algorithms and analysis. *Ind Eng Chem Res* **42**, 4789–4810.
- Maybee J and Quirk J (1969) Qualitative problems in matrix theory. *SIAM Rev* **11**, 30–51.
- McCluskey E and Clegg F (1971) Fault equivalence in combinatorial logic networks. *IEEE T Comput* **C-20**, 1286–1293.

- *Melas IN, *Samaga R, Alexopoulos LG, and Klamt S (2013) Detecting and removing inconsistencies between experimental data and signaling network topologies using integer linear programming on interaction graphs. *PLoS Comput Biol* **9**, e1003204.
- Mendoza L and Xenarios I (2006) A method for the generation of standardized qualitative dynamical systems of regulatory networks. *Theor Biol Med Model* **3**, 13.
- Mendoza MC, Er EE, and Blenis J (2011) The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation. *Trends Biochem Sci* **36**, 320–328.
- Meyer T and Stryer L (1988) Molecular model for receptor-stimulated calcium spiking. *Proc Natl Acad Sci U S A* **85**, 5051–5055.
- Michalopoulos GK (2007) Liver regeneration. *J Cell Physiol* **213**, 286–300.
- Michalopoulos GK and Khan Z (2005) Liver regeneration, growth factors, and amphiregulin. *Gastroenterology* **128**, 503–506.
- Mirschel S, Steinmetz K, Rempel M, Ginkel M, and Gilles ED (2009) PROMOT: modular modeling for systems biology. *Bioinformatics* **25**, 687–689.
- Mitsos A, Melas IN, Siminelakis P, Chairakaki AD, Saez-Rodriguez J, and Alexopoulos LG (2009) Identifying drug effects via pathway alterations using an integer linear programming optimization formulation on phosphoproteomic data. *PLoS Comput Biol* **5**, e1000591.
- Montagner A, Yart A, Dance M, Perret B, Salles JP, and Raynal P (2005) A novel role for Gab1 and SHP2 in epidermal growth factor-induced Ras activation. *J Biol Chem* **280**, 5350–5360.
- Morris MK, Saez-Rodriguez J, Sorger PK, and Lauffenburger DA (2010) Logic-based models for the analysis of cell signaling networks. *Biochemistry* **49**, 3216–3224.
- Morris MK, Saez-Rodriguez J, Clarke DC, Sorger PK, and Lauffenburger DA (2011) Training signaling pathway maps to biochemical data with constrained fuzzy logic: quantitative analysis of liver cell responses to inflammatory stimuli. *PLoS Comput Biol* **7**, e1001099.
- Mukhopadhyay NK, Price DJ, Kyriakis JM, Pelech S, Sanghera J, and Avruch J (1992) An array of insulin-activated, proline-directed serine/threonine protein kinases phosphorylate the p70 S6 kinase. *J Biol Chem* **267**, 3325–3335.
- Murphy LO, Smith S, Chen RH, Fingar DC, and Blenis J (2002) Molecular interpretation of ERK signal duration by immediate early gene products. *Nat Cell Biol* **4**, 556–564.
- Naldi A, Carneiro J, Chaouiya C, and Thieffry D (2010) Diversity and plasticity of Th cell types predicted from regulatory network modelling. *PLoS Comput Biol* **6**, e1000912.

- Naldini L, Vigna E, Ferracini R, Longati P, Gandino L, Prat M, and Comoglio PM (1991) The tyrosine kinase encoded by the met proto-oncogene is activated by autophosphorylation. *Mol Cell Biol* **11**, 1793–1803.
- Nelander S, Wang W, Nilsson B, She QB, Pratilas C, Rosen N, Gennemark P, and Sander C (2008) Models from experiments: combinatorial drug perturbations of cancer cells. *Mol Syst Biol* **4**, 216.
- Newton AC (2003) Regulation of the ABC kinases by phosphorylation: protein kinase C as a paradigm. *Biochem J* **370**, 361–371.
- Novák B and Tyson JJ (2008) Design principles of biochemical oscillators. *Nat Rev Mol Cell Biol* **9**, 981–991.
- Oda K, Matsuoka Y, Funahashi A, and Kitano H (2005) A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol Syst Biol* **1**, 2005.0010.
- Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, and Kanehisa M (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **27**, 29–34.
- Okabayashi Y, Kido Y, Okutani T, Sugimoto Y, Sakaguchi K, and Kasuga M (1994) Tyrosines 1148 and 1173 of activated human epidermal growth factor receptors are binding sites of Shc in intact cells. *J Biol Chem* **269**, 18674–18678.
- Olayioye MA, Graus-Porta D, Beerli RR, Rohrer J, Gay B, and Hynes NE (1998) ErbB-1 and ErbB-2 acquire distinct signaling properties dependent upon their dimerization partner. *Mol Cell Biol* **18**, 5042–5051.
- Olayioye MA, Beuvink I, Horsch K, Daly JM, and Hynes NE (1999) ErbB receptor-induced activation of Stat transcription factors is mediated by Src tyrosine kinases. *J Biol Chem* **274**, 17209–17218.
- Olayioye MA, Neve RM, Lane HA, and Hynes NE (2000) The ErbB signaling network: receptor heterodimerization in development and cancer. *EMBO J* **19**, 3159–3167.
- Oyeleye OO and Kramer MA (1988) Qualitative simulation of chemical process systems: steady-state analysis. *AIChE J* **34**, 1441–1454.
- Oyeleye OO (1990) Qualitative modeling of continuous chemical processes and applications to fault diagnosis. Sc.D. Thesis. Massachusetts Institute of Technology.
- Papin JA and Palsson BO (2004) Topological analysis of mass-balanced signaling networks: a framework to obtain network properties including crosstalk. *J Theor Biol* **227**, 283–297.
- Pawson T and Warner N (2007) Oncogenic re-wiring of cellular signaling pathways. *Oncogene* **26**, 1268–1275.

- Pfeiffer T, Sánchez-Valdenebro I, Nuño JC, Montero F, and Schuster S (1999) META-TOOL: for studying metabolic networks. *Bioinformatics* **15**, 251–257.
- Pinkas-Kramarski R, Shelly M, Guarino BC, Wang LM, Lyass L, Alroy I, Alimandi M, Kuo A, Moyer JD, Lavi S, Eisenstein M, Ratzkin BJ, Seger R, Bacus SS, Pierce JH, Andrews GC, Yarden Y, and Alimandi M (1998) ErbB tyrosine kinases and the two neuregulin families constitute a ligand-receptor network. *Mol Cell Biol* **18**, 6090–6101.
- Plahte E, Mestl T, and Omholt SW (1995) Feedback loops, stability and multistationarity in dynamical systems. *J Biol Syst* **3**, 409–413.
- Ponzetto C, Bardelli A, Maina F, Longati P, Panayotou G, Dhand R, Waterfield MD, and Comoglio PM (1993) A novel recognition motif for phosphatidylinositol 3-kinase binding mediates its association with the hepatocyte growth factor/scatter factor receptor. *Mol Cell Biol* **13**, 4600–4608.
- Ponzetto C, Bardelli A, Zhen Z, Maina F, Zonca P dalla, Giordano S, Graziani A, Panayotou G, and Comoglio PM (1994) A multifunctional docking site mediates signaling and transformation by the hepatocyte growth factor/scatter factor receptor family. *Cell* **77**, 261–271.
- Puto LA, Pestonjamas K, King CC, and Bokoch GM (2003) P21-activated kinase 1 (pak1) interacts with the grb2 adapter protein to couple to growth factor signaling. *J Biol Chem* **278**, 9388–9393.
- Radde N, Bar NS, and Banaji M (2010) Graphical methods for analysing feedback in biological networks—A survey. *Int J Syst Sci* **41**, 35–46.
- Radulescu O, Lagarrigue S, Siegel A, Veber P, and Borgne ML (2006) Topology and static response of interaction networks in molecular biology. *J R Soc Interface* **3**, 185–196.
- Raingaud J, Whitmarsh AJ, Barrett T, Dérijard B, and Davis RJ (1996) MKK3- and MKK6-regulated gene expression is mediated by the p38 mitogen-activated protein kinase signal transduction pathway. *Mol Cell Biol* **16**, 1247–1255.
- Raue A, Schilling M, Bachmann J, Matteson A, Schelke M, Kaschek D, Hug S, Kreutz C, Harms BD, Theis FJ, Klingmüller U, and Timmer J (2013) Lessons learned from quantitative dynamic modeling in Systems Biology. *PLoS ONE* **8** (9), e74335.
- Resat H, Ewald JA, Dixon DA, and Wiley HS (2003) An integrated model of epidermal growth factor receptor trafficking and signal transduction. *Biophys J* **85**, 730–743.
- Richard A and Comet JP (2011) Stable periodicity and negative circuits in differential systems. *J Math Biol* **63** (3), 593–600.

- Robinson MJ and Cobb MH (1997) Mitogen-activated protein kinase pathways. *Curr Opin Cell Biol* **9**, 180–186.
- Rodrigues GA, Falasca M, Zhang Z, Ong SH, and Schlessinger J (2000) A novel positive feedback loop mediated by the docking protein Gab1 and phosphatidylinositol 3-kinase in epidermal growth factor receptor signaling. *Mol Cell Biol* **20**, 1448–1459.
- Rosário M and Birchmeier W (2003) How to make tubes: signaling by the Met receptor tyrosine kinase. *Trends Cell Biol* **13**, 328–335.
- Rossman KL, Der CJ, and Sondek J (2005) GEF means go: turning on RHO GTPases with guanine nucleotide-exchange factors. *Nat Rev Mol Cell Biol* **6**, 167–180.
- Rumschinski P, Streif S, and Findeisen R (2012) Combining qualitative information and semi-quantitative data for guaranteed invalidation of biochemical network models. *Int J Robust Nonlinear Control* **22**, 1157–1173.
- Ryll A, Samaga R, Schaper F, Alexopoulos LG, and Klamt S (2011) Large-scale network models of IL-1 and IL-6 signalling and their hepatocellular specification. *Mol Biosyst* **7**, 3253–3270.
- Sachs K, Perez O, Pe'er D, Lauffenburger DA, and Nolan GP (2005) Causal protein-signaling networks derived from multiparameter single-cell data. *Science* **308**, 523–529.
- Saez-Rodriguez J, Simeoni L, Lindquist JA, Hemenway R, Bommhardt U, Arndt B, Haus UU, Weismantel R, Gilles ED, Klamt S, and Schraven B (2007) A logical model provides insights into T cell receptor signaling. *PLoS Comput Biol* **3**, e163.
- Saez-Rodriguez J, Goldsipe A, Muhlich J, Alexopoulos LG, Millard B, Lauffenburger DA, and Sorger PK (2008) Flexible informatics for linking experimental data to mathematical models via DataRail. *Bioinformatics* **24**, 840–847.
- Saez-Rodriguez J, Alexopoulos LG, Epperlein J, Samaga R, Lauffenburger DA, Klamt S, and Sorger PK (2009) Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol Syst Biol* **5**, 331.
- Sakkab D, Lewitzky M, Posern G, Schaeper U, Sachs M, Birchmeier W, and Feller SM (2000) Signaling of hepatocyte growth factor/scatter factor (HGF) to the small GTPase Rap1 via the large docking protein Gab1 and the adapter protein CRKL. *J Biol Chem* **275**, 10772–10778.
- Samaga R and Klamt S (2013) Modeling approaches for qualitative and semi-quantitative analysis of cellular signaling networks. *Cell Commun Signal* **11**, 43.

- Samaga R, Saez-Rodriguez J, Alexopoulos LG, Sorger PK, and Klamt S (2009) The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. *PLoS Comput Biol* **5**, e1000438.
- Samaga R, von Kamp A, and Klamt S (2010) Computing combinatorial intervention strategies and failure modes in signaling networks. *J Comput Biol* **17**, 39–53.
- Santos SDM, Verveer PJ, and Bastiaens PIH (2007) Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate. *Nat Cell Biol* **9**, 324–330.
- Sarbassov DD, Guertin DA, Ali SM, and Sabatini DM (2005a) Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. *Science* **307**, 1098–1101.
- Sarbassov DD, Ali SM, and Sabatini DM (2005b) Growing roles for the mTOR pathway. *Curr Opin Cell Biol* **17**, 596–603.
- Sato Ki, Nagao T, Kakumoto M, Kimoto M, Otsuki T, Iwasaki T, Tokmakov AA, Owada K, and Fukami Y (2002) Adaptor protein Shc is an isoform-specific direct activator of the tyrosine kinase c-Src. *J Biol Chem* **277**, 29568–29576.
- Sato S, Fujita N, and Tsuruo T (2004) Involvement of 3-phosphoinositide-dependent protein kinase-1 in the MEK/MAPK signal transduction pathway. *J Biol Chem* **279**, 33759–67.
- Scheid MP and Woodgett JR (2003) Unravelling the activation mechanisms of protein kinase B/Akt. *FEBS Lett* **546**, 108–112.
- Schilling M, Maiwald T, Bohl S, Kollmann M, Kreutz C, Timmer J, and Klingmüller U (2005) Quantitative data generation for systems biology: the impact of randomisation, calibrators and normalisers. *Syst Biol (Stevenage)* **152**, 193–200.
- Schlatter R, Schmich K, Avalos Vizcarra I, Scheurich P, Sauter T, Borner C, Ederer M, Merfort I, and Sawodny O (2009) ON/OFF and beyond—a Boolean model of apoptosis. *PLoS Comput Biol* **5**, e1000595.
- Schlesinger TK, Fanger GR, Yujiri T, and Johnson GL (1998) The TAO of MEKK. *Front Biosci* **3**, D1181–D1186.
- Schoeberl B, Eichler-Jonsson C, Gilles ED, and Müller G (2002) Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat Biotechnol* **20**, 370–375.
- Schoeberl B, Pace EA, Fitzgerald JB, Harms BD, Xu L, Nie L, Linggi B, Kalra A, Paragas V, Bukhalid R, Grantcharova V, Kohli N, West KA, Leszczyniecka M, Feldhaus MJ, Kudla AJ, and Nielsen UB (2009) Therapeutically targeting ErbB3: a key node in ligand-induced activation of the ErbB receptor-PI3K axis. *Sci Signal* **2**, ra31.

- Schulze WX, Deng L, and Mann M (2005) Phosphotyrosine interactome of the ErbB-receptor kinase family. *Mol Syst Biol* **1**, 2005.0008.
- Schürmann A, Mooney AF, Sanders LC, Sells MA, Wang HG, Reed JC, and Bokoch GM (2000) P21-activated kinase 1 phosphorylates the death agonist Bad and protects cells from apoptosis. *Mol Cell Biol* **20**, 453–461.
- Schuster S, Dandekar T, and Fell DA (1999) Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol* **17**, 53–60.
- Sears R, Nuckolls F, Haura E, Taya Y, Tamai K, and Nevins JR (2000) Multiple Ras-dependent phosphorylation pathways regulate Myc protein stability. *Genes Dev* **14**, 2501–2514.
- Shelly M, Pinkas-Kramarski R, Guarino BC, Waterman H, Wang LM, Lyass L, Alimandi M, Kuo A, Bacus SS, Pierce JH, Andrews GC, and Yarden Y (1998) Epregrulin is a potent pan-ErbB ligand that preferentially activates heterodimeric receptor complexes. *J Biol Chem* **273**, 10496–10505.
- Shin SY, Rath O, Choo SM, Fee F, McFerran B, Kolch W, and Cho KH (2009) Positive and negative-feedback regulations coordinate the dynamic behavior of the Ras-Raf-MEK-ERK signal transduction pathway. *J Cell Sci* **122**, 425–435.
- Shinar G and Feinberg M (2012) Concordant chemical reaction networks. *Math Biosci* **240**, 92–113.
- Shinar G and Feinberg M (2013) Concordant chemical reaction networks and the species–reaction graph. *Math Biosci* **241**, 1–23.
- Shiomura T (1995) Comparative statics in the large based on a fixed-point algorithm. *Journal Of Economics-Zeitschrift Fur Nationalokonomie* **62**, 77–92.
- Shoyab M, Plowman GD, McDonald VL, Bradley JG, and Todaro GJ (1989) Structure and function of human amphiregulin: a member of the epidermal growth factor family. *Science* **243**, 1074–1076.
- Siegel A, Radulescu O, Borgne ML, Veber P, Ouy J, and Lagarrigue S (2006) Qualitative analysis of the relation between DNA microarray data and behavioral models of regulation networks. *Biosystems* **84**, 153–174.
- Snooussi EH (1998) Necessary conditions for multistationarity and stable periodicity. *J Biol Syst* **6**, 3–9.
- Sontag ED (2005) Molecular systems biology and control. *European Journal Of Control* **11**, 396–435.
- Soulé C (2003) Graphic requirements for multistationarity. *ComplexUs* **1**, 123–133.

- Stokoe D, Engel K, Campbell DG, Cohen P, and Gaestel M (1992) Identification of MAPKAP kinase 2 as a major enzyme responsible for the phosphorylation of the small mammalian heat shock proteins. *FEBS Lett* **313**, 307–313.
- Stork PJS (2003) Does Rap1 deserve a bad Rap? *Trends Biochem Sci* **28**, 267–275.
- Sun H, Charles CH, Lau LF, and Tonks NK (1993) MKP-1 (3CH134), an immediate early gene product, is a dual specificity phosphatase that dephosphorylates MAP kinase in vivo. *Cell* **75**, 487–493.
- Swameye I, Müller TG, Timmer J, Sandra O, and Klingmüller U (2003) Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. *Proc Natl Acad Sci U S A* **100**, 1028–1033.
- Tall GG, Barbieri MA, Stahl PD, and Horazdovsky BF (2001) Ras-activated endocytosis is mediated by the Rab5 guanine nucleotide exchange activity of RIN1. *Dev Cell* **1**, 73–82.
- Tamás P, Solti Z, Bauer P, Illés A, Sipeki S, Bauer A, Faragó A, Downward J, and Buday L (2003) Mechanism of epidermal growth factor regulation of Vav2, a guanine nucleotide exchange factor for Rac. *J Biol Chem* **278**, 5163–5171.
- Tan Y, Rouse J, Zhang A, Cariati S, Cohen P, and Comb MJ (1996) FGF and stress regulate CREB and ATF-1 via a pathway involving p38 MAP kinase and MAPKAP kinase-2. *EMBO J* **15**, 4629–4642.
- Taub R (2004) Liver regeneration: from myth to mechanism. *Nat Rev Mol Cell Biol* **5**, 836–847.
- Terfve CDA, Cokelaer T, Henriques D, Macnamara A, Gonçalves E, Morris MK, van Iersel M, Lauffenburger DA, and Saez-Rodriguez J (2012) Cellnoptr: a flexible toolkit to train protein signaling networks to data using multiple logic formalisms. *BMC Syst Biol* **6**, 133.
- Thieffry D (2007) Dynamical roles of biological regulatory circuits. *Brief Bioinform* **8**, 220–225.
- Thiele S, Cerone L, Saez-Rodriguez J, Siegel A, Guziolowski C, and Klamt S (2015) Extended notions of sign consistency to relate experimental data to signaling and regulatory network topologies. *BMC Bioinformatics* **16**, 345.
- Thomas R (1973) Boolean formalization of genetic control circuits. *J Theor Biol* **42**, 563–585.
- Thomas R and D’Ari R (1990) *Biological feedback*. Boca Raton: CRC Press.
- Thomas R and Kaufman M (2001a) Multistationarity, the basis of cell differentiation and memory. I. Structural conditions of multistationarity and other nontrivial behavior. *Chaos* **11**, 170–179.

- Thomas R and Kaufman M (2001b) Multistationarity, the basis of cell differentiation and memory. II. Logical analysis of regulatory networks in terms of feedback circuits. *Chaos* **11**, 180–195.
- Tian J and Karin M (1999) Stimulation of Elk1 transcriptional activity by mitogen-activated protein kinases is negatively regulated by protein phosphatase 2B (calcineurin). *J Biol Chem* **274**, 15173–15180.
- Tibbles LA, Ing YL, Kiefer F, Chan J, Iscove N, Woodgett JR, and Lassam NJ (1996) MLK-3 activates the SAPK/JNK and p38/RK pathways via SEK1 and MKK3/6. *EMBO J* **15**, 7026–7035.
- Tolias KF and Cantley LC (1999) Pathways for phosphoinositide synthesis. *Chem Phys Lipids* **98**, 69–77.
- Tyson JJ, Chen KC, and Novak B (2003) Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr Opin Cell Biol* **15**, 221–231.
- Vacratis PO, Phinney BS, Gage DA, and Gallo KA (2002) Identification of in vivo phosphorylation sites of MLK3 by mass spectrometry and phosphopeptide mapping. *Biochemistry* **41**, 5613–5624.
- Vanhaesebroeck B, Leever SJ, Panayotou G, and Waterfield MD (1997) Phosphoinositide 3-kinases: a conserved family of signal transducers. *Trends Biochem Sci* **22**, 267–272.
- Veber P, Guziolowski C, Borgne ML, Radulescu O, and Siegel A (2008) Inferring the role of transcription factors in regulatory networks. *BMC Bioinformatics* **9**, 228.
- Vogelstein B, Lane D, and Levine AJ (2000) Surfing the p53 network. *Nature* **408**, 307–310.
- Wang LM, Kuo A, Alimandi M, Veri MC, Lee CC, Kapoor V, Ellmore N, Chen XH, and Pierce JH (1998) ErbB2 expression increases the spectrum and potency of ligand-mediated signal transduction through ErbB4. *Proc Natl Acad Sci U S A* **95**, 6809–6814.
- Wang RS, Saadatpour A, and Albert R (2012) Boolean modeling in systems biology: an overview of methodology and applications. *Phys Biol* **9**, 055001.
- Weng QP, Kozlowski M, Belham C, Zhang A, Comb MJ, and Avruch J (1998) Regulation of the p70 S6 kinase by phosphorylation in vivo. Analysis using site-specific anti-phosphopeptide antibodies. *J Biol Chem* **273**, 16621–16629.
- Wick MJ, Wick KR, Chen H, He H, Dong LQ, Quon MJ, and Liu F (2002) Substitution of the autophosphorylation site Thr516 with a negatively charged residue confers

- constitutive activity to mouse 3-phosphoinositide-dependent protein kinase-1 in cells. *J Biol Chem* **277**, 16632–16638.
- Wiley HS and Cunningham DD (1981) A steady state model for analyzing the cellular binding, internalization and degradation of polypeptide ligands. *Cell* **25**, 433–440.
- Wiley HS, Shvartsman SY, and Lauffenburger DA (2003) Computational modeling of the EGF-receptor system: a paradigm for systems biology. *Trends Cell Biol* **13**, 43–50.
- Wittmann DM, Krumsiek J, Saez-Rodriguez J, Lauffenburger DA, Klamt S, and Theis FJ (2009) Transforming Boolean models to continuous models: methodology and application to T-cell receptor signaling. *BMC Syst Biol* **3**, 98.
- Wiuf C and Feliu E (2013) Power-law kinetics and determinant criteria for the preclusion of multistationarity in networks of interacting species. *SIAM J Appl Dyn Syst* **12** (4), 1685–1721.
- Wofsy C, Goldstein B, Lund K, and Wiley HS (1992) Implications of epidermal growth factor (EGF) induced egf receptor aggregation. *Biophys J* **63**, 98–110.
- Won JK, Yang HW, Shin SY, Lee JH, Heo WD, and Cho KH (2012) The crossregulation between ERK and PI3K signaling pathways determines the tumoricidal efficacy of MEK inhibitor. *J Mol Cell Biol* **4**, 153–163.
- Wynn ML, Consul N, Merajver SD, and Schnell S (2012) Logic-based models in systems biology: a predictive and parameter-free network analysis method. *Integr Biol (Camb)* **4**, 1323–1337.
- Xu TR, Vyshemirsky V, Gormand A, von Kriegsheim A, Girolami M, Baillie GS, Ketley D, Dunlop AJ, Milligan G, Houslay MD, and Kolch W (2010) Inferring signaling pathway topologies from multiple perturbation measurements of specific biochemical species. *Sci Signal* **3**, ra20.
- Yeung K, Seitz T, Li S, Janosch P, McFerran B, Kaiser C, Fee F, Katsanakis KD, Rose DW, Mischak H, Sedivy JM, and Kolch W (1999) Suppression of Raf-1 kinase activity and MAP kinase signalling by RKIP. *Nature* **401**, 173–177.
- Yu CF, Roshan B, Liu ZX, and Cantley LG (2001) ERK regulates the hepatocyte growth factor-mediated interaction of Gab1 and the phosphatidylinositol 3-kinase. *J Biol Chem* **276**, 32552–32558.
- Zhang R, Shah MV, Yang J, Nyland SB, Liu X, Yun JK, Albert R, and Loughran Jr TP (2008) Network model of survival signaling in large granular lymphocyte leukemia. *Proc Natl Acad Sci U S A* **105**, 16308–16313.

- Zhang SQ, Tsiaras WG, Araki T, Wen G, Minichiello L, Klein R, and Neel BG (2002) Receptor-specific regulation of phosphatidylinositol 3'-kinase activation by the protein tyrosine phosphatase Shp2. *Mol Cell Biol* **22**, 4062–4072.
- Zimmermann S and Moelling K (1999) Phosphorylation and regulation of Raf by Akt (protein kinase B). *Science* **286**, 1741–1744.

Cited Webpages

- [1] *BioCarta*. URL: <http://www.biocarta.com>.
- [2] *Cell Signaling Technology*. URL: <http://www.cellsignal.com>.
- [3] *CellNetAnalyzer-related Tools and Packages*. URL: <http://www2.mpi-magdeburg.mpg.de/projects/cna/etcdownloads.html>.
- [4] *Chemical Reaction Network Toolbox*. URL: <http://www.crnt.osu.edu/CRNTWin>.
- [5] *GUROBI*. URL: <http://www.gurobi.com>.
- [6] *HepatoSys*. URL: <http://www.hepatosys.de>.
- [7] *IBM Academic Initiative*. URL: <http://www-03.ibm.com/ibm/university/academic/pub/page/membership>.
- [8] *The Virtual Liver Network*. URL: <http://network.virtual-liver.de/en/>.