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

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 Informationstechnik 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 Jérémy 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-advice 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

Li	List of Figures			vii	
Li	List of Tables				
Zı	Zusammenfassung				xiii
1	Intr	oducti	on		1
2	Sys	tems E	Biology o	f Signal Transduction Networks	5
	2.1	Cell S	bignaling		5
	2.2	Liver	Regenera	tion	6
	2.3	Signa	ling throu	igh the EGF Ligand Family and through HGF \ldots .	6
		2.3.1	The Sig	naling System of the EGF Ligand Family	6
		2.3.2	Hepato	cyte Growth Factor Signaling	10
	2.4	Mode	ling App	roaches for Signaling Networks	11
		2.4.1	Interact	ion 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	Ordinai	y 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	Dat	a-Drive	en Interro	ogation and Training of Signaling Network Structures	5	
	Based on Interaction Graphs 33					
3.1 Data Analysis Based on the Dependency Matrix					34	
	3.2	Detect	tion and Removal of Inconsistencies between Data and Signaling			
		Network Topologies Using the Concept of Sign Consistency				
		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	Appli	cation to	ERBB Signaling	48	
		3.3.1	The Inte	eraction Graph Model of ERBB Signaling	49	
		3.3.2	Analysis	s 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	Discu	ssion		74	
4	Log	ical Mo	dels of S	Signal Transduction Networks	79	
	4.1	Specie	es Equival	lence Classes	79	
	4.2	Comp	utation o	f Intervention Strategies and Failure Modes	81	
		4.2.1	Definitio	on of Minimal Intervention Sets and General Algorithm	81	
		4.2.2	Method	s 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	Analy	sis of Exp	perimental Data	90	
	4.4	The L	ogic of El	RBB Signaling	91	
		4.4.1	From a s	Stoichiometric Model to a Logical Model	92	
		4.4.2	Theoret	ical 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 Ar	alysis	104
	4.5	Discus	ssion		110
5	From	n Stru	cture to	Dynamics: Combined Interaction Graph and OD	E
	Мос	leling	Unravels	Network Structure of HGF Signaling	115
	5.1	An In	teraction	Graph Model of HGF Signaling	117
	5.2	Exper	imental E	Data	118
	5.3	Select	ion of Mi	nimal Model Structures	121
	5.4	Transl	ation into	Ordinary Differential Equation Models	128
	5.5	Ordin	ary Diffe	rential Equation Model Selection	130
	5.6	Exper	imental V	alidation and Model Predictions	134
		5.6.1	Experim	nental Validation of a Negative Crosstalk	134
		5.6.2	Inhibito	r Combination: Model Predictions and Experimental	
			Validatio	on	136
	5.7	Discus	ssion		139
6	Pre	diction	of Quali	tative Dynamics from Network Structure	141
	6.1	Pertur	bation A	nalysis for General ODE Systems	142
		6.1.1	General	Framework	142
		6.1.2	Determi	nants and Graphs	146
		6.1.3	A Graph	n-Theoretic Criterion for Initial Response	149
		6.1.4	Determi	nant Criteria for Ultimate System Response	149
		6.1.5	Some G	raph-Theoretic Implications	153
	6.2	Pertur	bation A	nalysis for Chemical Reaction Networks	155
		6.2.1	Dynami	c Chemical Reaction Networks	156
		6.2.2	Determi	nant Criteria for Ultimate Response of Chemical	
			Reactior	Networks	160
		6.2.3	Graphic	al Representation of Dynamic Chemical Reaction	
			Networl	ks	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	Discus	ssion	191			
7	Con	clusion	15	195			
A	Integer Linear Programming Formulations for Interaction Graph Train-						
	ing			197			
	A.1	Implei	mentation	197			
	A.2	ILP Fc	prmulation of Sign Consistency	197			
	A.3	ILP Fc	prmulation of SCEN_FIT	199			
	A.4	ILP Fc	prmulation of MCoS	200			
	A.5	ILP Fc	prmulation of OPT_SUBGRAPH	201			
В	Mod	lel Con	pression for Interaction Graphs	203			
с	Doc	umenta	ation of ERBB Models	205			
	C.1	List of	Species	205			
	C.2	List of	Interactions of the Logical ERBB Model (M1)	211			
	C.3	Intera	ctions that are only included in ERBB interaction graph model .	227			
	C.4	Descri	ption of Model M2	228			
D	Sen	sitiviti	es of Binarization of HepG2 Data to Chosen Parameters	229			
Е	Doc	umenta	ation 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	Doc	umenta	ation of HGF ODE Models	239			
	F.1	Reacti	ons in the ODE Models	239			
		F.1.1	Reactions in the Core Model	239			
		F.1.2	Candidate mechanisms	241			
	F.2	Param	eter Names and Values of the Final Model Structure	242			
G	Proc	ofs from	n Chapter 6	245			
Bi	Bibliography 249						

List of Figures

1.1	Different modeling formalisms for cellular signaling	4
2.1 2.2 2.3	Interaction graph of the EGF/NRG1 network example Dependency matrix of the EGF/NRG1 example model Logical model of the EGF/NRG1 network example	9 16 20
3.1 3.2 3.3	A simple example network used for illustration purposes. Interaction Graph Model of ERBB Signaling Data set of TGFα-stimulated primary human hepatocytes and HepG2 cells	37 50 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	
3.7	and ERBB model predictions derived from the dependency matrix Comparison between experimental data from HepG2 cells and ERBB	57
	model predictions derived from the dependency matrix	58
3.8	Interaction graph model of the ERBB signaling network, processed	
3.9	for ILP approach	64
	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	
	РкоМоТ	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 $\ensuremath{M2}$.	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 $\ldots \ldots \ldots \ldots$	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 Stuktur 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 Komplexitiät es ermöglicht, wichtige Erkenntnisse über die Struktur und Funktion von Signaltransduktionswegen zu erhalten. Als Anwendungsbeispiele wurden Modelle der Signalwege des Epidermalen Wachtumsfaktors (EGF) und des Hepatozyten-Wachstumsfaktors (HGF) erstellt und analysiert, was neue Einblicke in diese Singalwege 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 verschiendener 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 widergespiegelt 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.