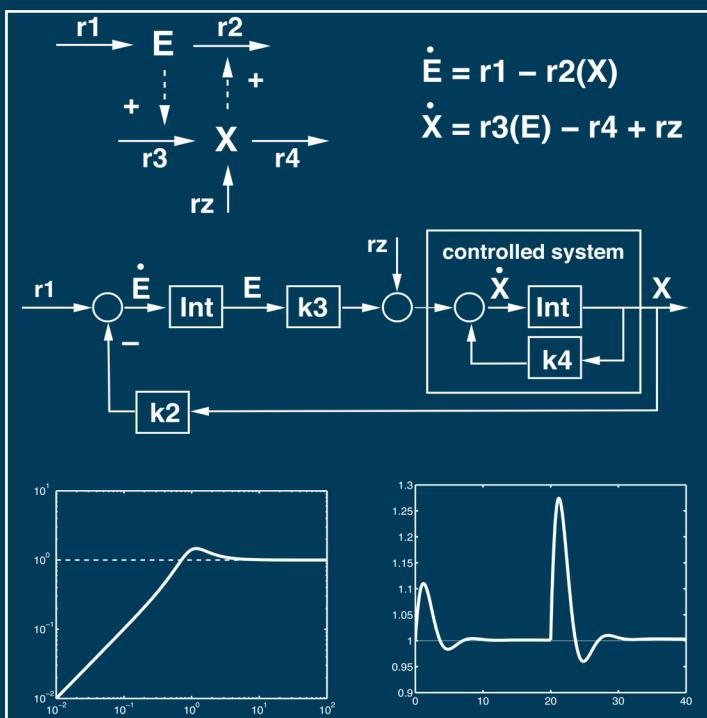


SYSTEMS BIOLOGY

Mathematical Modeling and Model Analysis



Andreas Kremling



CRC Press
Taylor & Francis Group

A CHAPMAN & HALL BOOK

SYSTEMS BIOLOGY

Mathematical Modeling
and Model Analysis

CHAPMAN & HALL/CRC

Mathematical and Computational Biology Series

Aims and scope:

This series aims to capture new developments and summarize what is known over the entire spectrum of mathematical and computational biology and medicine. It seeks to encourage the integration of mathematical, statistical, and computational methods into biology by publishing a broad range of textbooks, reference works, and handbooks. The titles included in the series are meant to appeal to students, researchers, and professionals in the mathematical, statistical and computational sciences, fundamental biology and bioengineering, as well as interdisciplinary researchers involved in the field. The inclusion of concrete examples and applications, and programming techniques and examples, is highly encouraged.

Series Editors

N. F. Britton
Department of Mathematical Sciences
University of Bath

Xihong Lin
Department of Biostatistics
Harvard University

Hershel M. Safer
School of Computer Science
Tel Aviv University

Maria Victoria Schneider
European Bioinformatics Institute

Mona Singh
Department of Computer Science
Princeton University

Anna Tramontano
Department of Biochemical Sciences
University of Rome La Sapienza

Proposals for the series should be submitted to one of the series editors above or directly to:
CRC Press, Taylor & Francis Group
3 Park Square, Milton Park
Abingdon, Oxfordshire OX14 4RN
UK

Published Titles

- Algorithms in Bioinformatics: A Practical Introduction**
Wing-Kin Sung
- Bioinformatics: A Practical Approach**
Shui Qing Ye
- Biological Computation**
Ehud Lamm and Ron Unger
- Biological Sequence Analysis Using the SeqAn C++ Library**
Andreas Gogol-Döring and Knut Reinert
- Cancer Modelling and Simulation**
Luigi Preziosi
- Cancer Systems Biology**
Edwin Wang
- Cell Mechanics: From Single Scale-Based Models to Multiscale Modeling**
Arnaud Chauvière, Luigi Preziosi, and Claude Verdier
- Cellular Potts Models: Multiscale Extensions and Biological Applications**
Marco Scianna and Luigi Preziosi
- Clustering in Bioinformatics and Drug Discovery**
John D. MacCuish and Norah E. MacCuish
- Combinatorial Pattern Matching Algorithms in Computational Biology Using Perl and R**
Gabriel Valiente
- Computational Biology: A Statistical Mechanics Perspective**
Ralf Blossey
- Computational Hydrodynamics of Capsules and Biological Cells**
C. Pozrikidis
- Computational Neuroscience: A Comprehensive Approach**
Jianfeng Feng
- Computational Systems Biology of Cancer**
Emmanuel Barillot, Laurence Calzone, Philippe Hupé, Jean-Philippe Vert, and Andrei Zinovyev
- Data Analysis Tools for DNA Microarrays**
Sorin Draghici
- Differential Equations and Mathematical Biology, Second Edition**
D.S. Jones, M.J. Plank, and B.D. Sleeman
- Dynamics of Biological Systems**
Michael Small
- Engineering Genetic Circuits**
Chris J. Myers
- Exactly Solvable Models of Biological Invasion**
Sergei V. Petrovskii and Bai-Lian Li
- Game-Theoretical Models in Biology**
Mark Broom and Jan Rychtář
- Gene Expression Studies Using Affymetrix Microarrays**
Hinrich Göhlmann and Willem Talloen
- Genome Annotation**
Jung Soh, Paul M.K. Gordon, and Christoph W. Sensen
- Glycome Informatics: Methods and Applications**
Kiyoko F. Aoki-Kinoshita
- Handbook of Hidden Markov Models in Bioinformatics**
Martin Gollery
- Introduction to Bioinformatics**
Anna Tramontano
- Introduction to Biological Networks**
Alpan Raval and Animesh Ray
- Introduction to Bio-Ontologies**
Peter N. Robinson and Sebastian Bauer
- Introduction to Computational Proteomics**
Golan Yona
- Introduction to Proteins: Structure, Function, and Motion**
Amit Kessel and Nir Ben-Tal
- An Introduction to Systems Biology: Design Principles of Biological Circuits**
Uri Alon

Published Titles (continued)

Kinetic Modelling in Systems Biology

Oleg Demin and Igor Goryanin

Knowledge Discovery in Proteomics

Igor Jurisica and Dennis Wigle

Meta-analysis and Combining

Information in Genetics and Genomics

Rudy Guerra and Darlene R. Goldstein

Methods in Medical Informatics:

Fundamentals of Healthcare

Programming in Perl, Python, and Ruby

Jules J. Berman

Modeling and Simulation of Capsules and Biological Cells

C. Pozrikidis

Niche Modeling: Predictions from Statistical Distributions

David Stockwell

Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems

Qiang Cui and Ivet Bahar

Optimal Control Applied to Biological Models

Suzanne Lenhart and John T. Workman

Pattern Discovery in Bioinformatics: Theory & Algorithms

Laxmi Parida

Python for Bioinformatics

Sebastian Bassi

Quantitative Biology: From Molecular to Cellular Systems

Sebastian Bassi

Spatial Ecology

Stephen Cantrell, Chris Cosner, and Shigui Ruan

Spatiotemporal Patterns in Ecology and Epidemiology: Theory, Models, and Simulation

Horst Malchow, Sergei V. Petrovskii, and Ezio Venturino

Statistics and Data Analysis for Microarrays Using R and Bioconductor, Second Edition

Sorin Drăghici

Stochastic Modelling for Systems Biology, Second Edition

Darren J. Wilkinson

Structural Bioinformatics: An Algorithmic Approach

Forbes J. Burkowski

Systems Biology: Mathematical Modeling and Model Analysis

Andreas Kremling

The Ten Most Wanted Solutions in Protein Bioinformatics

Anna Tramontano

SYSTEMS BIOLOGY

Mathematical Modeling
and Model Analysis

Andreas Kremling
Technische Universität München
Germany



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

Originally published in the German language by Vieweg + Teubner, 65189 Wiesbaden, Germany as Kremling, Andreas: Kompendium Systembiologie, Mathematische Modellierung und Modellanalyse, ©Vieweg+Teubner-Springer Fachmedien Wiesbaden GmbH 2012, Springer Fachmedien is part of Springer Science + Business Media.

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2014 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Version Date: 20130910

International Standard Book Number-13: 978-1-4665-6790-0 (eBook - PDF)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
<http://www.taylorandfrancis.com>

and the CRC Press Web site at
<http://www.crcpress.com>

Contents

Preface	xiii
About the Author	xv
I Fundamentals	1
1 Introduction	3
Bibliography	9
2 Biological Basics	11
2.1 The Cell — an Introduction	11
2.2 Cell Division and Growth	12
2.3 Basics of Metabolism	14
2.3.1 Energy generation	14
2.3.2 Metabolic pathways	16
2.3.3 Enzymes	17
2.3.4 Degradation of carbohydrates	17
2.3.5 Respiratory chain	19
2.3.6 Mass transport into the cell	19
2.4 Replication, Transcription and Translation	22
2.4.1 Regulation of enzyme activity and enzyme synthesis .	24
2.4.2 Signal transduction	27
2.4.3 Cell-cell communication	29
3 Fundamentals of Mathematical Modeling	31
3.1 Definition — Overview of Different Model Classes	31
3.1.1 System	31
3.1.2 Model	32
3.1.3 Process of modeling	34
3.2 Basics of Reaction Engineering	36
3.3 Stochastic Description	40
3.3.1 Stochastic simulation using the Gillespie algorithm .	45
3.4 Deterministic Modeling	50
3.4.1 Relationship between specific growth rate μ and rates r_{ai}^0	54
3.4.2 Intracellular reaction networks	56

3.5 Qualitative Modeling and Analysis	59
3.5.1 Cycles, loops and elementary modes	62
3.6 Modeling on the Level of Single Cells — the Population Balance	63
3.7 Data Driven Modeling	67
3.8 Thermodynamics	68
3.8.1 Fundamentals	68
3.8.2 Relation between ΔG and flux distributions	72
Exercises	78
Bibliography	85
4 Model Calibration and Experimental Design	87
4.1 Regression	87
4.1.1 Least square method & maximum likelihood estimation	87
4.1.2 Linear models	89
4.2 Model and Parameter Accuracy	90
4.2.1 Nonlinear models	96
4.2.1.1 Linearization	97
4.2.1.2 Bootstrapping	97
4.2.2 Experimental design	99
4.3 Dynamic Systems	101
4.4 Identifiability of Dynamic Systems	103
Exercises	106
Bibliography	107
II Modeling of Cellular Processes	109
5 Enzymatic Conversion	111
5.1 Fundamentals of Enzyme Kinetics	111
5.2 Models for Allosteric Enzymes	115
5.3 Influence of Effectors	119
5.3.1 Competitive inhibition	119
5.3.2 Essential activation	120
5.4 The Hill Equation	122
5.5 Multi-substrate Kinetics	124
5.6 Transport Processes	127
5.7 The Wegscheider Condition	128
5.8 Alternative Kinetic Approaches	130
5.9 Thermodynamic of a Single Reaction	132
Exercises	134
Bibliography	137
6 Polymerization Processes	139
6.1 Macroscopic View	139
6.2 Microscopic View	140
6.2.1 Calculation of the transcription rate	141
6.2.2 Calculation of the translation rate	143

6.2.3 Combination of equations	145
6.3 Influence of Regulatory Proteins (Transcription Factors, Repressors)	146
6.3.1 Induction model	146
6.3.2 Repression model	148
6.3.3 Extended induction model	148
6.4 Interaction of Several Regulators	151
6.4.1 AND – OR gates	151
6.4.2 Hierarchical approach	152
6.5 Replication	156
Exercises	160
Bibliography	161
7 Signal Transduction and Genetically Regulated Systems	163
7.1 Simple Schemes of Signal Transduction	163
7.2 Oscillating Systems	171
7.3 Genetically Regulated Networks	173
7.3.1 Inputs/outputs	173
7.3.2 Negative and positive feedback reactions	174
7.3.3 Toggle switch	177
7.4 Spatial Gradients by Signal Transduction	180
7.5 Analysis of Signaling Pathways by Heinrich	183
Exercises	187
Bibliography	188
III Analysis of Modules and Motifs	189
8 General Methods of Model Analysis	191
8.1 Analysis of Time Hierarchies	191
8.1.1 Linear systems	192
8.1.2 Classification of fast and slow reactions	197
8.1.3 Complete procedure for model reduction	199
8.1.4 Singular perturbation theory	201
8.2 Sensitivity Analysis	204
8.2.1 Definition of the parametric sensitivity w_{ij}	205
8.2.2 Sensitivity analysis using Hearne's method	207
8.2.3 Sensitivity matrix	209
8.2.4 Sensitivities for dynamic systems	210
8.3 Robustness in Stoichiometric Networks	211
8.4 Metabolic Control Analysis	216
8.4.1 Control coefficients	216
8.4.2 Summation and connectivity theorems for linear pathways	218
8.4.3 General summation and connectivity theorems	220
8.5 Biochemical Systems Theory	222

8.5.1	Building a theoretical framework	223
8.5.2	Setting up a power-law model	225
8.5.3	S-systems	227
8.5.4	Design principles and end-product inhibition	232
8.5.5	Design space analysis	233
8.6	Structured Kinetic Modeling	238
8.7	Model Reduction for Signal Proteins	242
Exercises	245
Bibliography	251
9	Aspects of Control Theory	253
9.1	Observability	253
9.2	Monotone Systems	258
9.2.1	Monotony of networks	258
9.2.2	Multiple stationary states	262
9.2.3	From nonmonotone systems to monotone systems	264
9.3	Integral Feedback	267
9.3.1	Bacterial chemotaxis	270
9.4	Robust Control	273
Exercises	278
Bibliography	279
10	Motifs in Cellular Networks	281
10.1	Feed-forward loop (FFL)	283
10.1.1	Logic switching pattern	286
10.1.2	Dynamic behavior	286
10.2	Feed-forward Loops in Metabolic Networks	287
10.2.1	Carbohydrate uptake and control	288
10.3	FFL in Signaling Systems: Two-component Signal Transduction	293
10.4	Further Signaling Motifs	296
10.4.1	Quorum sensing system	296
10.4.2	Transfer of phosphate groups in a cascade	298
Exercises	300
Bibliography	300
IV	Analysis of Cellular Networks	303
11	Metabolic Engineering	305
11.1	Reconstruction of Metabolic Network	305
11.2	Tasks and Problem Definition	306
11.3	Subspaces of Matrix N	309
11.3.1	Singular value decomposition of matrix N	312
11.4	Methods to Determine Flux Distributions	314
11.4.1	Stationary flux distribution	314
11.4.2	Redundant metabolic network	319

11.4.3 Optimality criteria	321
11.4.4 Linear optimization	323
11.5 Strain Optimization	325
Exercises	329
Bibliography	331
12 Topological Characteristics	333
12.1 Network Measures	333
12.1.1 Connectivity and cluster measures	333
12.2 Topological Overlap	336
12.3 Formation of Scale Free Networks	337
Bibliography	338
Appendix A Collection of mathematical approaches	343
Index	361

This page intentionally left blank

Preface

Systems biology is more and more perceived as an independent research field that is characterized by highly interdisciplinary activities. As in other fields, this is reflected by the emergence of textbooks to support teaching in newly designed study courses. This textbook contributes to theoretical lectures given in the field.

When I started teaching about ten years ago, one central question was how to structure all the material I collected for the students. A thorough analysis of the textbooks available on the market showed that a book fulfilling all the criteria and providing all the material that I thought it should was not available. So, in the end I decided to provide my own collection of material as a printed text for the students. This was in 2007. Meanwhile, the number of publications and textbooks has exploded and a broad spectrum of ideas is available. Therefore, now it is indispensable to extract selected methods from this huge “reservoir” and to present it in a compact form to the students, supported with exercises and numerical simulation studies. The scope of this book is not to present all the methods that cover all aspects of theoretical systems biology—there are definitely more extensive textbooks on the market—but to present selected methods in more detail, especially in the way mathematical formulas are derived. Moreover, all numerical studies that are shown in various figures can be downloaded from the publisher’s web site, and a solutions manual is available from the publisher for qualifying instructors. The book can be used for self-consistent studies and at best as supporting material for a lecture in theoretical systems biology. The book is dedicated to students from engineering sciences and mathematics; however, I believe that students from life science or microbiology can also use the book *after* an intense introduction in basic engineering mathematics. Such courses are now offered at many universities to facilitate interdisciplinary studies.

The concept of the book can be summarized as follows: process - function - behavior.

As a control engineer I was always interested in the design of processes and possibilities to structure the knowledge that we have on the process. Accordingly, the first part of the book is dedicated to possibilities to describe cellular systems, and the focus is on three major approaches — a stochastic approach, a quantitative deterministic approach and a qualitative deterministic approach — which is reflected in the structure of equations systems (also different, depending on the approach: a chemical master equation, a differen-

tial equation or a graph). In Part II, in order to complete system description, kinetics are assigned to the reaction rates that characterize the temporal processes in the system. In Parts III and IV, the understanding of the system by different approaches of model analysis, that is, to infer the function of smaller networks and to observe the behavior of the cellular population is the main focus.

As an extension to the German version of the book (Vieweg & Teubner, 2011) additional chapters were included. Especially, I wish to thank my co-workers Katharina Pflüger-Grau and Alberto Marin Sanguino, who contributed with their own parts (Chapter 2 [KPF] and Section 8.5 [AMS]). I also wish to thank the many people who contributed to the preparation of the English version of this book: my co-workers at the Technische Universität München for reading and correcting the manuscript, M. Zaki, L. Tran, and H. Rosner for preparing the material, and B. Auerbach and M. Valderrama who supported the preparation of exercises and numerical studies.

München, July 2013
A. Kremling

About the Author

Andreas Kremling studied engineering cybernetics at University Stuttgart, Germany, and started his PhD project at the Institute of Systems Dynamics and Control (Prof. Gilles). He became a member of the newly founded Max Planck Institute for Dynamics of Complex Technical Systems in Magdeburg, Germany, where he was a leader of several research projects dedicated to systems biology. Here, he also started to give lectures in the research field for engineers and life science students. After completing his PhD at University Stuttgart and his habilitation at University Magdeburg he was appointed as a Professor for Systems Biotechnology at the Technische Universität München in 2010.

This page intentionally left blank

Part I

Fundamentals

This page intentionally left blank

Chapter 1

Introduction

Systems biology has shown a strong emergence in the last few years. This is visible in forms of various novel research programs, the founding of new departments as well as offers of new lectures at universities. However, there are numerous approaches in experimental and theoretical systems biology, so that a current clear definition of the “systems biologist” does not exist. Research in systems biology is based on an interdisciplinary approach. This implicates that theoretical, experimental and computational methods are applied. Systems biology as used in this compendium focuses on mathematical modeling. There are many reasons for this: Biological systems (here in particular cellular systems) are based on interactions between various components. The relations between the components are not linear, as for example in a linear chain of biochemical reactions, but are characterized by feedback and feed-forward loops and their behavior is not intuitively understandable. Mathematical modeling is used in the first step to encompass existing knowledge on the system under investigation. In the next step, models are analyzed to acquire a better understanding. There are various methods available based on a deterministic explanation, and some of them are covered in this book. Furthermore, these models are applied systematically, e.g., in biotechnology to use biological resources efficiently to increase quality and quantity of an interesting product. Therefore, it is required that the models are predictive, that is, it should be possible to formulate hypotheses that are testable first in simulations and later on in the real experiment.

Two different but complementary approaches that lead to a mathematical model. In the bottom-up approach a small subsystem is chosen as starting point. For this system biological knowledge from literature is collected and used for a first model. Figure 1.1 left shows the steps that are passed through to come to a meaningful description of the system. Then the model can be combined with further submodels to describe a larger unit.

In the experimental oriented top-down approach an overall picture of the cellular activity is available, for example, in form of cDNA-array-data (Figure 1.1 right side). The data are analyzed in the first step, and subsequently combined/integrated with other data, proteome data or metabolome data. Mathematical models can then be built on these data based on various techniques; very common are statistical approaches like clustering, factorization techniques and factor analysis and stochastical approaches based on Bayesian networks.

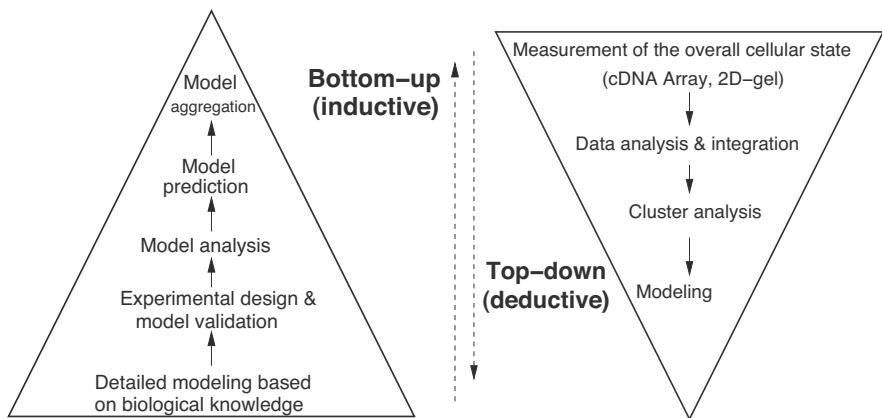


FIGURE 1.1: Two approaches in the systems biology. Left: Bottom-up approach. Right: Top-down approach. In systems biology, both approaches complement each other.

Mathematical modeling goes hand in hand with molecular biological experiments, as Figure 1.2 shows. Starting point is a biological experiment, an observation or an unexplainable phenomenon. For using a mathematical modeling approach it is required that the system at hand and the problem can be described mathematically. During problem definition, the researcher may make hypotheses on solutions of the problem. In the modeling process, it is important that the model explains, either qualitatively or quantitatively, the presented data. This is possible by predicting future behavior if different input conditions are chosen in follow-up experiments. Consequently, the model can also be used to make proposals for further experiments that can be tested and help to improve the model (experimental design). The problem is processed in an iterative sequence of model improvements and new experiments.

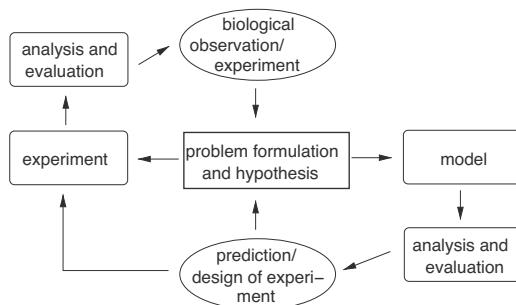


FIGURE 1.2: Iterative cycle of problem definition, biological experiment and modeling.

Despite the fact that the presented methods are applicable very generally, most of the examples originate from the fields of microbiology. The bacterium *Escherichia coli* is chosen as a “model” organism, which is one of the most researched microorganism today. Thus, biological basics will shortly be covered in order to apply presented methods specifically on metabolism, signaling, gene expression and control. Then, basics in mathematical modeling will be discussed, considering different explanation methods. The focus lies on the deduction of deterministic models that will be used in analysis and verification. By balancing mass flow and energy flow a system of differential equations is achieved as a basic structure of a deterministic approach. Stochastic modeling is applied when the number of reactants is very small; this is the case for an interaction of a regulator protein with its respective DNA binding sites whereas both partners, the protein and the binding sequence of the protein, are only available in small copy numbers. The counterpart of the mass balance equation is the (chemical) master equation, which indicates the probability of one component in the cell having a specific number of molecules. Deterministic models can also be discrete and describe the system qualitatively. Graphs are very popular entities that can be used here to describe and analyze also larger biochemical networks. Also thermodynamic aspects are described in this part of the book.

In a subsequent chapter, the kinetic parameters of deterministic models are in the focus. This is due to the fact that a quantitative description of experimental data using mathematical models requires an estimation of the kinetic parameter that is in general unknown. However, the focus here is not on optimization methods, as these are covered in special lectures and other books (see a complete list of recommended books at the end of this chapter). More important is the assessment of the estimated parameter and models as well as improvement of the parameter accuracy. To start with, methods of linear regression are covered, and differences between linear and nonlinear regression are discussed. Then input variables are determined to improve estimation accuracy during experimental design.

In the second part intracellular processes are discussed. Here, first emphasis is on enzymatic reactions. Various mechanisms are used to describe the transformation of substrates in products. Additionally, there are regulatory influences caused by inhibitors or activators that affect the reaction rate. Simple network structures that have a certain signal response behavior are analyzed in signal transduction systems. Examples for more complex networks are two component signal transduction, phosphotransferase system and phosphate transfer cascades, which will be covered in a chapter on motifs. Spatial gradients are considered, which emerge when phosphorylation of a component on an membrane occurs, with dephosphorylation occurring inside of the cell. This leads to a different type of differential equations, partial differential equation. This type describes the alterations of the state variables over time and along space coordinates.

Polymerization processes are another important class of reaction. Inter-

action between regulatory proteins and specific DNA-binding sites plays an important role during initiation of mRNA synthesis. The complete synthesis of macromolecules can also be modeled in detail and compared to measured data. Detailed models of polymerization can be deduced from macroscopical observations. Furthermore, basic regulatory schemes known in bacteria like induction and repression are discussed. During replication in bacterial cells, DNA synthesis is observed, even though the current synthesis was not yet completed. As a result, the cells have different copy numbers of the genes depending on the location on the DNA. Therefore, a few relations are given to identify copy numbers dependent on growth rate and location on the DNA.

The concept of this book is to show that cooperation of individual processes lead to a certain functionality of the subnetwork inside the cell, which results in a certain behavior of the entire cell population. The sequence process–function–behavior is an important structural element of chapter divisions. Modeling and analysis of signal transduction units play a mediating role between process and function. A simple connection of a few elements can already cause functionality. If such connections occur more frequently than others it can be assumed that the structure is an important one. Such connection patterns are known as motives.

Cellular systems are also characterized by the fact that parts of the biochemical reaction network can be separated and described independently. These parts are known as modules and can be defined according to different criteria. On one hand it is possible to classify them for example by their metabolic functionality such as providing precursors for other parts of the network; on the other hand, classification by the organization of genes in operons or regulons is also possible. Lastly, components and reactions can be structured by signal transfer and processing, so that elements controlled by a common signal transduction system are encompassed in one unit. These ideas are discussed in a very interesting and recommendable publication [1].

A summary of modules and motives is demonstrated in Figure 1.3. Glycolysis is shown in part **A**, which encompasses an almost linear chain of reactions that break a hexose in a triose. Glycolysis can be called a module, as it functions to exploit ATP. Glycolysis is almost the only ATP producing unit if the organisms grow under anaerobic conditions. A part of the network shown in part **B** in detail is a motif. The motif shows the activation of the pyruvate kinase reaction by the metabolite fructose-1,6-bisphosphate as known from *E. coli*. The forward loop has a certain function which will be discussed in the respective chapter. Since this part of glycolysis is connected to a signaling pathway to activate transcription factor Crp via the phosphotransferase system an interesting example for the connection of metabolism and signal processing is presented. In the upper part of the figure the uptake systems for lactose **C** and arabinose **D** are shown. They build a module as the corresponding genes form an operon for the degradation of lactose (genes *lacZYA*) and arabinose (genes *araFGH*). The lactose uptake system has a certain function. It is known that if an inducer causing gene expression is provided a bistable behavior can be

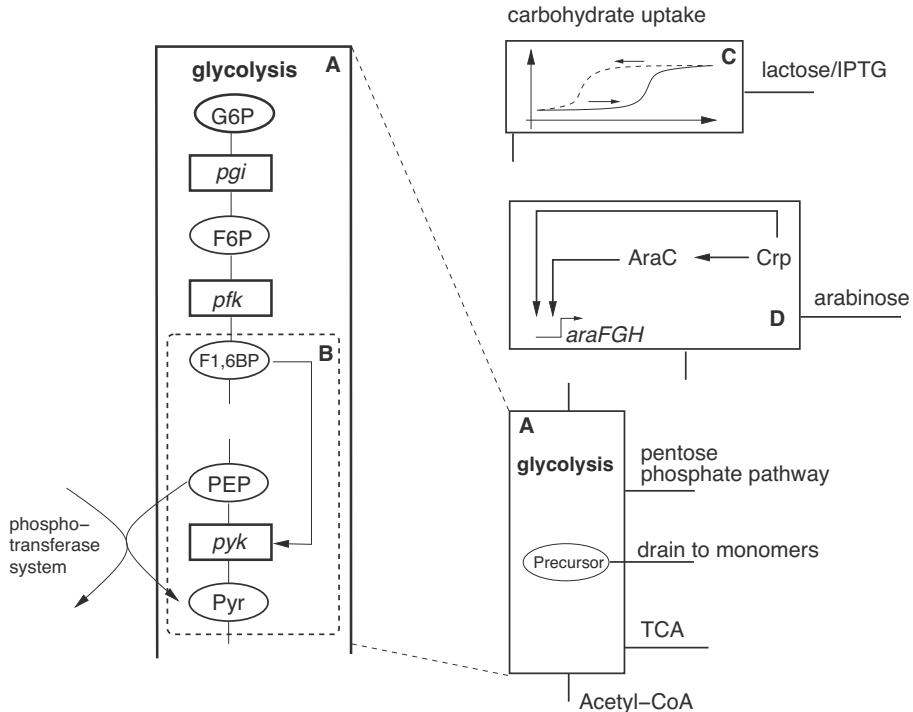


FIGURE 1.3: Representation of modules and motifs with the example of carbohydrate uptake in *E. coli*.

observed when the concentration of the inducer is varied. Increasing the concentration of the inducer results in enhanced protein synthesis after reaching a threshold value. By reducing the inducer's concentration, a switch off takes place at a different concentration, as can be seen in the figure. This functionality has the advantage of not having to switch gene expression on and off in case of fluctuations in nutrient supply. The lactose uptake system is widely used in the book to demonstrate various modeling and analysis tools. In D one can see the connection between two transcription factors affecting the *ara*-operon. It is apparent that transcription factor Crp regulates both transcription factor AraC and genes in the operon. A feed-forward loop takes place on the level of gene regulation. This motif occurs very frequently and represents one of the most important motifs in *E. coli*.

The third part covers analysis of modules and motifs by presenting theoretical methods addressing the dynamics of subsystems and the behavior near a steady state. Techniques of determining different time scales are presented. Time scales are characteristic for systems and are used for model reduction/simplification. Sensitivity analysis is another important standard method in systems biology. Determining sensitivities is an important tool in

ranking the importance of kinetic parameters and also in parameter estimation. A well known method for analysis of sensitivities is metabolic control theory. It defines coefficients that connect the local characteristics of individual reaction steps with global characteristics of the module like the steady state flux through the network. A very similar method, biochemical systems theory, also allows specific model analysis by using approximative kinetics. Structural kinetic modeling is an analysis of a system's dynamic, which will be subsequently covered. In an additional chapter a method is introduced that allows to determine if some of the state variables show a robust behavior in the sense that their steady state concentrations are not affected if the system is perturbed (absolute concentration robustness). When signal proteins and effectors interact, a large number of different conformations can result. If only the physiological units are relevant, one can reduce the number of conformations by using reduction techniques.

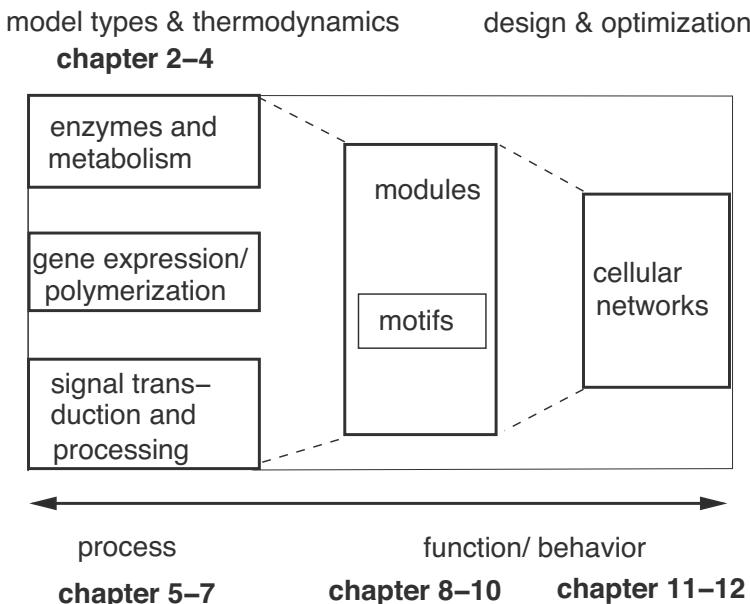


FIGURE 1.4: Book organization.

Theoretical control engineering aspects have received much more attention recently. For example in bacterial chemotaxis it was shown that feedback loops are important for the understanding of the experimental observations. The feedback shows a typical behavior that is known as integral feedback in control theory and will be examined in detail. Biochemical networks are characterized by the fact that there is insufficient quantitative information for many parts to parameterize a dynamic model. Consequently, the class of monotone systems represents a connector between both fields — quantitative

and qualitative methods. Monotony of systems can be tested with graphical methods. If the subsystems demonstrate a monotone behavior, conclusion for the dynamic behavior of the entire system can be drawn without detailed qualitative knowledge. At the end of this part a method for robust control is presented.

In the fourth part, networks are discussed while considering two aspects. Analysis of stoichiometric networks is a standard method for biotechnologists. These methods allow the determination of flux maps in the cell, that is, the distribution of incoming fluxes through the network. Networks can also be analyzed graphically. In this case, questions on connection frequency, shortest connection between two components and cluster characteristics can be answered. Current research in systems biology provides a large number of new theoretical methods for modeling and analysis of cellular systems. So, it is impossible to describe and summarize all recent findings. Hence, a selection is presented in this book where we feel that important aspects are covered. Figure 1.4 summarizes the structure of the book.

Bibliography

- [1] D. M. Wolf and A. P. Arkin. Motifs, modules and games in bacteria. *Current Opinion in Microbiology* 6: 125-134, 2003.
-

Further Readings

The following books represent important literature in systems biology.

- Alberts, R. A. Thermodynamics of Biochemical Reactions; John Wiley, & Sons, Inc, 2003.
Comment: Emphasize thermodynamic topics in biochemical reactions.
- Alon, U. An Introduction to Systems Biology – Design Principles of Biological Circuits; Chapman & Hall, 2006.
Comment: Focus in on the analysis of network motifs. Chemotaxis as interesting example is considered.
- Bailey, J. E. Biochemical Engineering Fundamentals, McGraw-Hill Professional, 1988
Comment: Classic textbook for Metabolic Engineering.

- Beard, D. and Hong, Q. *Chemical Biophysics – Quantitative Analysis of Cellular Systems*; Cambridge University Press, 2008
Comment: Focus also on thermodynamic aspects.
- Bergethon, P. R. *The Physical Basis of Biochemistry – The Foundations of Molecular Biophysics*; Springer Berlin, 1998
- Bisswanger, H. *Theorie und Methoden der Enzymkinetik*. Wiley-VCH, 1994.
- Cooper, S. *Bacterial Growth and Division*; Academic Press, 1991
- Dochain, D. [Ed.] *Automatic Control of Bioprocesses*; ISTE Ltd and John Wiley, 2008
Comment: Collection of texts for process engineers.
- Heinrich, R. and Schuster, S. *The Regulation of Cellular Systems*, Springer, Berlin, 1996
Comment: Basic introduction in many topics of systems biology with focus on static network analysis and metabolic control theory.
- Klipp, E. et al. *Systems Biology: A Textbook*, Wiley-VCH, 2009
- Neidhardt, F. *Physiology of the Bacterial Cell : A Molecular Approach*, Sinauer Associates, 1990
- Nelson, D. L. and Cox, M. M. *Lehninger Principles of Biochemistry*. Palgrave Macmillan, 2008.
Comment: Biological basics on *E. coli*.
- Nielsen, J. et al. *Bioreaction Engineering Principles*, Springer Netherlands, 2002
Comment: Emphasizes modeling and analysis of bioengineering processes.
- Palsson, Bernhard O. *Systems Biology – Properties of Reconstructed Networks*, Cambridge University Press, 2006
Comment: Focus on the analysis of stoichiometric networks.
- Palsson, Bernhard O. *Simulation of dynamic network states*, Cambridge University Press, 2011
- Segel, Irwin H. *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-state Enzyme Systems*; Wiley Classics, 1993
- Stephanopoulos, G. N. et al. *Metabolic Engineering: Principles and Methodologies*; Elsevier, 1998
Comment: Here also Metabolic Engineering is the main focus.
- Voit, Eberhard O. *Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists*; Cambridge University Press, 2000
- Voit, Eberhard O. *A First Course in Systems Biology*, Garland Science, Taylor & Francis Group, 2012

Chapter 2

Biological Basics

2.1 The Cell — an Introduction

The smallest viable unit of organic matter is a cell. Its fundamental components are identical in all organisms: DNA (deoxyribonucleic acid), RNA (ribonucleic acid), proteins, lipids and phospholipids. However, there are considerable differences between different groups of organisms regarding the microstructure and composition of each cell. Generally, it is distinguished between eukaryotic (animals, plants, algae and protozoans) and prokaryotic cells (Bacteria and Archaea). The main component, about 70-85%, of a bacterial cell is water, which means that their dry mass is about 10-30% of their fresh weight. Proteins make up about 50% of the fresh weight, the cell wall accounts for 10-20%, the entire RNA for another 10-20%, the DNA for 3-4% and approximately 10% of the fresh weight are lipids. In contrast to the eukaryotic cell, the prokaryotic cell does not possess a nucleus but the entire genome is found in a ring shaped DNA strand freely floating in the cytoplasm, which is called the bacterial chromosome.

The DNA encodes all the information necessary for cellular growth. Additionally, the cell might contain smaller, circular DNA structures, the so-called plasmids. These plasmids however are not essential but usually equip the cell with additional properties. Regarding their shape, prokaryotes are quite uniform. They can either be spherical or even curved barrels. The spherical ones are called cocci (singular: coccus), the others rods. Microorganisms propagate through cell division, where usually two identical daughter cells are formed. Prokaryotes are extremely diverse also regarding their habitats. A distinction is made between aerobic microorganisms, those that need oxygen for life, and anaerobic microorganisms, those that can only survive in the absence of oxygen.

Furthermore microorganisms are found in a great range of temperature (-15°C to 113°C), a broad range of pH (pH=0.06 to pH=12) or at different pressures (for example in the deep-sea) and at various salt concentrations (in saturated saline solutions). These extreme habitats show that microorganisms are widely spread and can be found in almost all possible environments. Microorganisms colonizing these extreme habitats are called extremophiles. Extremophilic microorganisms are of great significance in biotechnology since

they possess enzymes that are still functional under the most extreme conditions (for example high temperatures).

This broad variety of life forms is reflected by their metabolism. The metabolism describes all biochemical reactions occurring in an organism that are necessary for energy production and therefore essential for growth and cell division. In general, it is differentiated between respiration, photosynthesis and fermentation. For respiration, either oxygen (aerobic respiration) or another anorganic molecule, like SO₄²⁻ (anaerobic respiration), serves as the terminal electron acceptor. In doing so, a proton gradient is formed across the cell membrane, which is then used for energy production. In contrast, during photosynthesis light energy is used for ATP generation, which means that light energy is transformed into chemical energy. Fermentation describes a process, in which under anaerobic conditions organic compounds serve as both electron donors as well as acceptors, and energy is exclusively gained by substrate level phosphorylation.

2.2 Cell Division and Growth

Growth describes the increase of living matter, which means a gain in cell mass and cell number. The specific growth rate (μ) describes the increase of the number of bacterial cells in a distinct period of time. Prokaryotic cells usually grow by symmetric cell division, so that two identical daughter cells are formed. Prior to the cell division, the cell elongates and therefore increases its biomass and duplicates its chromosome and existing plasmids. The time that is needed for a duplication of the cell number is called generation time, whereas the period for duplication of the cell mass is called doubling time (τ). Usually generation and doubling time are identical, because the cell mass doubles at the same rate as the cell number. The doubling time can be determined from the specific growth rate μ . Setting up the following equation for describing the change in biomass B with a constant rate of growth μ :

$$\dot{B} = \mu B \quad \longrightarrow \quad B(t) = B_0 e^{\mu t}, \quad (2.1)$$

the doubling time τ can be determined from the equation:

$$2 B_0 = B_0 e^{\mu \tau}. \quad (2.2)$$

This leads to the following correlation between τ and μ :

$$\tau = \frac{\ln 2}{\mu}. \quad (2.3)$$

The specific growth rate of a microorganism depends on various factors, but should be constant under the very same conditions. Apart from temperature,

oxygen supply and medium composition have a significant influence on the growth rate. The influence of the carbon source on the growth rate, for example, can be analyzed by varying the substrates in an otherwise identical medium. By this, good carbon sources can be differentiated from less favorable C-sources.

In general, growth of a bacterial culture can be divided into several growth phases, which becomes obvious when plotting the growth against time. Therefore, the optical density (a dimension for measuring the cell number) of a static culture is plotted on the ordinate and the time on the abscissa. The result is a growth curve, which is typical for bacterial cells. Growth curves should be presented semilogarithmically, as this allows a good presentation of a high number of cell divisions and exponential curves. A typical growth curve is shown in Figure 2.1.

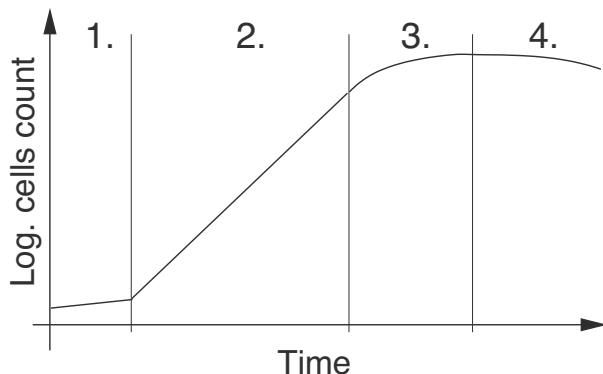


FIGURE 2.1: Typical growth curve of a bacterial culture: 1. lag phase, 2. exponential phase, 3. stationary phase and 4. death phase.

The starting period, called lag phase, is the time from the initial inoculation until the maximum division rate is reached. During this time, the bacteria adapt to the new growth conditions. After this adaptation period, the exponential growth phase begins. In this phase, the bacteria grow at maximum speed and the cell size, cell components and the material flow are usually constant. Therefore, most physiological experiments are performed during that stage. Afterwards, the stationary phase begins. Now, hardly any (net) growth happens. The transition from the exponential to the stationary phase occurs gradually and is caused by limiting factors like the decrease of the substrate concentration, high population densities, low oxygen partial pressure or the accumulation of toxic metabolites. However, cells at that stage are still metabolically active and can produce so-called secondary metabolites (e.g., penicillin). Many biotechnological applications target exactly at these secondary metabolites. In biotechnology, the first growth- or nutrition phase is therefore called trophophase and the following stage of production is called

idiophase. The stationary phase is followed by a final phase, the death phase, in which a gradual decline of the number of living cells in the culture occurs until all cells are dead. These different phases describe events in a bacterial population and therefore the terms lag-phase, exponential phase, stationary phase and death phase should only be applied to populations of cells, not to individual cells.

2.3 Basics of Metabolism

Metabolism refers to all the biochemical reactions that are essential for growth, energy supply and cell division. Nutrients are taken up from the environment and subsequently degraded by the activity of enzymes in different metabolic pathways. This results in new chemical compounds, which serve as precursors for cellular constituents and hence are used for cellular growth. This process, in which nutrients are transformed into cellular constituents, is called anabolism. The synthesis of new cellular material is referred to as biosynthesis. Biosynthetic reactions are energy dependent, which means that each single cell must have the ability to generate energy. Energy is not only needed for biosynthesis, but also for other cellular functions, like motility or nutrient transport. Energy sources are delivered, as also the nutrients, by the environment. Generally one can differentiate between two different energy sources: Light and chemical compounds. Organisms that use light as energy source make photosynthesis and are referred to as phototrophic organisms. The other group of organisms, which produces energy through chemical degradation, is called chemotrophic. All metabolic processes that lead to the production of energy are summarized in the term catabolism. The global structure of metabolism is shown in Figure 2.2.

The chemical composition of a cell is mainly made up of four components: Carbon, oxygen, hydrogen and nitrogen. These four elements are the main building blocks of all macromolecules (DNA/RNA, proteins, lipids, polysaccharides), and also of small organic molecules (amino acids, nucleotides, sugar). Apart from these, apparently further elements are found in the cell which are also essential for the metabolism but present in far lower concentrations. These cover phosphate, potassium, calcium, magnesium, sulfur, iron, zinc, manganese, copper, molybdenum, cobalt and others, depending on the organism. All these elements have to be taken up by the cell from the environment in order to be available for the synthesis of macromolecules.

2.3.1 Energy generation

During catabolism, nutrients are degraded by the metabolism in order to generate energy. If this was to happen in one single step, the cell would most

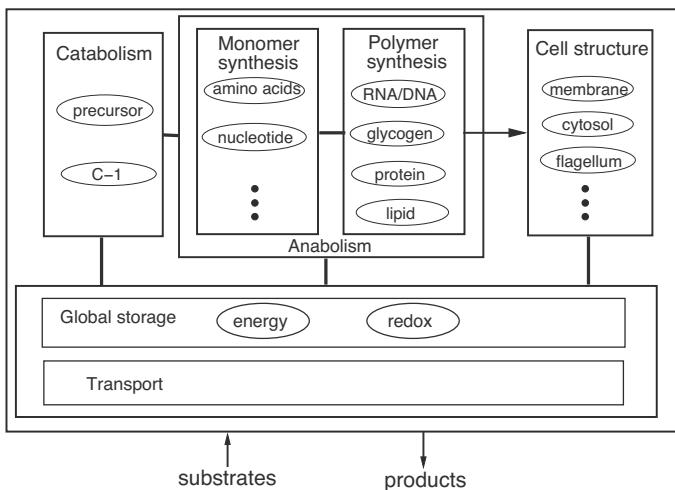


FIGURE 2.2: The metabolism is split into catabolism and anabolism. Transportation processes are needed for many components and are therefore distributed among all modules. Redox and energy equivalents are involved in reactions from all modules.

likely burn because of the heat development. Therefore, energy is transmitted in a stepwise process to a carrier substance, thereby forming adenosine triphosphate (ATP), a nucleotide. ATP is the most phosphorylated form, as shown in Figure 2.3.

Reactions that lead to the generation of ATP do not occur spontaneously and are called endergonic. The generated ATP can then be used in anabolic or transport reactions (exergonic reactions). In cells, ATP can be generated in three ways: by photosynthetic phosphorylation, respiratory chain phosphorylation or by substrate-level phosphorylation. During the first two processes, the proton gradient across the membrane is used for the generation of ATP by the activity of an enzyme called ATP synthase. The substrate-level phosphorylation occurs during some reactions in the central metabolism, in which a phosphate molecule is transferred directly to ADP by the enzyme catalyzing the reaction. An example for this is the conversion of phosphoenolpyruvate (PEP) by the Pyruvate-Kinase. Thereby ATP and pyruvate are generated:



Organisms, which gain their energy through fermentation processes, have a far lower energy yield than organisms performing respiration or photosynthesis, as ATP can only be generated through substrate-level phosphorylation.

Energy currency of the cell: ATP

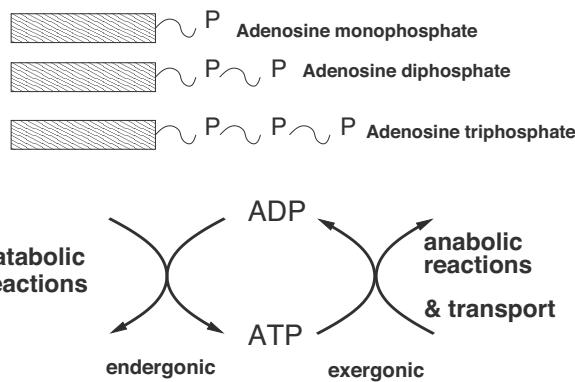


FIGURE 2.3: The energy currency of the cell: ATP.

2.3.2 Metabolic pathways

A metabolic pathway consists of metabolites and reactions that connect the metabolites. The major metabolic pathways, like the degradation of sugars or the respiratory chain, are almost identical in all organisms. However, in some groups of bacteria, the basic schemes have changed and some routes have shortened or are rudimentary, while others predominate. Metabolic pathways have special functions in the cell like, for example, the supply of energy equivalents in the glycolysis (degradation of sugar, catabolism) or the supply with precursors for the generation of monomers (anabolism). Metabolites can be connected in various ways. In Figure 2.4 several possibilities are shown.

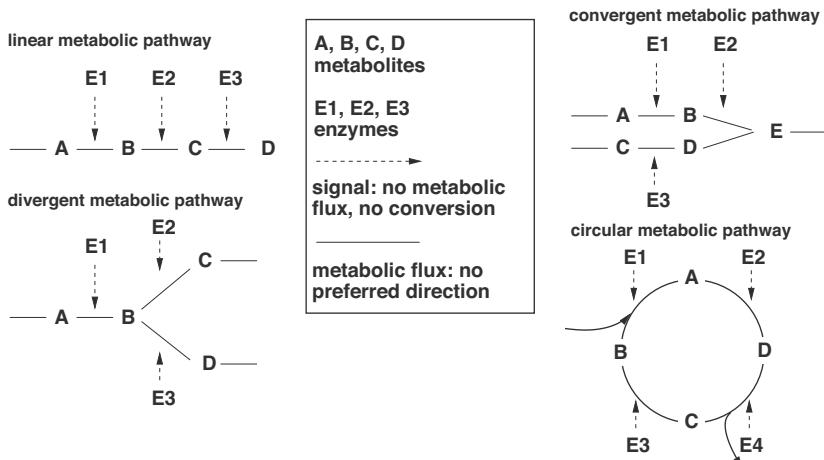


FIGURE 2.4: Different representations of metabolic pathways.

2.3.3 Enzymes

Metabolic reactions generally do not occur by themselves, but need to be catalyzed by enzymes. The enzymes themselves are not consumed during the chemical conversion and are therefore available for further reactions. Enzymes form a specific class of proteins. Generally they consist of different functional domains and catalyze one specific reaction, which means the conversion of one metabolite into another. First, the enzyme has to recognize the respective metabolite and subsequently it catalyzes the reaction. The participation of enzymes in reactions makes them controllable. However, only the velocity of the reaction can be modified, not the equilibrium. Enzymes are also called biocatalysts, as they lower the activation energy of a reaction. The enzymatic reaction is influenced by temperature and the pH. Each enzyme generally reacts only with one specific metabolite, which is referred to as the enzyme's substrate. The enzyme catalyzes the conversion of the substrate into the product of the reaction until an equilibrium is reached. Thus, each enzyme is specific for its substrate (substrate specificity). The enzymatic reaction starts with the recognition and binding of the substrate to the so-called catalytic site (a specific site) of the according enzyme. Thereby the enzyme-substrate complex is formed. After the conversion of the substrate has occurred, the product dissociates from the enzyme, which in turn is again ready for the next round of reactions. A relevant characteristic of enzymes for the functionality of metabolism is the possibility to regulate their catalytic activity. By varying the reaction velocity, an adaptation of the metabolic fluxes to changes in the environmental situation is achieved. Therefore, this leads to an optimal utilization of the available resources.

Those regulated enzymes usually have a second binding site, the so-called regulatory site. A widely spread regulatory mechanism is the end-product inhibition. There, the end-product of a metabolic pathway binds to the regulatory site of one of the enzymes of the reaction chain and thereby downregulates the velocity of the enzymatic reaction. By this the accumulation of a specific metabolite can be avoided. Those metabolites that lead to changes in the velocity of an enzymatic reaction are called effectors. If their production leads to a repression of the enzyme activity they are referred to as negative effectors or repressors and if their production leads to an increase of enzyme activity, they are called positive effectors or activators.

2.3.4 Degradation of carbohydrates

The majority of microorganisms degrade carbohydrates. Macromolecules (e.g., starch) are usually digested outside the cell by secretory enzymes. These enzymes degrade the macromolecules into monomeric building blocks (e.g., glucose), which are then taken up by transport systems into the cell and channelled into the metabolism.

A schematic representation of the general structure of the metabolism of

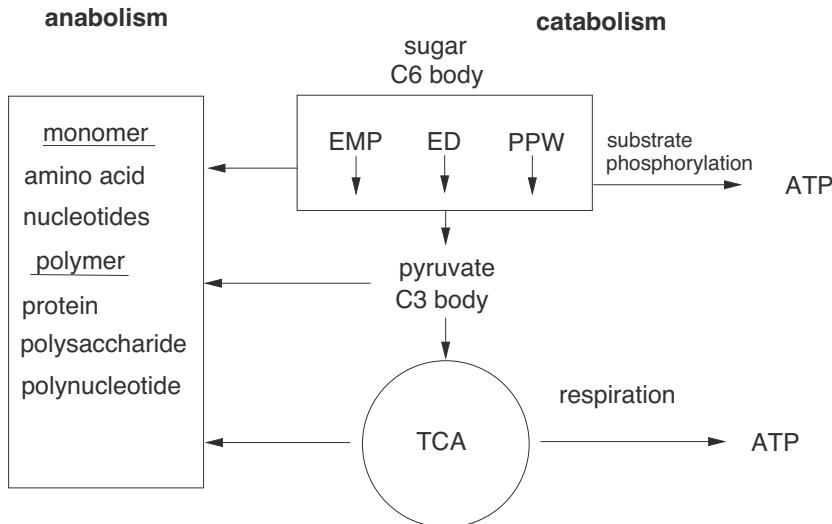


FIGURE 2.5: General structure of central metabolism. Emden-Meyerhof-Parnas pathway (EMP), also known as glycolysis. Entner-Douderoff pathway (ED), pentose phosphate cycle (PPW) and tricarboxylic acid cycle (TCA).

sugar degradation by respiratory organisms is shown in Figure 2.5. Hexoses (monosaccharides with six carbon-atoms) are taken up by the cell and degraded into pyruvate by different metabolic pathways. Pyruvate is a central metabolite in intermediary metabolism. The possible metabolic pathways from a C6-body (sugar) down to the C3-body (pyruvate) are the Emden-Meyerhof-Parnas pathway (EMP), also called glycolysis, the Entner-Douderoff pathway (ED) or the pentose phosphate cycle (PPW). All these metabolic pathways converge at the level of pyruvate. In the course of the reactions of the upper pathways, energy is produced in form of ATP through substrate level phosphorylation. Next, pyruvate is decarboxylated and the newly formed C2-body binds to an acceptor molecule, oxaloacetate (OAA), and thereby enters the citric acid cycle (TCA, tricarboxylic acid cycle), where a successive oxidation down to carbon dioxide occurs. During this cyclic process, OAA is regenerated.

The reduction equivalents, that are released by the TCA cycle in form of NADH, are used in the respiratory chain for energy generation. These above-mentioned metabolic pathways are not only responsible for the generation of energy (catabolism) but also provide precursors for the de novo synthesis of cellular material (anabolism). In Figure 2.6, the precursors of central metabolic pathways and the monomers that are produced are listed.

<u>precursor</u>	<u>amino acid</u>	<u>RNA nucleotide</u>	<u>DNA nucleotide</u>
glucose 6-phosphate	alanine	ATP (Adenin)	dATP
fructose 6-phosphate	arginine	GTP (Guanin)	dGTP
ribose 3-phosphate	asparagine	CTP (Cytosin)	dCTP
3-phosphoglycerate	aspartate	UTP (Uracil)	dUTP (Thymin)
phosphoenolpyruvate	cysteine		
pentose 5-phosphate	glutamate		
erythrose 4-phosphate	glutamine		
oxalacetate	glycin		
succinyl CoA	histidine		
α -ketoglutarate	isoleucine		
	leucin		
	lysin		
	methionin		
	phenylalanin		
	proline		
	serine		
	threonine		
	tryptophane		

FIGURE 2.6: List of precursors and monomers. Monomers are basic components for proteins, RNA, and DNA.

2.3.5 Respiratory chain

The reduction equivalents, which are gained during the TCA cycle in the form of NADH, are used in the respiratory chain. During this process, NADH is reduced to NAD and the released protons are transported across the membrane to the extracellular space, while the electrons are transferred inside the membrane from one specific enzyme to the other, until they reach the terminal electron acceptor, which is oxygen. In the course of this respiratory chain, further proton translocation occurs from the cell to the exterior. Consequently a proton gradient is generated across the membrane. This proton motive force (PMF) is used by a membrane-bound enzyme, the ATP synthase. The protons enter the cell through a channel formed by the ATP-synthase and the energy originating in the proton gradient is stored in form of ATP (Figure 2.7). Theoretically, 38 molecules of ATP can be generated per molecule glucose that is degraded in metabolism by a respiratory cell.

2.3.6 Mass transport into the cell

The cytoplasmic membrane protects the cell from the environment. This implies, however, that nutrients that are metabolized by the cell need to overcome this barrier in order to enter the cell. Small molecules and ions usually can diffuse freely across the membrane, whereas bigger molecules cannot. To guarantee nutrient uptake, several transport systems exist. Basically three carrier systems can be distinguished: So-called uniporters, which transport a single substance from one side of the membrane to the other, symporters and

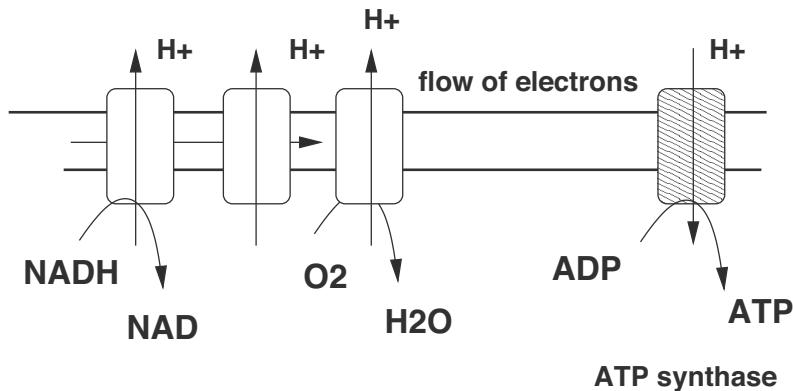


FIGURE 2.7: Simplified diagram of the respiratory chain.

antiporters. Both last mentioned transporters have in common that they require a second substance (energy donor) for the transport of the substance of interest. The second substance is co-transported together with the one of interest, thereby providing the energy. Symporters co-transport both substances in the same direction, whereas antiporters transport the second substance in the opposite direction (see Figure 2.8).

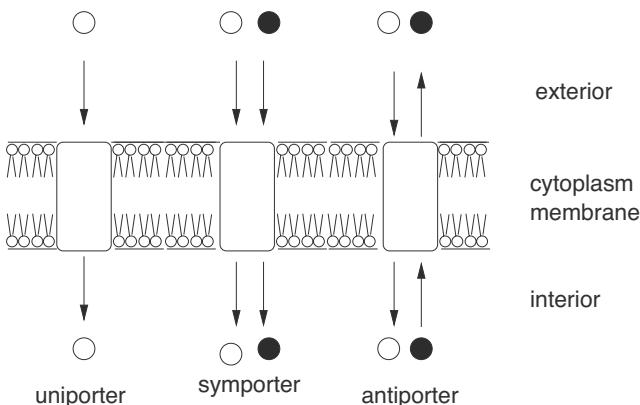


FIGURE 2.8: Cellular transport systems.

Furthermore, it can be distinguished between passive diffusion, active transport and group translocation. Passive diffusion describes the unspecific permeation of a substance into the cell along its concentration gradient. Therefore, an accumulation of the substance in the cell is impossible, as the rate of uptake and the intracellular level are dependent on the extracellular concentration of the substance. An accumulation of a desired substance can only be achieved by active transport, which means that the transport process is cou-

pled to the expenditure of energy. The energy can either be derived from the proton gradient or from ATP. In active transport as well as in group translocation processes the transport itself is mediated by proteins that are located in the membrane and serve as the actual carriers. In contrast to active transport, in a group translocation process the substance is chemically modified during the passage across the membrane. Therefore no concentration gradient is formed across the membrane, as the transported substance is chemically modified in the cell and therefore no longer identical to the extracellular substrate. One of the most prominent examples of group translocation systems is the phosphotransferase system (PTS), which is shown in Figure 2.9. By means of the PTS, sugars are taken up by the cell and concomitantly phosphorylated and thereby activated for further degradation. The phosphate group, which is translocated to the transported sugar, is provided by an enzyme's cascade (EI, HPr and EIIC) to the actual transporter (EIIC).

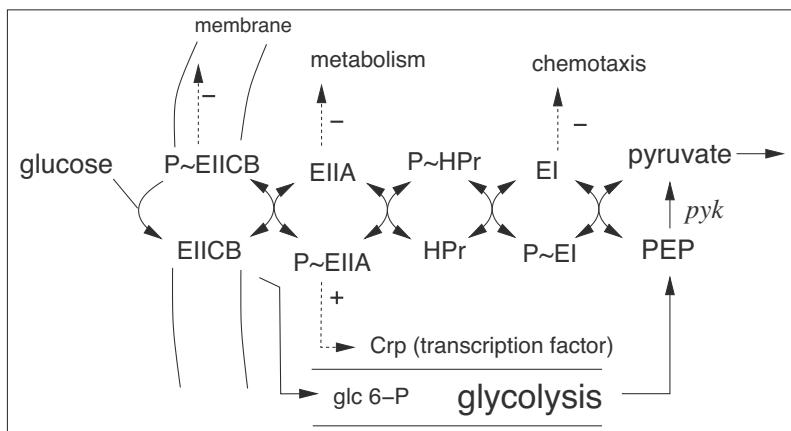


FIGURE 2.9: The bacterial PTS serves as transporter and sensor system. Upon entry glucose is phosphorylated to glucose 6-phosphate. The energy necessary is derived from PEP and transferred in a stepwise reaction to glucose. The involved enzymes EI, HPr and EI_A can be found in both forms, phosphorylated and nonphosphorylated. The phosphorylation state is a descriptor of the availability of nutrients and serves as a signal for the regulation of metabolic enzymes.

Apart from transporting nutrients into the cell, transporters can also fulfill sensory functions. The PTS for example serves as a sensor of the metabolic fluxes in the central metabolic pathway (glycolysis). This information finally leads to different regulatory outputs in order to perfectly adapt the cellular functions to a given situation.

2.4 Replication, Transcription and Translation

A gene is a piece of DNA (Desoxyribonucleic acid) that codes for a specific protein. Genes are located on the chromosome. The backbone of the DNA is a polymer of alternating sugar and phosphate molecules, to which the nitrogen bases are covalently bound. The DNA has four bases: adenine, thymine, guanine and cytosine. Adenine and thymine, as well as guanine and cytosine, are complementary, which means that they can form specific base pairs by hydrogen bonding. As a result the DNA appears as a double strand inside the cell. The complete DNA molecule consisting of the sugar, the phosphate and the specific base is called a nucleotide (see Figure 2.10). If several nucleotides covalently bind to each other the polynucleotide is formed.

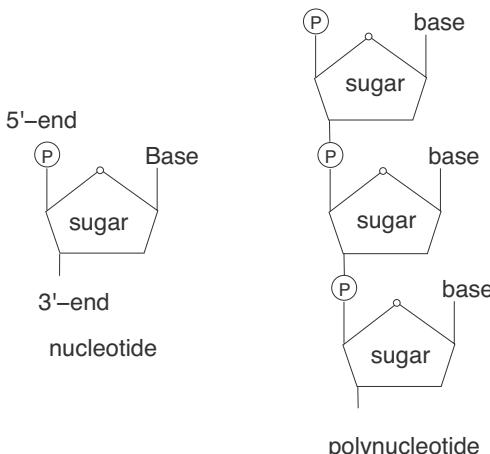


FIGURE 2.10: Schematic representation of a nucleotide and polynucleotide.

In a DNA single strand it can be differentiated between the 5'-end and the 3'-end. The 5'-end is the end of the polynucleotide strand to which the phosphate molecule is bound to the C5 atom of the sugar molecule, and the 3'-end is the side without phosphate. The information for the encoded protein is stored in the sequence of the DNA bases. The protein itself is built by a sequence of amino acids. Each amino acid (protein building block) is encoded on the chromosome by a base triplet (codon), a sequence of three bases.

During cell division, the DNA has to be copied. This results in two identical daughter chromosomes, which are passed to the daughter cells, so that in the end two genetically identical cells arise, which are called clones. The duplication of the DNA before cell division is called replication. An enzyme, the DNA polymerase, is responsible for this process (see Figure 2.11).

Additionally to the DNA, there is a second polynucleotide inside the cell,

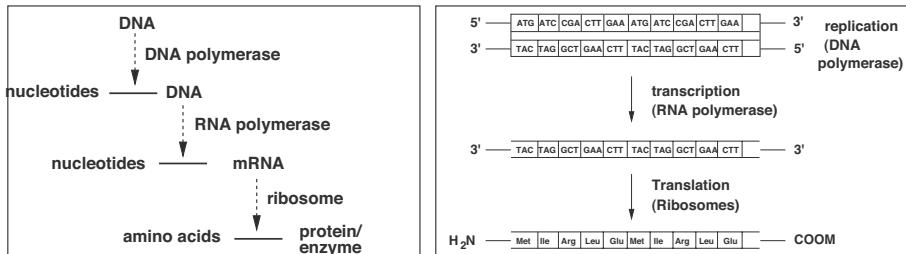


FIGURE 2.11: Dogma of molecular biology: The information stored on the DNA is converted into proteins by two steps, transcription and translation. Left: signal-orientated representation; Right: molecular biology-orientated representation.

the RNA (ribonucleic acid). The RNA fulfills three functions in the cell: As messenger-RNA (mRNA), it serves as a messenger of genetic information between the DNA and the ribosomes (the protein synthesis machineries of the cell). As transfer-RNA (tRNA), it delivers the building blocks for proteins to the ribosomes. The tRNA is a molecule with two specificities, one for the mRNA codon and the other for the corresponding amino acid; it therefore directly translates the base sequence into the protein sequence. And as ribosomal RNA (rRNA), it forms an important structural and catalytic element of the ribosomes. The DNA and the RNA backbone differ in the composition of the sugar molecule: DNA is made up from deoxyribose while RNA is made up from ribose. Furthermore, RNA is always single stranded and contains the base uracil instead of thymine, which can only be found in DNA (apart from very few exceptions). The process in which a part of the DNA sequence is transformed into RNA by an enzyme, the RNA-polymerase, is referred to as transcription.

The RNA-polymerase binds to the promoter, a specific sequence in the DNA. The promoter lies upstream of one or several genes and controls their transcription. The part of the sequence that is transcribed into mRNA is called operon. The process of transcription is shown in Figure 2.12 in the example of the *lac*-operon. The RNA-polymerase binds to the *lac*-promoter, which is located upstream of the genes *lacZ*, *lacY* and *lacA*. These three genes are transcribed into one RNA transcript and therefore build an operon, the *lac* operon. The first step of transcription, the binding of the RNA-polymerase to the promoter region, is called initiation. The initiation is followed by the elongation, the process in which the RNA-polymerase moves down the DNA chain and the mRNA transcript is formed. When a termination site is reached, it dissociates from the DNA and the mRNA is released. The mRNA has a quite short half life, and has therefore only a short time span available as matrix for protein synthesis by the ribosomes, before it gets degraded again.

The process of protein biosynthesis, i.e., the conversion of the mRNA into

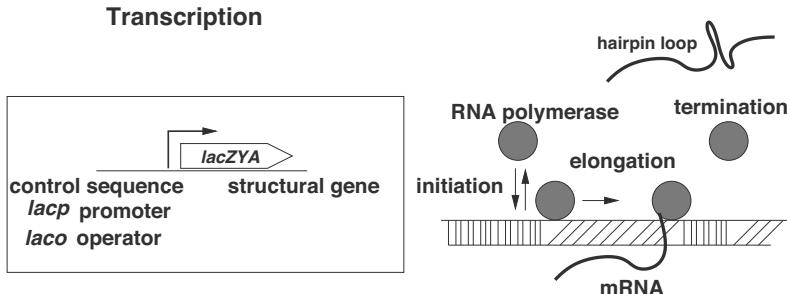


FIGURE 2.12: Transcription: Transcription of the DNA information into mRNA. The mRNA can have secondary structures, like a hairpin loop.

the amino acid sequence, is called translation. This process is responsible for the correct arrangement of amino acids in the polypeptide chain and hence for the synthesis of the correct and functional protein. The site of protein biosynthesis is the ribosomes. They bind to a specific sequence in the mRNA, the ribosome binding site (RBS) or Shine-Dalgarno sequence. Additionally, the tRNA is needed for translation. It is charged with its specific amino acid. In analogy to the transcription process the single steps in protein biosynthesis are referred to as initiation, elongation and termination (Figure 2.13). The binding of the ribosome to the Shine-Dalgarno sequence of the mRNA is called initiation. In the next step, the elongation, a stepwise elongation of the polypeptide chain occurs. Protein biosynthesis starts at AUG, the translation initiation codon, which means that the first amino acid in the growing polypeptide chain is always methionine. During elongation the ribosome moves along the mRNA and the correct charged tRNA is brought into position by the interaction between the mRNA codon and the anticodon (the complementary site of the tRNA). Three nucleotides code for one amino acid. Subsequently the amino acid is linked to the adjacent by a peptide bond and the free tRNA is released by the ribosome. Termination occurs after a stop-codon is reached. A stop-codon is a codon for which no charged tRNA exists. Instead of the tRNA, other proteins bind that cleave the attached polypeptide and help the dissociation of the ribosome. In many cases the correct folding of the proteins is dependent on proteins that bind to the nascent polypeptide chain and lead to a proper folding of the protein. These enzymes are called chaperones.

2.4.1 Regulation of enzyme activity and enzyme synthesis

During the bacterial cell cycle, hundreds of enzymatic reactions occur simultaneously. Of course not all of them take place with the same turnover rates or at the same speed, but they are subjected to well-balanced control mechanisms. Enzymatic reactions leading to components that are needed in higher amounts have to run faster or to a greater extent, whereas reactions

Translation

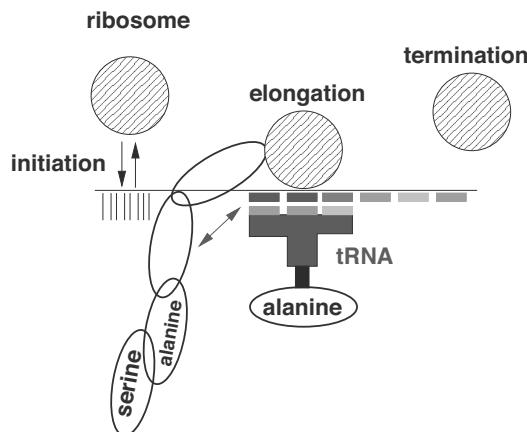


FIGURE 2.13: Translation: the process of protein synthesis.

leading to products that are rarely needed can run slower. The control of enzymatic reactions allows the cell to use its resources most efficiently.

The regulation of enzymatic activity happens on two levels: On one hand, the activity of the enzyme can be regulated; on the other hand, the amount of the enzyme can be modified. The first mechanism occurs on the protein level; therefore it is referred to as post-translational regulation. The control of the amount of an enzyme can either be achieved on the transcriptional level (via the amount of mRNA) or on the translational level, i.e., the level of protein synthesis.

The regulation of enzyme activity on the protein level can be achieved in various ways. One option is the so-called product inhibition. In this case the enzyme is repressed by its own reaction product when this accumulates to higher concentrations. Another widespread mechanism is the so-called feedback inhibition, which is frequently found in entire biosynthetic pathways. In feedback inhibition usually the end-product of a biosynthetic pathway (for example a specific amino acid) represses the first enzyme of this pathway. This repression occurs through reversible binding of the final product (inhibitor) to the regulatory center of the enzyme. Usually this causes a conformational change of the enzyme, so that the substrate can no longer bind to the enzyme, and the reaction therefore no longer takes place. When the concentration of the inhibitor decreases, it dissociates from the enzyme, which then again is in its active form. In addition to this, regulation of enzyme activity can also be achieved by covalent modification. Thereby, usually a small molecule covalently binds to the enzyme which leads to a change in the activity. This can happen through phosphorylation or methylation but also through binding of AMP or ADP. Regulation of enzyme activity on the protein level generally is a quite fast mechanism. It usually takes only a few seconds or less and is

therefore a quite effective way for adaptation to changes in the environmental conditions.

Due to energetic reasons the cell does not have all possible enzymes always available in the cytoplasm, but only synthesizes the enzymes which are necessary at a given time point. This is achieved by regulation on the transcriptional level. This process is quite slow, but allows the cell a long-term adaptation to changes in the environment. The two most common regulatory mechanisms on transcriptional level by prokaryotes are induction and repression. During induction, the genes of the induced operon are transcribed at a higher rate, and therefore the enzyme is produced, when the substrate is present. In contrast, during repression, transcription of the genes of the operon is repressed and therefore the production of the corresponding enzyme is blocked, when the product is present in high concentrations.

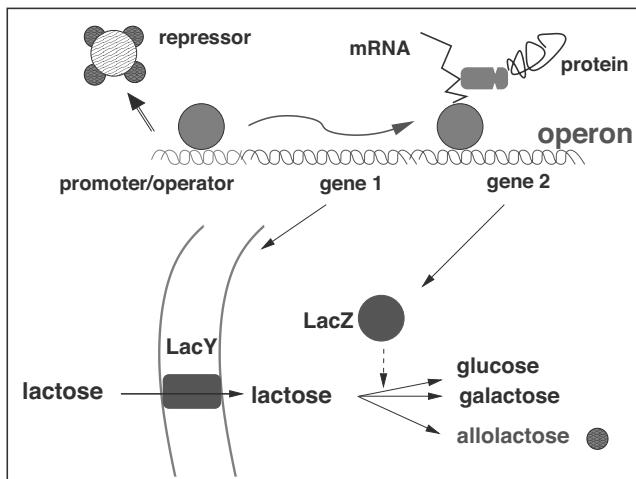


FIGURE 2.14: Induction of the *lac* operon.

In the following, the principle of induction is explained in more detail on the example of the *lac* operon of *Escherichia coli* (Figure 2.14). The enzymes, which are necessary for the uptake and degradation of lactose, are produced to a high extent in the presence of lactose. This is achieved by a regulatory sequence, the so-called operator, that is located together with the promoter in the *lac* operon. In the absence of lactose, a repressor binds to this operator and thereby inhibits the binding of the RNA-polymerase and the transcription of the operon. If lactose is present in the environment, a few molecules are taken up by the cell and are converted into allolactose and other products. This transport is possible as there is always a basal transcription of the *lac* operon, which ensures the presence of a small amount of the enzyme at all times. The allolactose then acts as an inducer and binds to the repressor, which in turn dissociates from the operator. This clears the way for the RNA-

polymerase, which binds to the promoter and initiates the transcription of the genes of the *lac* operon. Hence, it comes to an increased production of the enzymes responsible for the uptake and degradation of lactose. In this way an autocatalytic loop is established since the proteins themselves enhance their rate of synthesis.

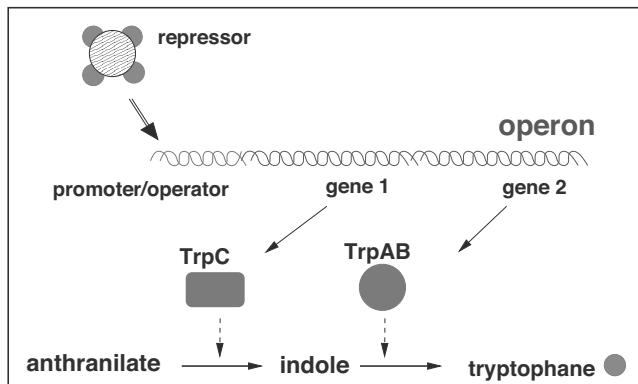


FIGURE 2.15: Repression of the *trp* operon.

During repression, an activating molecule binds to the regulatory protein. Binding of this activating molecule enables the regulator to bind to the operator, which then leads to repression of the transcription of the operon genes. One example for repression is the regulation of the tryptophan operon (Figure 2.15). The proteins necessary for the biosynthesis of tryptophan are only produced in the absence of tryptophan, thus in conditions where the cell has to synthesize tryptophan by itself. If tryptophan is present in the medium, it is taken up by the cell and binds the regulator of the *trp* operon. The regulator in turn can now bind the operator and thereby inhibits transcription of the genes responsible for tryptophan synthesis. When tryptophan gets scarce in the environment, a lack of tryptophan in the cytoplasm is the consequence and the repressor is present in its free form. It can therefore no longer bind the DNA, so that RNA-polymerase binds and transcription and production of the tryptophan biosynthesis enzymes starts.

2.4.2 Signal transduction

Regulation of bacterial metabolism is dependent on a variety of environmental influences and stimuli, as the availability of nutrients, temperature, the pH or the osmolarity of the surrounding medium. This information is sensed by the cell and transduced to the corresponding receptor through signal transduction chains. This usually results in modifications in the transcription rate of the according genes. Signal transduction systems are structured in a similar way as technical systems: Sensors convert extracellular information to intra-

cellular signals, which are processed and result in a cellular response. Such a simple signal transduction chain is shown in Figure 2.16.

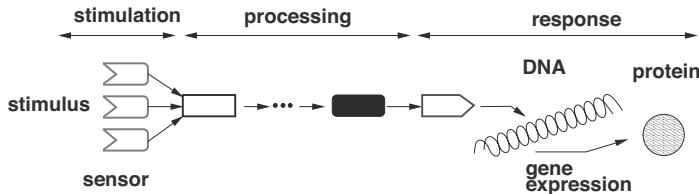


FIGURE 2.16: Structure of a signal transduction device. The sensor senses the extracellular stimuli and converts them into an intracellular signal. The signal is processed and generally leads to the activation or repression of a transcription factor which activates (turn on) or inhibits (turn off) protein synthesis.

A typical example for such a signal transduction chain is the so-called two-component system (TCS). TCSs consist of two proteins, the sensor kinase and the response regulator. The sensor kinase is located in the membrane and detects an extracellular signal. It gets autophosphorylated, that is the sensor kinase phosphorylates itself at a specific site on its cytoplasmatic side. Then, the phosphate group is transferred to the second protein of the two-component system, the response regulator. This regulatory protein usually is a DNA-binding protein, which regulates transcription. Thus, the phosphorylated response regulator can act as a repressor, while the nonphosphorylated protein cannot bind to the DNA. But also the opposite is found in which the phosphorylated response regulator induces transcription by binding to the DNA. In the last few years more and more evidence arose that the different two-component systems do not act completely independent, but that they are connected by a group of proteins on a post-translational level. These proteins serve as connecting elements between the different two-component systems and are referred to as TCS connectors. The synthesis of those TCS connectors is usually regulated by signals that are different from those of the involved TCS. This allows the interconnection of different systems. TCS connectors modulate the activity of the response regulators by tackling the steps that lead to the phosphorylation of the response regulator. For instance, this can happen by inhibition of the autophosphorylation of the sensor kinase or by inhibition of the dephosphorylation of the response regulator.

The TCS connectors act as a further level of control, which enables the cell to combine the responses to several different stimuli and therefore contributes to the fine-tuning of metabolism.

2.4.3 Cell-cell communication

Some microorganisms can communicate with each other through chemical messengers. This allows them to coordinate their behavior in a group dependent manner. *Vibrio fischeri*, for instance, is able to emit light (bioluminescence). This, however, only makes sense when a certain number of cells is present, because otherwise the emitted light would not be noticed at all. *V. fischeri* lives in symbiosis with some eukaryotes (for example *Euprymna scolopes*, a dwarf squid living along the Hawaiian shore), which have a specific light organ that is populated and thereby illuminated by *V. fischeri*.

The mechanism of sensing the own cell density is called quorum sensing. Quorum sensing enables bacteria to coordinate the expression of specific genes to the actual cell density. Bacteria that are able to perform quorum sensing constantly produce signaling molecules (so-called autoinductors), which are secreted and accumulate in the environment. Additionally they are equipped with a receptor for exactly this autoinducer. When the concentration of external autoinductors reaches a specific threshold (caused by a sufficient number of microorganisms in the surrounding area), the autoinductor binds to the receptor. This results in the transcription of specific genes, including those genes that are responsible for the production of the autoinductor itself. Consequently, its production is increased resulting in an amplification, so that in the end the receptor is completely activated and the group of genes is simultaneously transcribed in all cells. This makes a coordinated group behavior of microorganisms possible.

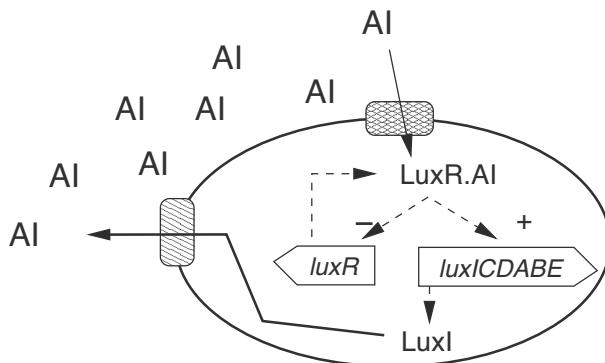


FIGURE 2.17: Quorum sensing system of *Vibrio fischeri*.

In Figure 2.17 the quorum sensing system of *V. fischeri* is shown. The genes, which are responsible for bioluminescence, are *luxCDABE*. They are located in one operon together with *luxI*, coding for the enzyme responsible for the synthesis of the autoinductor. Divergently transcribed is the gene *luxR*, which encodes the receptor protein. An increase in the cell density leads to an increase in the concentration of the autoinductor. After a specific threshold is

reached, the autoinductor binds to the receptor LuxR. The LuxR autoinductor complex in turn binds to the *luxICDABE* promoter and leads to induction of the transcription of the lux operon. This further increases autoinductor synthesis (positive feedback loop), as more LuxI is produced, and in manifestation of bioluminescence. Simultaneously, the LuxR autoinductor complex represses the *luxR* promoter, to prevent an overshoot of the system as a result of the positive feedback loop. Quorum sensing is not only employed by bacteria to emit bioluminescence, but is also used for other concerted actions as the coordinated expression of pathogenic factors or for the production of biofilm.

Chapter 3

Fundamentals of Mathematical Modeling

3.1 Definition — Overview of Different Model Classes

The two terms “system” and “model” are central keywords which are necessary for understanding the further proceedings. They will be explained in the following.

3.1.1 System

A system describes a number of components which are interacting with each other. An important step in distinguishing a system from its surroundings is the definition of the system borders. In most cases a certain volume or volume element is considered with a barrier that allows only a defined exchange of mass and energy with the surroundings. Usually the exchange of information is greater inside a system than with the surroundings, thus a certain autonomy of the system is granted. A system is characterized by two relevant factors: Inputs u influence the system and activate it, while state variables x are the physical-chemical factors that are influenced. State variables like components can be characterized by

- the entire number n_X of molecules X inside the system or inside the cell, respectively;
- the entire mass m_X of molecule X inside the system or inside the cell, respectively;
- the concentration c_X of the component X (here we have to define the reference state variable: In process engineering the volume V is typically used while in bioprocess engineering and systems biology the cell dry mass (DM) is used as a reference for intracellular components and mol/gDM is the respective unit);
- state on/off (the meaning is that the component is present or not);
- a probability P_n that says that we can expect n molecules of species X inside the system.

However, there are other state variables z like the cell age, cell size or other morphological parameters that might be of interest. Here, one might look at the overall population and write down a number density function $f(z)$ that characterizes it.

Depending on the choice a specific modeling approach is required as described in this book. Outputs $y = g(x)$ are state variables (or combinations of state variables) which can be measured and therefore are externally observable.

3.1.2 Model

A mathematical model shows the relationship of inputs u and state variables x with mathematical equations and can therefore be seen as an abstract and simplified picture of reality. The modeling of very complex systems nowadays is computer aided. This means that the existing knowledge is structured in an adequate manner and must be stored in a knowledge base. Therefore a model can also be seen as a formal representation of knowledge. A model must always be validated by suitable experiments. However, until today, there exists no measure for the validity of a model: The accuracy of a model can therefore only be assessed through its use. If a simple model is used as basis of process control and fair enough results are obtained, the process of modeling can be seen as complete. If a model shall be used for validating a biological hypothesis, it is referred to as “good” as long as all experiments can be described through it or hypotheses can be formulated and checked. A model can therefore be disproved with experiments, which can not be described (falsification). Figure 3.1 compares different model classes which differ in their mathematical description. The single subtypes are now described following the path on deterministic models; in doing so some classes are not explained in detail.

The deterministic approach is the way which is most often used for system description. The basis is chemical-physical fundamentals like the law of conservation of mass and the description usually occurs on a macroscopic level where the conversion of components is considered (and not individual elements like carbon C , oxygen O_2 , etc.). The cell is described with one or more compartments, whereby exchange flows between the compartments must be considered. As a result, one gets mass balance equations in the form of differential equations. Here, the interactions are described in a deterministic manner without stochastic events. In a stochastic approach events are described by random processes. This can be necessary if insufficient knowledge is available or if the rules, applicable for deterministic modeling, do not apply anymore. For example if the number of molecules involved is very small, the concurrence of molecules is modeled using a stochastic approach and not using the concentrations of the partners. A probability $P_n(t)$ is given which tells that the respective component is present with exactly n molecules at the time t . If a higher number of components is considered, it appears that the probability

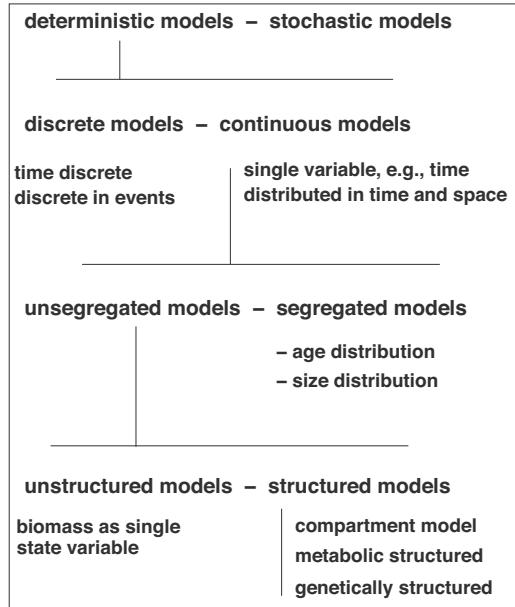


FIGURE 3.1: Overview over different model classes.

$P_{n1,n2,\dots}(t)$, which is very difficult to determine. The applicable approach to derive mathematical equations is explained below. Considering the probability for a certain number of molecules inside a single cell allows taking a mean value that leads back to the consideration of an “average” cell which is the basis of the deterministic approach.

Another differentiation occurs between discrete and continuous models. So far, only the time t was looked at as a continuous variable. If further dimensions like space coordinates are introduced, diffusion of the components inside the cell should be considered. This leads to a distributed model depending on the time t and the locus z , compared to a concentrated model, which is only depending on the time t (an applicable example is given in the chapter on signaling). Concentration gradients occur inside the cell if the diffusion constant is rather small.

If biological knowledge is available for a system but sufficient knowledge of the kinetic parameters is missing, the analysis of qualitative characteristics comes into consideration whereby discrete models are used. For discrete models, the time and/or the range of values of the state variables is split in fixed values. This method makes sense if switching processes or genetic modifications with [on/off]/[gene available, not available] are considered. Graphs of cellular networks can be regarded as discrete models since in its simplest form interactions are present or not. Furthermore prefixes can be added to

the applicable arrows, which indicate an activating or inactivating function of one component onto another.

Segregated models are characterized by the examination of distinct properties like a size distribution or an age distribution. Therefore, the basis is a single cell and not the “average” population as above. Formally, a property space is looked at and the processes are written down for small volume elements. This includes cell division (one cell makes two) as well as growth processes (from one cell a smaller one and a larger one emerge; an older cell births a younger one). The difficulty usually lies in properly specifying the regularities since they are dependent on numerous aspects. Since one is interested in the temporal change of distributions, such models are also called statistical models. On the other hand statistical models are widely used for the analysis of high dimensional data sets where one tries to infer the structure of a model.

A specific term from bioprocess engineering is the differentiation between structured and unstructured models. In unstructured models, biomass is described with one single state variable whereby in structured models, equations are set up for single components of the cell (e.g., metabolites, protein fractions or DNA). As it was indicated above, the choice of the type of model is made by looking at the aim. Next to the better understanding of complex processes, the use and the application of the model is essential for the choice of a certain model type. Today, models also find a possible field of application in planning of experiments and in formulating hypotheses. Modeling approaches and simulation tools are therefore accordingly explained in literature [1].

To summarize there exist a number of possibilities to describe state variables for cellular systems. Table 3.1 shows in which way some of the descriptions can be converted.

3.1.3 Process of modeling

Modeling can be seen as an iterative process that is geared to specific aims. These aims can be in pure research as well as in biotechnology or medicine. In a first step, the definition of input and state variables is necessary. The state variables are of an extensive nature, in order to derive the corresponding mass balance equations. Usually one only gets a very rough description of the system with a first model, which must be refined in further steps.

A comparison with experimental data leads to alterations in the model structure and the kinetic parameters used. For many enzymes detailed knowledge on the individual reaction steps and the time constants are uncertain or unknown. Here, methods from structure and parameter identification come into play that allow to reconstruct (to a certain degree) the model structure and to estimate the kinetic parameters.

TABLE 3.1: Relation between state variables and description methods.

	Meaning	Relation	Interpretation
P_n	Number density function:	$\int P_n \, dn = 1;$	Integration leads to the secure event
	Probability to have n molecules of X inside the cell	$\int n \, P_n \, dn = n_X$	Integration leads to the average molecule number
f	Number density function:	$\int f(z) \, dz = N$	Integration leads to the overall number of cells
	Probability to have a number of cells with characteristic z	$\int z \, f(z) / \int f(z) = Z$	Integration leads to the average value of the cell descriptor
n_X	Molecule number	$\hat{n}_X = 0$ for $n_X < n_{X_s}$	Discretization leads to binary values

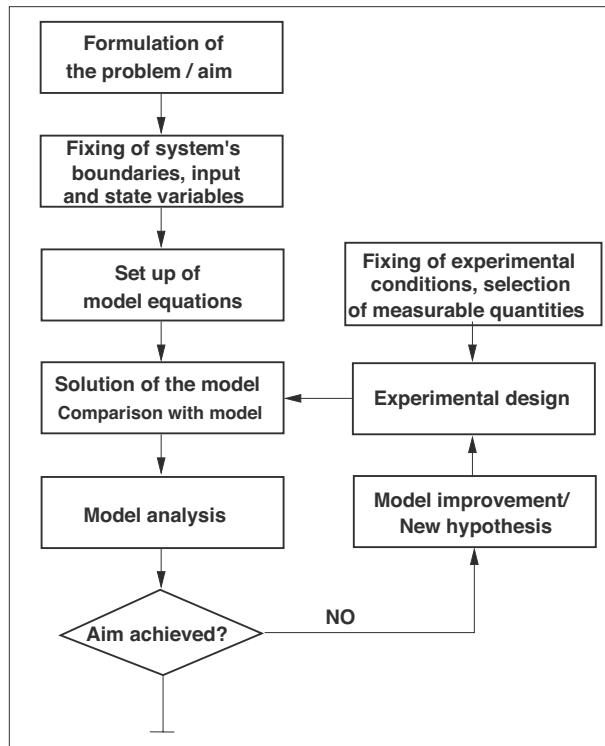


FIGURE 3.2: The single steps in the process of modeling: A basic step during this iterative process is model calibration, the comparison of the output of the model with experimental data. It serves as basis for following model analysis, which is only useful if the model is meaningful.

3.2 Basics of Reaction Engineering

Cellular networks are mainly characterized by the conversion of components by enzymes. Here, not only the numbers of molecules that are converted are important but also the velocity with which this conversion takes place. Therefore, basics of reaction engineering are shortly discussed here. A single reaction known for cellular networks is given in the following form:



Coefficients γ_i are called stoichiometric coefficients. It is generally accepted that the γ_i on the left side (substrates) are negative while the γ_j on the right side are positive (products). With the arrows given in the equation above, a mathematical treatment is hardly possible; therefore one uses the equal sign

to express the same as above and one gets after resorting and omitting the absolute signs:

$$0 = \gamma_A A + \gamma_B B + \gamma_C C + \gamma_D D. \quad (3.2)$$

In mathematical terms this is a linear equation which can be re-written in vector form with the two vectors \underline{n}^T with stoichiometric coefficients and \underline{K} with components:

$$\underline{n}^T = (\gamma_A \quad \gamma_B \quad \gamma_C \quad \gamma_D), \quad \underline{K} = \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} \quad (3.3)$$

and we get with the defined quantities:

$$0 = \underline{n}^T \underline{K}, \quad (3.4)$$

Having q reactions, new lines are added but the vector of components is unchanged. In this way all row vectors can be combined into a single matrix N^T and the overall reaction system can be written as:

$$0 = N^T \underline{K}. \quad (3.5)$$

with the $q \times n$ matrix N^T and the $n \times 1$ vector of components \underline{K} .

EXAMPLE 3.1 Reaction system with two reactions and three components.

The following reaction system is considered:



Writing both reactions as equations results in:

$$\begin{aligned} 0 &= -A - B + C \\ 0 &= -2A + B. \end{aligned} \quad (3.7)$$

With vector $\underline{K}^T = (A \quad B \quad C)$ matrix N^T reads:

$$N^T = \begin{pmatrix} -1 & -1 & 1 \\ -2 & 1 & 0 \end{pmatrix}. \quad (3.8)$$

For this representation it doesn't matter if the reaction is reversible or irreversible. This fact comes into play when the reaction velocity is considered (or if the reversible reaction is split into two single reactions).

Reaction engineering involves the question at which reaction rate r substance transformation occurs. The reaction rate is given in *mol/l time* or in *mol/gTM time* if considering intercellular networks. It depends on the following factors:

- temperature, pH value
- concentration of the participating reactants
- enzyme concentration

Assuming a large number of molecules that react with each other, the reaction rate is directly proportional to the participating substances' concentration. In the forward reaction r_{1f} from above:



the rate can be described as:

$$r_{1f} = k_{1f} c_A c_B \quad \text{since} \quad r_{1f} \sim c_A \quad \text{and} \quad r_{1f} \sim c_B, \quad (3.10)$$

with the reaction rate parameter k_{1f} . For the back reaction



one obtains:

$$r_{1b} = k_{1b} c_C. \quad (3.12)$$

For the reversible reaction combined from the two partial reactions above one obtains:

$$r_1 = k_{1f} c_A c_B - k_{1b} c_C. \quad (3.13)$$

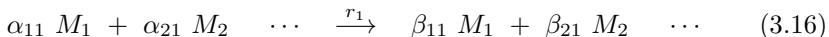
The signs indicate the direction of the reaction. A treatment analogous to above gives for the second reaction:



with the rate:

$$r_2 = k_{2f} c_A^2 - k_{2b} c_B. \quad (3.15)$$

In general a substance (or component) can be a substrate and a product at the same time. Furthermore, it is often desired to describe an elevating or decreasing influence on the reaction rate by additional components which in itself are not degraded by the reaction. Therefore, in an alternative description, reaction equations are considered as irreversible and are given in the following form:



where α_{ij} and β_{ij} are positive numbers; then stoichiometric coefficients are defined as $\gamma_{ij} = \beta_{ij} - \alpha_{ij}$. Note that in this notation, the first index i refers to

the compound and the second one to the reaction j . If a component appears on both sides with the same stoichiometry, it can be considered as a signaling input since the component is not consumed nor synthesized in this reaction but the reaction rate r then depends on the stoichiometric coefficients on the left hand side of the equation above:

$$r_j = k_j \prod_i c_i^{\alpha_{ij}}. \quad (3.17)$$

If the reaction is reversible, the equation for the backward reaction can be written in the same way.

In reaction equilibrium, forward and feedback rate — we now have to consider two reactions (index 1) and (index 2) — are equal and the equilibrium constant K_{Eq} results in:

$$K_{Eq} = \frac{k_1}{k_2} = \frac{\prod_i c_i^{\alpha_{i2}}}{\prod_i c_j^{\alpha_{i1}}}. \quad (3.18)$$

The equilibrium constant is defined as the inverse of a dissociation constant and describes the ratio of components, that is, the ratio of products and substrates.

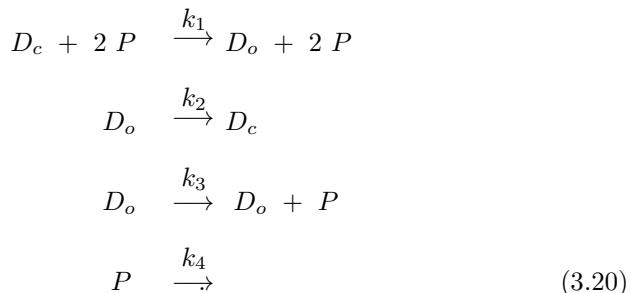
EXAMPLE 3.2 Reversible reaction from above.

For the reversible reaction given above in Equation (3.14) the equilibrium constant is given by:

$$K_{Eq} = \frac{k_{2f}}{k_{2b}} = \frac{c_B}{c_A^2}. \quad (3.19)$$

EXAMPLE 3.3 Simple model for gene expression induction.

The following reaction system is considered:



Here a protein P is synthesized depending on the state of the DNA which can be closed (c , no synthesis) or open (o). The reversibility is described with the

first two reactions while Protein P acts in an autocatalytical way to activate its own synthesis in the first reaction. The protein itself is synthesized in reaction three where the promoter binding site D_o is the driving force. The protein is finally degraded in reaction four. There are four reactions and three components involve. The stoichiometric matrix N^T therefore reads, taking into account the columns D_c , D_o and P and the stoichiometry given,

$$N^T = \begin{pmatrix} -1 & 1 & \underbrace{-2+2}_{=0} \\ 1 & -1 & 0 \\ 0 & \underbrace{1-1}_{=0} & 1 \\ 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{pmatrix}. \quad (3.21)$$

For the individual reaction rates the following is obtained:

$$\begin{aligned} r_1 &= k_1 D_c P^2, & r_2 &= k_2 D_o, \\ r_3 &= k_3 D_o, & r_4 &= k_4 P. \end{aligned} \quad (3.22)$$

From the first two rates, an equilibrium constant could be determined. In equilibrium both the reaction rates are equal and one gets:

$$r_1 = r_2 \rightarrow K_{Eq} = \frac{k_1}{k_2} = \left. \frac{D_o}{D_c P^2} \right|_{Eq}. \quad (3.23)$$

3.3 Stochastic Description

A stochastic system description is necessary if the number of interacting molecules and components in a cell is very small. Then, single reactions are considered as stochastic processes. Thereby they are characterized by a number of random events and the time course of the reaction partners is followed. In the stochastic approach the number of molecules of the component X is assigned a specific probability. It is written $\#X = n$ for the number of molecules n of X . The probability that n molecules of X are present at the time t inside the cell is described with $P_n(t)$.

Since one is interested in the temporal development of the components, the total time span is considered, and the probability how many molecules $\#X_{t_k} = n_k$ are present at distinct points of time t_k is of interest. Single events must have occurred over the whole period of time, so the probability of interest can be expressed the following way:

$$P(\#X_{t_0} = n_0 \wedge \#X_{t_1} = n_1 \wedge \#X_{t_2} = n_2 \cdots \#X_{t_k} = n_k). \quad (3.24)$$

The connection with the \wedge (and) symbol states that all events must have occurred. To express the equation verbally, one can say that we are interested in the probability that X has n_0 molecules at time t_0 , n_1 molecules at time t_1 , and so on. Further calculations are very difficult using the given form, so the system is simplified and we look back only one step of time; therefore we are interested in the probability:

$$P(\#X_{t_k} \wedge \#X_{t_{(k+1)}}). \quad (3.25)$$

This probability can be rewritten to a conditional probability (see Appendix for conditional probabilities):

$$P(\#X_{t_k} \wedge \#X_{t_{(k+1)}}) = P(\#X_{t_{(k+1)}} | \#X_{t_k}) P(\#X_{t_k}) \quad (3.26)$$

whereby each possible precondition $P(\#X_{t_k})$ must be considered and, therefore, the following term of sums results:

$$P(\#X_{t_{(k+1)}}) = \sum_m P(\#X_{t_{(k+1)}} = n | \#X_{t_k} = n_m) P(\#X_{t_k} = n_m). \quad (3.27)$$

A summation is possible since the preconditions ($n_m = 1, n_m = 2, \dots$) exclude each other: The system could have been only in one of the states before. Because of the stoichiometry of the reaction, the system simplifies even further.

Here, the idea arises to only look at a little step of time Δt . In this time step the stoichiometric conversion with stoichiometry γ happens (γ describes the number of molecules that is converted) or the system stays unchanged (time step Δt accounts for the time difference t_{k+1} and t_k). Then, the dependence on time of P means that one is interested in $P_n(t + \Delta t)$ as a function of $P_{n-\gamma}(t)$. This correlation looks like the following in a more compact way of writing the equation:

$$P_n(t + \Delta t) = \sum_{n-\gamma} P_{n-\gamma,n} P_{n-\gamma}(t) \quad \text{with} \quad (3.28)$$

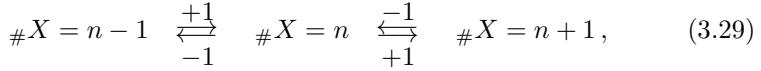
$$P_{n-\gamma,n} \equiv P(\#X(t + \Delta t) = n | \#X(t) = n - \gamma)$$

$$P_{n-\gamma} \equiv P(\#X(t) = n - \gamma).$$

The stochastic process described above is also referred to as Markov process. Thereby it is assumed that the changeovers are not dependent on time and the knowledge of the previous history is very limited. The prediction of the next step of time only happens on the base of the current state and the previous state.

EXAMPLE 3.4 Master equation for the scheme of a simple reaction.

In this example, only one component X is looked at, whereby the transfers only allow an increase or loss of one molecule of X . The following processes are considered ($\gamma = \pm 1$):



where the state $\#X = n$ is in the center of the equation. With this, the stoichiometry of the probabilities are $P_{m,n} = 0$, except for $m = n-1$, $m = n+1$ or $m = n$ (number of molecules is unchanged in the time interval).

First we start by considering all probabilities for degradation of X that must sum up to 1: If the state is n molecules, a molecule can be added during the next time step, a molecule can be lost or n molecules remain. With this, one obtains:

$$P_{n,n} + P_{n,n+1} + P_{n,n-1} = 1 \longrightarrow P_{n,n} = 1 - P_{n,n+1} - P_{n,n-1}. \quad (3.30)$$

In a second step, for example, using Equation (3.28) one gets:

$$\begin{aligned} P_n(t + \Delta t) &= P_{n-1,n} P_{n-1}(t) + P_{n+1,n} P_{n+1}(t) + P_{n,n} P_n(t) \\ &= P_{n-1,n} P_{n-1}(t) + P_{n+1,n} P_{n+1}(t) + \\ &\quad (1 - P_{n,n+1} - P_{n,n-1}) P_n(t). \end{aligned} \quad (3.31)$$

The transition probabilities must be stated for each single transition: Normally it is supposed that the transfers are proportional to the current number of molecules and the time interval Δt . With this, one obtains:

$$\begin{aligned} P_{n-1,n} &= k^+ (n - 1) \Delta t \\ P_{n+1,n} &= k^- (n + 1) \Delta t \\ P_{n,n-1} &= k^- n \Delta t \\ P_{n,n+1} &= k^+ n \Delta t. \end{aligned} \quad (3.32)$$

One gets further:

$$\begin{aligned} P_n(t + \Delta t) &= k^+ (n - 1) \Delta t P_{n-1} + k^- (n + 1) \Delta t P_{n+1} \\ &\quad + (1 - k^+ n \Delta t - k^- n \Delta t) P_n, \end{aligned} \quad (3.33)$$

so the difference quotient can be written as:

$$\begin{aligned} \frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} &= k^+ (n - 1) P_{n-1}(t) + k^- (n + 1) P_{n+1}(t) - \\ &\quad (k^+ + k^-) n P_n(t). \end{aligned} \quad (3.34)$$

For Δt towards 0, the following differential quotient results:

$$\frac{dP_n}{dt} = k^+ (n - 1) P_{n-1} + k^- (n + 1) P_{n+1} - (k^+ + k^-) n P_n. \quad (3.35)$$

Equation (3.35) is referred to as (Chemical) Master Equation (CME). In general form, the master equation looks like the following, if all reactions q in which the component is involved are regarded:

$$\frac{dP_n}{dt} = \sum_q a_q(n - \gamma_q) P_{n-\gamma_q} - a_q(n) P_n \quad (3.36)$$

whereby a_q represents propensities and γ_q stoichiometric coefficients of reaction q . Therefore, the propensities represent the transition probability in relation to the time. This system of differential equation (an equation must be written for each $n = 1, 2, \dots$) is barely solvable analytically for larger networks with a different number of molecules. Therefore, this is often managed using an approximation, whereby the point of time and the choice of each transition is simulated (see below).

To be able to compare with the deterministic approach, an arithmetic mean can be calculated from probabilities. The arithmetic mean is calculated using the relationship given in Table 3.1 from above:

$$c_X = \langle X(t) \rangle = \sum_{n=1}^{\infty} n P_n(t). \quad (3.37)$$

If the relations determined above are inserted, one obtains the following:

$$\begin{aligned} \langle X(t + \Delta t) \rangle &= \sum_{n=1}^{\infty} n P_n(t + \Delta t) \\ &= \sum_{n=1}^{\infty} n P_{n-1,n} P_{n-1} + n P_{n+1,n} P_{n+1} \\ &\quad + n (1 - P_{n,n+1} - P_{n,n-1}) P_n. \end{aligned} \quad (3.38)$$

If the sum is written out and terms are combined, one obtains:

$$\begin{aligned} \langle X(t + \Delta t) \rangle &= \sum_{n=1}^{\infty} (n + P_{n,n+1} - P_{n,n-1}) P_n(t) \\ &= \sum_{n=1}^{\infty} (n + k^+ n \Delta t - k^- n \Delta t) P_n(t) \end{aligned}$$

$$\begin{aligned}
&= (1 + (k^+ - k^-) \Delta t) \underbrace{\sum_{n=1}^{\infty} n P_n(t)}_{\langle X(t) \rangle} \\
&= \langle X(t) \rangle + (k^+ - k^-) \Delta t \langle X(t) \rangle . \quad (3.39)
\end{aligned}$$

Combining left and right side, we get:

$$\frac{\langle X(t + \Delta t) \rangle - \langle X(t) \rangle}{\Delta t} = (k^+ - k^-) \langle X(t) \rangle . \quad (3.40)$$

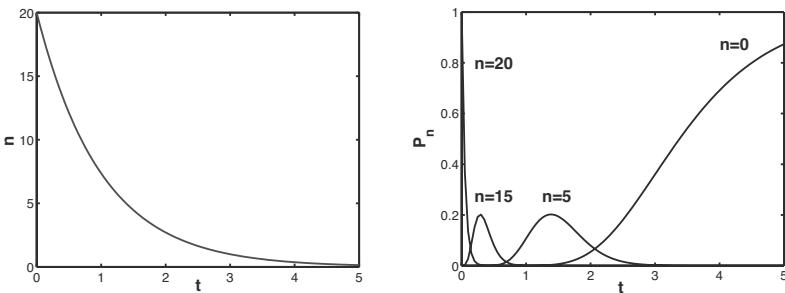


FIGURE 3.3: On the left: Mean value for X . On the right: Temporal trend of the probabilities $P_n(t)$.

So, the arithmetic mean represents the concentration c_X of X . Thus, one obtains the following differential equation for the limit $\Delta t \rightarrow 0$:

$$\frac{dc_X}{dt} = \underbrace{(k^+ - k^-)}_k c_X \quad (3.41)$$

whereby k is the reaction velocity constant. If one gets a positive value, more X is produced than used and X will accumulate. In the other case, X will be degraded. For this simple case, the differential equations (3.35) can be solved numerically. Figure 3.3 left shows the mean value for X (exponential degradation with $k^+ = 0$ and $k^- = 1$) and plot on right shows the temporal distribution of the probabilities P_n for $n = 0, 5, 15, 20$. At the point of time $t = 0$, the following starting condition was applied: $P_{20}(t = 0) = 1$.

It is possible to read off the probabilities that there are exactly n molecules of X at each point of time. Figure 3.4 on the left shows the behavior for chosen points of time $t = 0.5$ and $t = 2.5$ on the right and provides the according distribution of molecules. Note that for an increasing number of molecules the results are crucial since one can not simulate any high number and therefore the probabilities don't add up to 1.

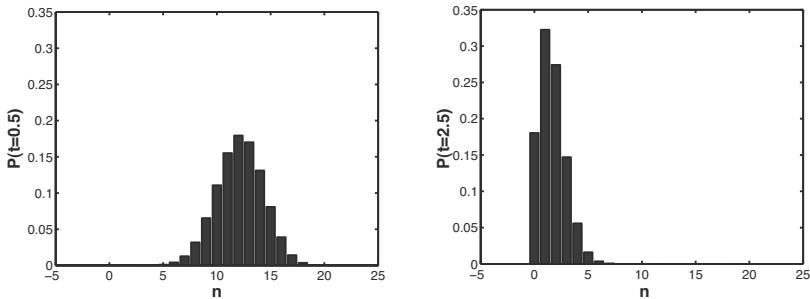


FIGURE 3.4: On the left: Distribution for time point $t = 0.5$. On the right: Distribution for time point $t = 2.5$.

3.3.1 Stochastic simulation using the Gillespie algorithm

As it was shown above, to simulate the master equation for one single species is already difficult. If the probability function P is now formulated for a small volume element V with metabolites S_1, S_2, \dots with

$$P_{s_1, s_2, \dots}(t), \quad (3.42)$$

no techniques are available today to derive a solution for larger reaction networks in a limited period of time. To circumvent this problem a Monte Carlo simulation procedure [2] was developed. Thereby the calculations are repeated several times to get a high statistical significance of the simulation results. Thereby two central questions arise:

- When does the next reaction occur?
- Which reaction will it be?

Using the approach introduced previously, one looks for the probability $P(\tau, \mu) d\tau$ of the given reaction system at time t , so the next reaction occurs during the time interval $[t + \tau, t + \tau + d\tau]$ and is reaction r_μ ($1 \leq \mu \leq q$) (no reaction occurs during the interval $t + \tau$). The natural value μ labels the number of the reaction. The approach is the following: Firstly, each reaction must be labelled with a number h_μ , which explains about the combinatorics of the individual reaction. Terms h_μ indicate the combinatorics of the molecules. If there are each 5 molecules of two kinds S_1 and S_2 in a reaction compartment, each molecule of S_1 can react with one of S_2 . There are 25 possibilities. However, if the molecules of S_1 react amongst each other, pairs out of 5 must be formed and 10 possibilities result. Generally stated, the terms h_μ result from the product of the binomial coefficients of the reaction partners involved with the amount of elements n_i and the stoichiometric coefficient γ_i , where

index i refers again to the component:

$$h_\mu = \prod_i \binom{n_i}{\gamma_i}. \quad (3.43)$$

EXAMPLE 3.5 Calculation of h_μ for elementary reactions.

For elementary reactions, these probabilities are the following (whereby the small letters indicate the number of molecules):

S	\longrightarrow	$M,$	$h_\mu = n_s$
$S_1 + S_2$	\longrightarrow	$M,$	$h_\mu = n_{s1} n_{s2}$
$2 S$	\longrightarrow	$M,$	$h_\mu = n_s \frac{n_s - 1}{2}$
$S_1 + 2 S_2$	\longrightarrow	$M,$	$h_\mu = n_{s1} n_{s2} \frac{n_{s2} - 1}{2}$
$3 S$	\longrightarrow	$M,$	$h_\mu = n_s \frac{n_s - 1}{2} \frac{n_s - 2}{3}$

The assignment of the average (reaction-) probability $c_\mu d\tau$ to each reaction occurs during the interval $d\tau$ that happens in a second step. c_μ matches the reaction constant k_μ of a deterministic system, whereby a conversion, which is described below, must occur. Then, the expression

$$h_\mu c_\mu d\tau = a_\mu d\tau \quad (3.44)$$

is the probability that the reaction r_μ occurs. It applies:

$$P(\tau, \mu) d\tau = P_0(\tau) a_\mu d\tau \quad (3.45)$$

with the probability $P_0(\tau)$ that no reaction occurs during the time interval $[t, t + \tau]$. This probability is calculated using an e-function with a negative exponent:

$$P_0(\tau) = e^{-\sum_j a_j \tau}. \quad (3.46)$$

The meaning is that the chance that nothing happens decreases exponentially with increasing time t and the following term results for the equation above:

$$P(\tau, \mu) = a_\mu e^{-\sum_j a_j \tau}. \quad (3.47)$$

The simulation algorithm splits the calculation into two parts:

$$P(\tau, \mu) = P_1(\tau) P_2(\mu|\tau) \quad (3.48)$$

whereby $P_1(\tau) d\tau$ states the probability that any reaction occurs during the time interval $[t + \tau, t + \tau + d\tau]$:

$$P_1(\tau) = \sum_j P(\tau, \mu) = a e^{-a\tau}, \quad (3.49)$$

with $a = \sum_j a_\mu = \sum_j h_\mu c_\mu$. Taking this into account one gets for P_2 :

$$P_2(\mu|\tau) = \frac{P(\tau, \mu)}{P_1(\tau)} = \frac{a_\mu}{a}. \quad (3.50)$$

Now two random numbers are drawn and used for determining time and reaction:

- Draw random number $rand_1$ (equally distributed) and calculate $\tau = \frac{1}{a} \ln \frac{1}{rand_1}$.
- Draw a second random number $rand_2$, and determine μ :

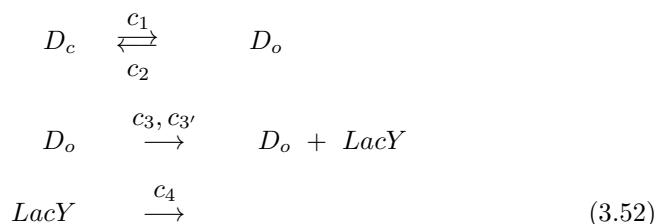
$$\sum_{j=1}^{\mu-1} a_j < rand_2 a \leq \sum_{j=1}^{\mu} a_j. \quad (3.51)$$

Value μ is set in as the value whose index was added last.

EXAMPLE 3.6 Gene expression of the lactose operon — autocatalytic loop.

The simple model to describe gene expression of the lactose operon presented here does not consider all known interactions but concentrates on the formation of the protein and describes the influence of the protein back onto the promoter dynamics. The scheme describes the promoter D in two states (Figure 3.5): open D_o and closed D_c .

The transition is realized by docking of a molecule of RNA polymerase that is, however, not considered as a state variable in our example. After the molecule has bound, the promoter is open and the synthesis of a mRNA molecule and the proteins, represented by the protein $LacY$, can take place. A simplified reaction scheme with a total of four reactions (represented by c_1 to c_4) illustrates the situation:



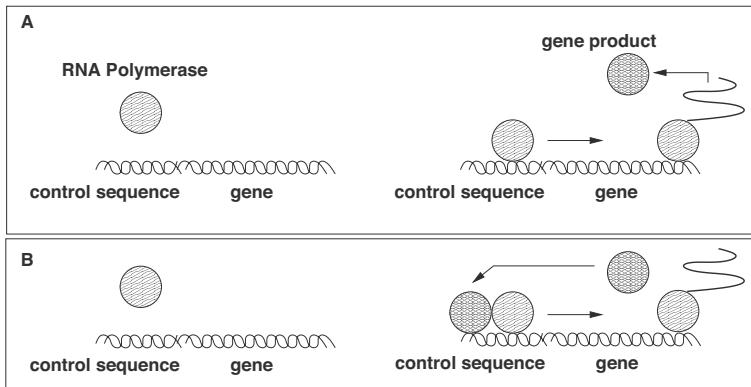


FIGURE 3.5: **A** Binding of the RNA polymerase induces gene expression. **B** Autocatalytic loop. The gene product is a transcription factor, speeding up its own synthesis (in the biological example indirectly via other components).

whereby both reactions occurring first describe the transition from closed to open promoter, the third one the synthesis of a protein and the last reaction the degradation of the protein (including dilution due to growth). In reaction 3 also a basal expression rate is considered independent of D_o (c'_3).

Simple rates are assumed for the reaction rates: The transition is proportional to the number of molecules available. The mRNA is not considered explicitly; therefore a synthesis in dependence of an open promoter is assumed for the gene product $LacY$. The lower part of the illustration shows the interesting case of a feedback loop. The gene product influences its own transcription positively. Looking at this case, the reaction rate for the transition from closed to open complex is additionally proportional to the gene product $\sim LacY^2$. This approach is justified since several inducer molecules are necessary to completely inactivate the repressor.

In this example the master equation will be set up and solved numerically; further on, the results are compared with a simulation study with the Gillespie algorithm. The derivation of the master equation can be simplified, so the sum of both promoter configurations is constant (D_t). The following illustration in Figure 3.6 gives an overview of all propensities, which must be considered.

All in all, eight transitions are necessary which describe the state in which D_o just has n molecules and $LacY$ has m molecules. The master equation is derived as above and all changes are summed up. One obtains the following differential equations for the probabilities that there are n molecules of D_o and m molecules of $LacY$ inside the cell ($P_{(n,m)}$ is the notation for the state

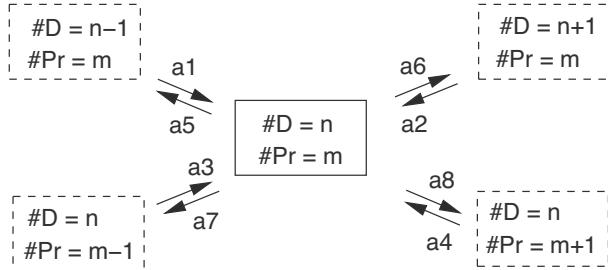


FIGURE 3.6: Illustration of the transitions, which must be considered if the two kinds of molecules D_o and $LacY$ are involved.

looked for):

$$\frac{dP_{\binom{n}{m}}}{dt} = a_1 P_{\binom{n-1}{m}} + a_2 P_{\binom{n+1}{m}} + a_3 P_{\binom{n}{m-1}} + a_4 P_{\binom{n}{m+1}} - (a_5 + a_6 + a_7 + a_8) P_{\binom{n}{m}} \quad (3.53)$$

with the propensities:

$$\begin{aligned} a_1 &= c_1 (D_t - D_o + 1) & a_2 &= c_2 (D_o + 1); \\ &\times LacY (LacY - 1)/2; \\ a_3 &= c_3 D_o + c'_3; & a_4 &= c_4 (LacY + 1); \\ a_5 &= c_2 D_o; & a_6 &= c_1 (D_t - D_o) LacY (LacY - 1)/2; \\ a_7 &= c_4 LacY; & a_8 &= c_3 D_o + c'_3. \end{aligned} \quad (3.54)$$

When looking at propensities, one must consider that the closed promoter is not stated explicitly but is considered in the total amount of promoter D_t . This becomes clear in the terms from a_1 to a_6 . The stochastic simulation using the Gillespie algorithm occurs using the same propensities. The simulation is repeated several times and the statistics is evaluated as shown in Figure 3.7.

In the case of the autocatalytic regulation, one sees that two populations form. In one partial population, only a basal gene expression occurs; in the other partial population, the system is fully induced. In a real experiment, one therefore would obtain a heterogeneous population, in which only a part is synthesizing the protein. The induction described in the chapter on biology is a typical example for this kind of regulation. For the lactose operon model, the calculated behavior was also observed experimentally [3].

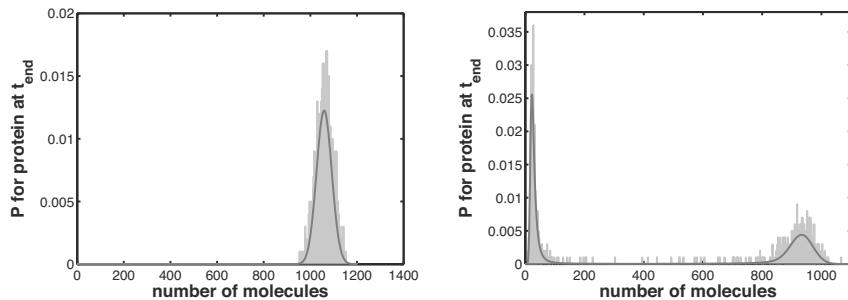


FIGURE 3.7: On the left: Histogram for the protein at t_{end} of the simulation study of the unregulated system (1000 simulations) and the solution of the master equation. On the right: Histogram of the regulated system (1000 simulations) and the solution of the master equation. Time considered is 1200 minutes, which is 20 hours.

3.4 Deterministic Modeling

In deterministic modeling, balance equations of extensive state variables are formed. A balance equation shows the variation of a state variable in dependence of mass flow and energy flow and is thus the fundamental equivalence structure. Since cells are influenced by their environment that itself is not always in a steady state, general equations will be derived for the example of the bio-reactor as illustrated in Figure 3.8.

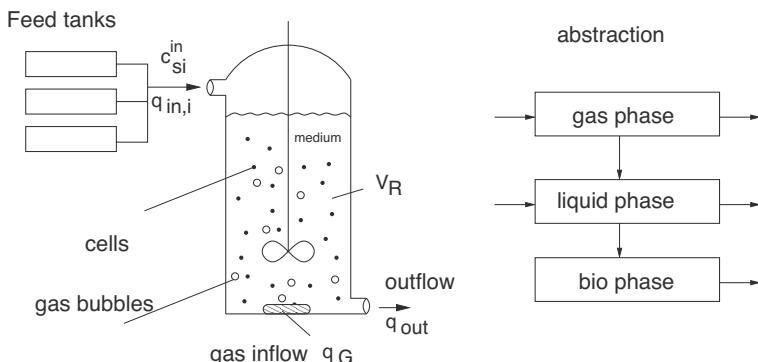


FIGURE 3.8: Bio-reactor with several inlets $q_{in,i}$ and feed concentration c_{Si}^{in} and outflow q_{out} . There is no biomass in the feed. The reactor can be considered abstractly as a system with three interacting phases. Density ρ is taken as a constant.

The bio-reactor is characterized by oxygen influx q_G and influx and outflow $q_{in,i}$, q_{out} of the substrates S_i , where concentration c_{Si}^{in} is the concentration in the feed. A fully homogeneous system (mass ρ) is also assumed. The volume of the reactor is V_R . The reactor is as shown in the figure made up of an interaction of three (thermodynamic) phases. The gas phase transfers to liquid phase and then to bio-phase. Thus, the cells cannot absorb oxygen directly from the gas bubbles.

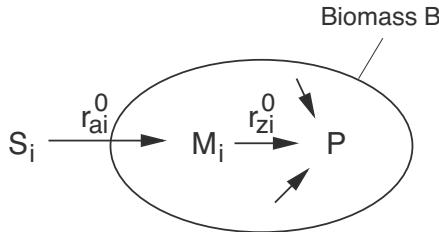


FIGURE 3.9: Cellular network to describe a part of metabolism.

A structured but not segregated approach should be used to deduct the equations. The biomass is described by an average cell as shown in Figure 3.9. Metabolism is shown by a small network in the cell, but can have any other structure. The network describes uptake of substrates S_i and further metabolism. The uptake rates of the substrates are r_{ai}^0 ; rates r_{zi}^0 describe the synthesis of a metabolite P from metabolites M_i . Component P is a main compartment of the cell, for example, proteins. Table 3.2 summarizes the state variables and input variables.

A mass balance equation

$$\dot{X} = J + Q \quad (3.55)$$

describes the change of an extensive state variable X as depending on mass and energy flow. One differentiates between flows J over the system boundaries and flows Q that arise from production/usage within the system boundaries. In the following, equations for all relevant masses/mol numbers of the substrates (extensive state variables) are derived.

Oxygen in the gas phase

$$\dot{n}_{O_2} = q_G (c_{O_2}^{in} - c_{O_2}) - k_L A (c_{FO_2}^* - c_{FO_2}) . \quad (3.56)$$

Here the mass transfer from gas to liquid is proportional to the concentration difference in the liquid and interface $c_{FO_2}^*$. Henry's law applies which determines the gas concentration at the interface between gas and liquid through partial pressure in the gas: $c_{FO_2}^* = p_{O_2}/H$ where H is the Henry coefficient.

TABLE 3.2: Overview of the used variables and units for the reactor system.

Gas speed in reactor	$[q_G] = \frac{l}{h}$	Oxygen conc. gas	$[c_{O_2}] = mol/l$
Mass transfer	$[k_L] = \frac{l}{m^2 h}$	Interface area	$[A] = m^2$
Feed i	$[q_{in,i}] = \frac{l}{h}$	Substrate conc.	$[c_S] = g/l$
Reactor out flux	$[q_{out}] = \frac{l}{h}$	Metabolite mass M	$[m_M] = mol$
Liquid	$[m_R] = g$	Metabolite conc.	$[c_M] = \frac{mol}{gDM}$
Reactor volume	$[V_R] = l$	Metabolite mass P	$[m_p] = mol$
Biomass	$[m_B] = gDM$	Metabolite conc.	$[c_p] = \frac{mol}{gDM}$
Biomass conc.	$[c_B] = g/l$	Rates	$[r_{ai}^0] = \frac{mol}{h}$
Substrate	$[m_S] = g$		$[r_{ti}^0] = \frac{mol}{h}$

By replacing the partial pressure by volume fraction v_{O_2} and total pressure by $p v_{O_2} = p v_{O_2}$ and then reformulating the equation for the gas volume fraction, one obtains:

$$\dot{v}_{O_2} = \frac{q_G}{V_G} (v_{O_2}^{in} - v_{O_2}) - k_L \frac{A}{V_G} \left(\frac{p v_{O_2}}{H} - c_{FO_2} \right) V_N. \quad (3.57)$$

The mass transfer number $k_L a = k_L A/V_F$ of the liquid can be determined in an experimental approach. One obtains:

$$\dot{v}_{O_2} = \frac{q_G}{V_G} (v_{O_2}^{in} - v_{O_2}) - k_L a \frac{V_F}{V_G} \left(\frac{p v_{O_2}}{H} - c_{FO_2} \right) V_N. \quad (3.58)$$

The volume fraction is measurable by an exhaust gas measuring device. As volume V_G can hardly be determined, it is replaced by $\epsilon_G = V_G/(V_G + V_R)$.

Volume of the bio-reactor

For the reactor we sum up all flows over the systems boarder:

$$\dot{m}_R = \sum q_{in,i} \rho - q_{out} \rho \longrightarrow \dot{V}_R = \sum q_{in,i} - q_{out}. \quad (3.59)$$

Density ρ is assumed to be constant.

Oxygen in the liquid phase

Oxygen is transferred with the same rate as above in the liquid phase and consumed by organisms with rate $r_{O_2}^0$. Oxygen in the feed must be also considered.

$$\dot{n}_{FO_2} = k_L A (c_{FO_2}^* - c_{FO_2}) - r_{O_2}^0 + q_{in} c_{FO}^{in} - q_{out} c_{FO_2}. \quad (3.60)$$

The same variables as above are used and re-formulated for the concentration; one obtains:

$$\dot{c}_{FO_2} = k_L a \left(\frac{p v_{O_2}}{H} - c_{FO_2} \right) - \frac{r_{O_2}^0}{V_R} + q_{in} (c_{FO_2}^{in} - c_{FO_2}). \quad (3.61)$$

Substrate in liquid phase

For the substrates in the liquid phase all transport over the systems boundaries as well as uptake/production must be taken into account. In the following equation, transport, that is uptake from the medium, is considered; therefore the respective term has a minus sign:

$$\dot{m}_{Si} = q_{in,i} c_{Si}^{in} - q_{out} c_{Si} - r_{ai}^0 mg_i. \quad (3.62)$$

The molecular weight mg_i is used because the rates r^0 are given in mole, while mass on the left side is often given in gram. To calculate the substrate concentration from the left side, the differential must be reformulated.

$$\dot{m}_{Si} = (V_R \dot{c}_{Si}) = V_R \dot{c}_{Si} + \dot{V}_R c_{Si}. \quad (3.63)$$

By setting the terms on the right side of the equation, one obtains:

$$\begin{aligned} V_R \dot{c}_{Si} + \dot{V}_R c_{Si} &= q_{in,i} c_{Si}^{in} - q_{out} c_{Si} - r_{ai}^0 mg_i \\ \dot{c}_{Si} &= \frac{1}{V_R} \left[q_{in,i} c_{Si}^{in} - q_{out} c_{Si} - r_{ai}^0 mg_i - \left(\sum q_{in,i} - q_{out} \right) c_{Si} \right] \\ \rightarrow \dot{c}_{Si} &= \frac{q_{in,i}}{V_R} c_{Si}^{in} - \frac{\sum q_{in,i}}{V_R} c_S - \frac{r_{ai}^0 mg_i}{V_R}. \end{aligned} \quad (3.64)$$

For one single substrate in the feed and in the medium the equation is simplified as:

$$\dot{c}_S = \frac{q_{in}}{V_R} (c_S^{in} - c_S) - \frac{r_a^0 mg}{V_R}. \quad (3.65)$$

Biomass

$$\dot{m}_B = \mu m_B - q_{out} c_B, \quad (3.66)$$

where μ is the specific growth rate that has to be described more in detail later. Analogous to above one obtains:

$$\begin{aligned}\dot{m}_B &= (V_R \dot{c}_B) = V_R \dot{c}_B + \dot{V}_R c_B \\ \dot{c}_B &= \frac{1}{V_R} \left[\mu m_B - q_{out} c_B - \left(\sum q_{in,i} - q_{out} \right) c_B \right] \\ \rightarrow \quad \dot{c}_B &= \mu c_B - \frac{\sum q_{in,i}}{V_R} c_B.\end{aligned}\tag{3.67}$$

Intracellular components

For intracellular components it is important to consider that the mass flow of components that leaves the reactor is proportional to the flow of outflowing biomass. The mass balance for components M_i and P are as follows:

$$\dot{m}_{Mi} = r_{ai}^0 - r_{zi}^0 - \underbrace{q_{out} c_B}_{\text{biomass}} c_{Mi}\tag{3.68}$$

$$\dot{m}_P = \sum r_{zi}^0 - q_{out} c_B c_P\tag{3.69}$$

Analogous to above one obtains for concentrations c_P and c_{Mi} (here based on biomass):

$$\begin{aligned}\dot{m}_P &= (m_B \dot{c}_P) = m_B \dot{c}_P + \dot{m}_x c_P \\ \dot{c}_P &= \frac{1}{m_B} \left[\sum r_{zi}^0 - q_{out} c_B c_P - (\mu m_B - q_{out} c_B) c_P \right] \\ \rightarrow \quad \dot{c}_P &= \frac{\sum r_{zi}^0}{m_B} - \mu c_P\end{aligned}\tag{3.70}$$

$$\text{analogous } \dot{c}_{Mi} = \frac{r_{ai}^0 - r_{zi}^0}{m_B} - \mu c_{Mi}.\tag{3.71}$$

From the last two equations an important relationship becomes obvious. The specific growth rate appears in each equation as a dilution term for the intra cellular components. This means that the components dilute when the cells are growing and if no new synthesis takes place. With each cell division the daughter cell receives half the molecules. This is an important fact considering proteins and other macromolecules: They do not have to be especially degraded which means that less energy has to be spent by the cell.

3.4.1 Relationship between specific growth rate μ and rates r_{ai}^0

In the equations above, the growth of the biomass has only been described by the specific growth rate μ . However, as the specific growth rate describes

the change of the biomass, there is a relationship between the uptake rate r_a and the specific growth rate. For determining this relationship, biomass is first considered, where it is assumed that all components make up the total biomass as in the figure shown above.

$$m_B = \sum m_{Mi} mg_i + m_P mg_P. \quad (3.72)$$

By taking the differential and putting in the terms from above, one obtains:

$$\begin{aligned} \dot{m}_B &= \sum \dot{m}_{Mi} mg_i + \dot{m}_{Pp} mg_P \\ &= \sum (r_{ai}^0 - r_{zi}^0) mg_i - q_{out} c_B \sum c_{Mi} mg_i + \\ &\quad \sum r_{zi}^0 mg_p - q_{out} c_B c_P mg_P. \end{aligned} \quad (3.73)$$

A simple stoichiometry is assumed for the reaction network above. Thus, the molecular weights should be equal for all metabolites: $mg_i = mg_P$. Consequently, one obtains in the following steps:

$$\begin{aligned} \dot{m}_B &= \sum r_{ai}^0 mg_i - q_{out} c_B \underbrace{\left(\sum c_{Mi} + c_P \right) mg_P}_1 \\ \longrightarrow \quad \dot{m}_B &= \sum r_{ai}^0 mg_i - q_{out} c_B. \end{aligned} \quad (3.74)$$

The bracket applies because concentration times molecular weight presents the individual mass fraction that adds up to 1. A comparison with above results in

$$\mu = \frac{\sum r_{ai}^0 mg_i}{m_B} \equiv \text{sum of all specific uptake rates.} \quad (3.75)$$

From the equation it becomes clear that all uptake and production rates contribute to a change of biomass. In the last equation, the rates are based on the biomass. One obtains for the rates:

$$\begin{aligned} r_{ai} &= \frac{r_{ai}^0}{m_B}, \quad r_{zi} = \frac{r_{zi}^0}{m_B} \quad \text{and therefore for the growth rate} \\ \mu &= \sum r_{ai} mg_i. \end{aligned} \quad (3.76)$$

The equations for intra cellular variable P and substrate S (only one influx) are simplified as:

$$\dot{c}_P = \sum r_{zi} - \mu c_P \quad (3.77)$$

$$\begin{aligned} \dot{c}_S &= \frac{q_{in}}{V_R} (c_S^{in} - c_S) - \frac{r_a m_B mg}{V_R} \\ &= \frac{q_{in}}{V_R} (c_S^{in} - c_S) - r_a mg c_B. \end{aligned} \quad (3.78)$$

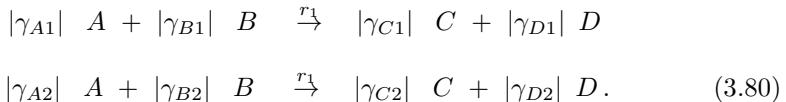
Normally not all transport/efflux systems are modeled in detail. This would require great effort, which is not possible to measure. Thus, the following reasonable approach can be used:

$$\mu = Y r_a mg \quad (3.79)$$

where Y is a rate of yield with the unit gTM/g . A typical value that one finds for the bacterium *E. coli* during growth of glucose is $Y_{Glc} = 0.5 \frac{gDM}{g}$. However, there are of course different possibilities to relate the growth rate to measured values, for example a relationship with intracellular components like protein or RNA content is possible as well.

3.4.2 Intracellular reaction networks

As already described above, biochemical reactions that take place in a cell with stoichiometric coefficients γ_{ij} and rate r_j can be written as:



As we have already seen, the reaction rate that appears in this equation is also used in the mass balance equations and the corresponding differential equations for the concentrations. If we look now for component A we see that in all reactions where A is involved, the respective rate has to be multiplied by the stoichiometric coefficient and has to be summed up (note that γ_{A1} , etc. are negative):

$$\begin{aligned} \dot{c}_A &= \gamma_{A1} r_1 + \gamma_{A2} r_2 - \mu c_A \\ \dot{c}_B &= \gamma_{B1} r_1 + \gamma_{B2} r_2 - \mu c_B \\ &\dots \end{aligned} \quad (3.81)$$

We see that the same coefficients appear in the rows as in the columns of matrix N^T . Therefore, matrix N itself can be used. To summarize the findings the differential equations for the intracellular components are:

$$\boxed{\dot{\underline{c}} = N \underline{r} - \mu \underline{c},} \quad (3.82)$$

with the specific growth rate μ . By putting in kinetic expressions in the rates $\underline{r}(\underline{c})$ in the Equation (3.82), then one obtains a nonlinear differential equation system in the form:

$$\boxed{\dot{\underline{c}} = f(\underline{c}).} \quad (3.83)$$

In some applications we are interested in subnetworks with high fluxes.

In this case, the dilution term is small in comparison with the fluxes and the equation simplifies:

$$\dot{c} = N r(c) = f'(c). \quad (3.84)$$

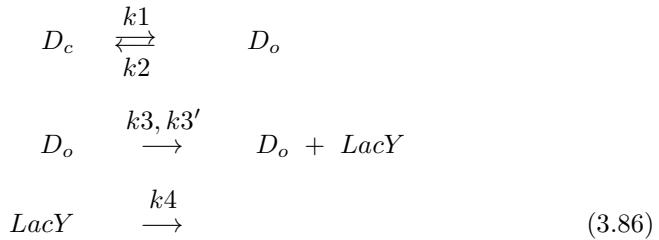
The steady state is a special case. In this case, there is no temporal change of the state variables anymore. This means that all reactions that produce the metabolite and those that use it are in equilibrium. This is written as:

$$0 = N r. \quad (3.85)$$

This equation is important for the analysis of the networks. The stoichiometric matrix N can be presented as a graph, where the metabolites are shown as nodes and the reactions as edges.

EXAMPLE 3.7 Gene expression of the lactose operon — autocatalytic loop.

The system introduced above shall now be analyzed deterministically. The same reaction equations as above apply; however, we omit the regulatory influences in the stoichiometry and write down the rate laws afterwards:



whereby only the deterministic reaction parameter k determines the speed of the reaction.

The following are three equations for the three metabolites (the term for the dilution is not considered for the promoter conformations, since the reactions occur very quickly; furthermore in k_4 the dilution for the protein is included):

$$\begin{aligned} \dot{D}_c &= -r_1 + r_2 = -k_1 D_c LacY^2 + k_2 D_o \\ \dot{D}_o &= r_1 - r_2 = k_1 D_z - k_2 D_o \\ \dot{LacY} &= r_3 - r_4 = k'_3 + k_3 D_o - k_4 LacY. \end{aligned} \quad (3.87)$$

The Master equation stated above is only based on the number of particles per cell; now, the variables describe concentrations with the sizes mol/gDM or mol/l whereby it is referred onto the cell mass or the cell volume. If attention is paid to the stoichiometry of the reactions, which influences the sizes, the

following correlation between the deterministic sizes k and the stochastic sizes c can be determined [4]:

$$c = \frac{k}{(N_A V)^{K-1}} \prod (l_i)! \quad \text{with} \quad K = \sum l_i \quad (3.88)$$

whereby l_i are the coefficients of the metabolites which cause the reaction (in general all stoichiometric coefficients and coefficients of modifiers), N_A the Avogadro number and V the volume. Usually, the coefficients l_i are the stoichiometric coefficients of the substrates. One obtains the following conversions for the example above:

$$\begin{aligned} \text{substrate } D_c, \text{ effector } LacY: & \quad K_1 = 3; \quad c_1 = 2 \frac{k_1}{(N_A V)^2} \\ \text{substrate } D_o: & \quad K_2 = 1; \quad c_2 = k_2 \\ \text{effector } D_o: & \quad K_3 = 1; \quad c_3 = k_3 \\ \text{no substrate, no effector:} & \quad K'_3 = 0; \quad c'_3 = k'_3 N_A V \\ \text{substrate } LacY: & \quad K_4 = 1; \quad c_4 = k_4. \end{aligned} \quad (3.89)$$

For the analysis of the steady states, the left side must be equalized to zero and solved for the applicable components. One obtains for the promoter confirmation D_o :

$$D_o = \frac{k_1 D_t LacY^2}{k_1 LacY^2 + k_2} = \frac{D_t LacY^2}{LacY^2 + K_B} \quad (3.90)$$

with the binding constant $K_B = k_2/k_1$. For determining the steady state for the protein, the two rates $r_3(LacY)$ and $r_4(LacY)$ can be plotted against the concentration of the protein. With the parameters c from above and converted into k -values, one obtains Figure 3.10.

The arrows indicate that there are three equilibrium states. Left of the first equilibrium, the rate of synthesis is higher than the degradation, meaning that the amount of $LacY$ will increase. The same applies for the values right of the second equilibrium. Right of the first and right of the third equilibrium, the degradation rate is higher. The number of molecules will decrease. Using the deterministic simulation, depending on the start value, one ends either in the first or third equilibrium state.

Using the deterministic simulation, the result is clearly defined by the starting conditions. The observed bimodal behavior in the stochastic simulation is now manifested in two possible stable steady states for a defined range of the input as can be seen in Figure 3.10.

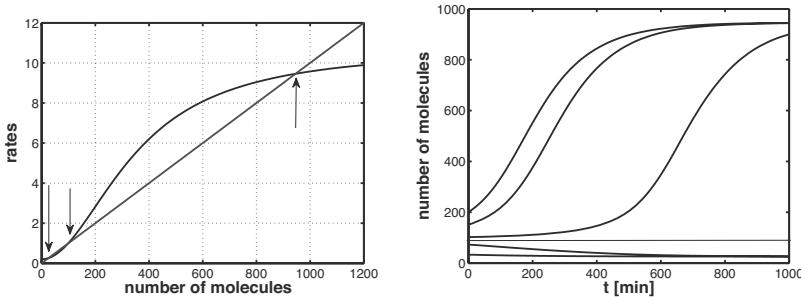


FIGURE 3.10: On the left: Plotting of the rates $r_3(\text{LacY})$ and $r_4(\text{LacY})$ as a function of LacY . The number of intersections represents the number of steady states. The illustration shows the particle number. Arrows indicate the individual steady states. On the right: Simulation study for different initial conditions for LacY .

3.5 Qualitative Modeling and Analysis

In qualitative modeling the focus is on interactions between components without describing these interactions in detail. Therefore, it is sufficient to make the interaction clear, e.g., A influences B . The interaction can be presented best in a graph [5]. In the graph G a disjunctive set of vertices V and edges E is presented. For example in Figure 3.11 there are 4 vertices and 6 edges. The set of vertices and edges for graph G is thus:

$$\begin{aligned} V &= \{1, 2, 3, 4\} \\ E &= \{\{1, 2\} \{1, 3\} \{2, 3\} \{1, 4\} \{2, 4\} \{3, 4\}\}. \end{aligned} \quad (3.91)$$

Graph **A** is complete as all vertices are linked together. Direct graphs are characterized by arrows (plot **B**).

Subsets can be formed from the node and edge set, which then presents only parts of the graph. This can be used to present unions, intersections and complements between two graphs. All neighbors of one vertex, i.e., all nodes that are linked to one vertex, also form a subset of the vertex set. The number of connections of one node, which is referred to as connectivity, can be called degree $d(v)$, when v is an element of V . By dividing this number by the number of nodes, one obtains the average degree. Later, it will be important to define “path” and “circuit.” Formally, a path as in the figure, plot **C** above, from node 1 to node 2 can be written as:

$$V = \{1, 2, 3, 4\}, \quad E = \{\{1, 4\} \{4, 2\}\}. \quad (3.92)$$

A circuit as in the figure, plot **D** above, can be described first by the definition

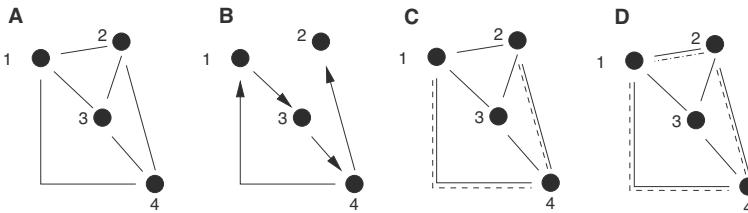


FIGURE 3.11: **A** Undirected graph with 4 nodes and 6 edges. **B** Directed graph. The direction of the interaction is shown by arrows. **C** Path from node 1 to node 2 via node 4. **D** The second connection between node 1 and node 2 forms a circuit.

of a path. A further edge $\{2, 1\}$ is then added. The length of the path can be deducted by counting the edges. The distance between two nodes is the length of the shortest path.

Graphs can be presented in other ways than those presented above. In the adjacency matrix nodes are compared and receive value 1, when the corresponding nodes are connected. The main diagonal has zeros, unless an edge connects a node to itself. In the incidence matrix the nodes are written in the rows and the edges in the columns. In an undirected graph the start node and the end node receive the value 1. In the example in Figure 3.11 (**A**) one obtains the following adjacency matrix A (a 4×4 Matrix) and incidence matrix I (a 4×6 matrix):

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}, \quad I = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}. \quad (3.93)$$

Metabolism networks and signal networks can be presented with the methods above as in Figure 3.12. However, problems can occur as in plot **B**. Enzymes react on enzymatic reactions in lactose degradation. This is shown as an edge that points to an edge, which cannot be presented properly in the graph. The tripartite graph is a particular illustration [6] (plot **C**). Interactions like enzymatic conversion and signal transmission are presented as nodes. The three disjunctive node sets are now connected by edges, where an edge can never be between nodes of the same set. A tripartite graph can illustrate metabolic processes that are linked to signaling processes.

Qualitative modeling of biochemical networks is based on the analysis of graphs. Depending on which interactions are illustrated in the graph, nodes and edges can have different meanings. The following networks are interesting in systems biology.

Metabolism network: describes the enzymatic conversions of metabolites by enzymes.

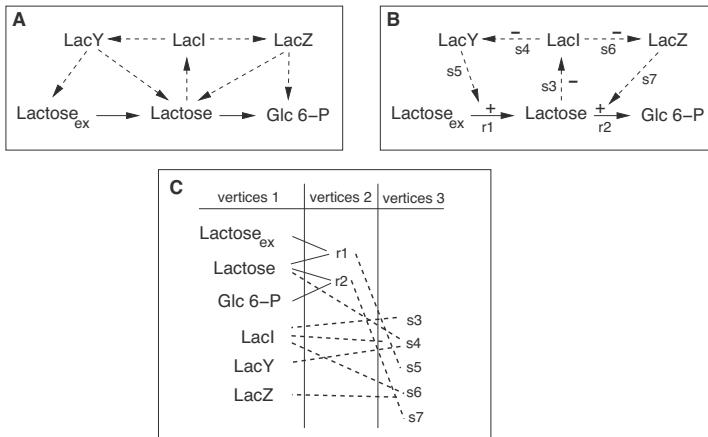


FIGURE 3.12: **A** Presentation of lactose metabolism as a directed graph. The edges indicate which component interacts with others. **B** Presentation of metabolism where the network is presented in more detail. **C** Tripartite graph. Nodes in a set are not connected (solid lines indicate metabolism, dashed lines indicate signaling).

Protein-protein interaction: describes the process and transmission of information in signal transduction networks.

Transcription network: describes the interaction of transcription factors with their DNA-binding site.

In metabolic networks nodes are metabolites and edges represent biochemical reactions or transformations. Directed graphs can be used depending on whether information on the reversibility of reactions exists. If a reaction is irreversible, the arrow is shown in one direction. Graphs can be used to determine how many connections/links a node has. Values can be assigned to the edges to express the rate at which the metabolite is being transformed.

Signal and transcription networks [7] differ in that the edges in the graphs do not represent mass flow or reaction rates, but signaling interactions between two components. One component influences the other by activation or inhibition of the other component. Such a graph is called interaction graph. For very large networks detailed kinetic information does not exist, so that the interactions are assigned with positive or negative (+, -) signs and are thus weighed. In interaction graphs, nodes represent components such as, e.g., receptors, ligands, effectors, genes or transcription factors.

EXAMPLE 3.8 Simple interaction graph.

Figure 3.13 shows a simple diagram. There is a directed graph and incidence matrix **I** is shown in an extended form. It is assigned to the value -1 when

the components have been ‘used,’ which corresponds to the starting point of the edge, and the value 1 if a component is ‘formed,’ which corresponds to the arrow of the edges. In this case the incidence matrix reads:

$$I = \begin{pmatrix} -1 & 0 \\ 1 & 1 \\ 0 & -1 \end{pmatrix} \quad (3.94)$$

where the rows belong to metabolites A , B and C , while the first column represents the first interaction and the second one the second interaction. Often the interaction graph is not sufficient to describe all interactions. If it is known that two components are necessary to activate component B as in Figure 3.13 on the right side, then a hypergraph must be applied. It has two starting points and only one arrow. The interaction is characterized by AND: A AND (NOT C) are needed to activate B . Hypergraphs are not considered here in detail.

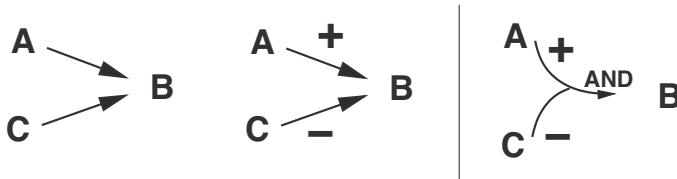


FIGURE 3.13: Left: Example of interaction graphs. On the right side: Logical interaction hypergraph.

3.5.1 Cycles, loops and elementary modes

An important characteristic of the incidence matrix can be obtained by calculating the null space (see Appendix). A null space vector \underline{c} fulfills the condition:

$$I \underline{c} = \underline{0}. \quad (3.95)$$

The vector \underline{c} is a conservation relation and is thus a particular path through the network. Calculating the null space serves to determine feedback loop and, thus, circuits in the graph. However, for some applications, a more strict definition of a feedback loop is required. To characterize the feedback loop the entries in vector \underline{c} must possess only elements larger or equal to zero; moreover, it is required that the loop cannot be decomposed further. To check the second property one defines the index set of the elements of the loop. Then there should not be a subset of this index set that already is a loop. This is illustrated in the Figure 3.14 with a network with three nodes and five links (**A**).

In **B**, index set $P = \{2, 3\}$ and in **C** index set $P = \{1, 2, 5\}$ the black arrows

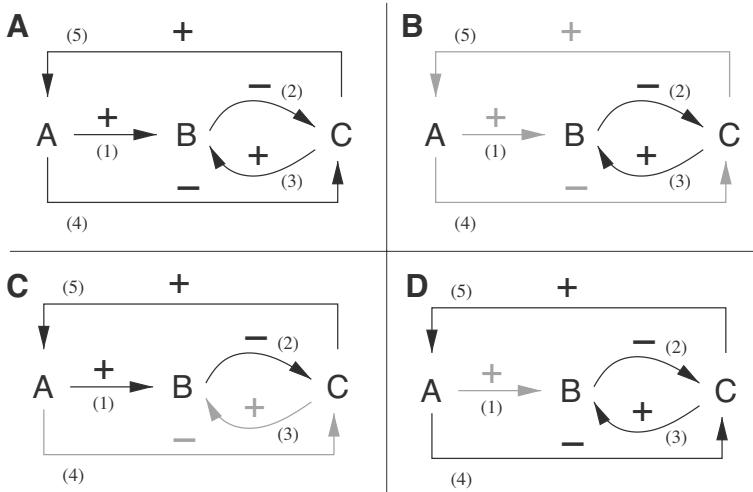


FIGURE 3.14: **A** Network of three nodes and five links. **B** and **C** Feedback loops. **D** No feedback loop.

indicate the feedback loops. In **D** the index set of the loop is $P = \{2, 3, 4, 5\}$. However, there is a subset $P = \{2, 3\}$ that is already a feedback loop; therefore in **D** the loop is not a feedback loop.

Closely related to the definition of a feedback loop is the definition of a path through the network. For the analysis of flux distributions in biochemical networks steady state conditions are often analyzed and a question that is treated is the number of pathways that guarantee a steady state of the system. In contrast to the definition of loops from above, also uptake/input fluxes and drain/output fluxes are taken into account. As above, it is required that a path cannot be decomposed in smaller paths through the network. Such a path that cannot be decomposed is called an elementary mode. Figure 3.15 shows all elementary modes for the given network. If the direction of the fluxes is known (one can assume that the reaction is irreversible) the number of elementary modes is reduced.

3.6 Modeling on the Level of Single Cells — the Population Balance

Until now only the population was considered as a whole, where each cell is seen as an “average” cell. The focus was on the temporal changes of components within a cell. Now there are further characteristics that might be

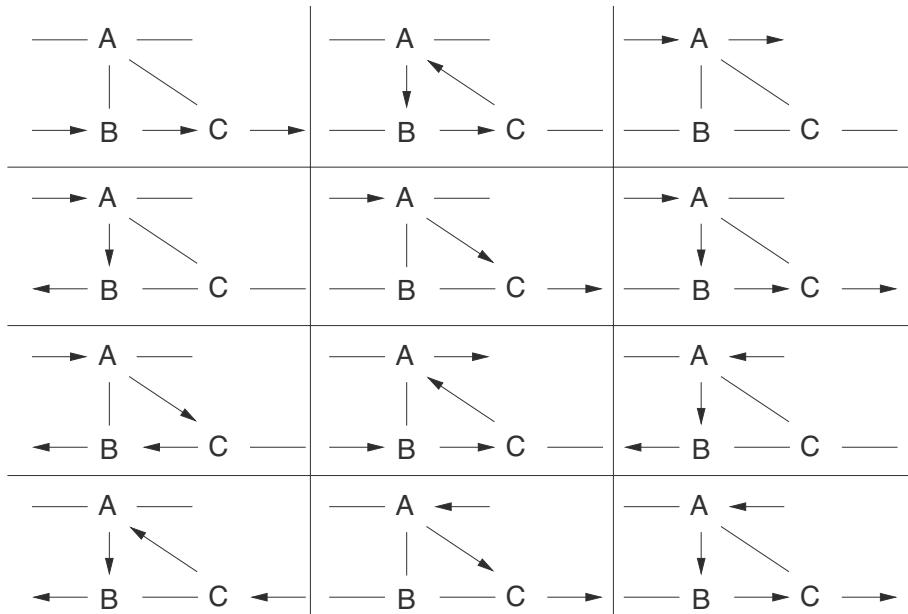


FIGURE 3.15: Metabolic network of three metabolites and seven reactions. Shown are all elementary modes. Note that all links could work in both directions.

important. For example, size and age distribution changes with time. Models that can represent these characteristics are called segregated and statistical models, as they describe the temporal changes of a distribution. The term population balance often used in literature can be misleading in this case as populations were considered until now, but the focus will now be on individual cells.

In model equations characteristic descriptor variables are considered and analyzed. When size and age distributions are considered, the process of cell division plays an important role as the cells fall out from one partition into another partition as shown in Figure 3.16. Therefore, it is necessary to write functions for these transitions.

The total number of cells is determined for discrete distributions by summing up the corresponding variables. In modeling it is assumed that the distribution is continuous. The integral is formed to obtain the total number of cells N :

$$N = \int_{VE} f(E, t) dE, \quad (3.96)$$

where function $f(E, t)$ is the number density function and variable E is the descriptor variable. Thus, the integral ranges over the whole area of the volume

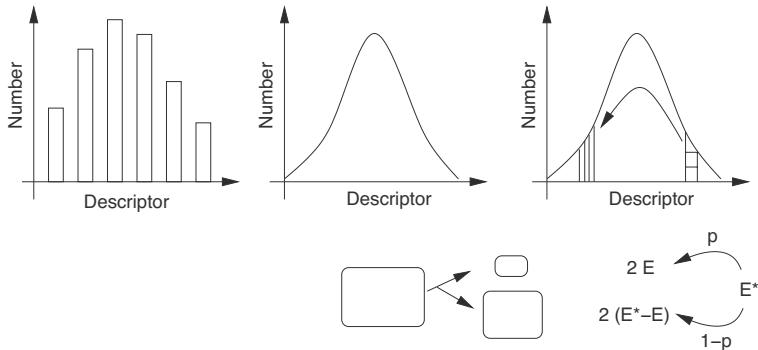


FIGURE 3.16: On the left side: Size distribution in discrete form. In the middle: Continuous distribution function. On the right side: Cells that can reach a certain size are probable to form into two smaller cells.

VE of the property space. Consequently, the function f has the unit dimension [<#cellsdescriptor)].

The following will be based on the presentation of Nielsen et al. [8] by an equation for the special case that a continuous reactor with the flow rate D is considered. The equation describes the number density function's temporal changes; in this case only one descriptor variable is considered:

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial E}(r(E,t) f(E,t)) = h(E,t) - D f(E,t). \quad (3.97)$$

The left hand side of the equation is a result of taking the derivative (Leibniz integral rule) of Equation (3.96). Function r describes f 's temporal changes by cellular processes, e.g., growth, and h is the value that results from cell division. Thus, function r has the dimension [descriptor/time], while h has the dimension [<#cellsdescriptor/time)]. Function h describes cell division process and is normally divided into two parts (h^+ und h^-). The first part describes the cells that fall in that segment after division, and the second part describes the cells that fall out from the segment. A partitioning function $p(E, E^*, t)$ is needed for the first part. It gives the probability that cells with characteristic E^* are divided into the two areas E and $E^* - E$. Also, time constant $b(E, t)$ is defined that expresses the breakage frequency. In order to determine all cells that finally have characteristic E , the sum of the characteristic space is formed and multiplied by 2, as there are 2 daughter cells.

$$h(E,t) = h^+ - h^- = 2 \int_{VE} b(E^*, t) p(E, E^*, t) f(E^*, t) dE^* - b(E, t) f(E, t). \quad (3.98)$$

The integro-differential equation is the basic equation of models on the level of single cells. Cell mass m is considered as the characteristic. The following

suggestions are made for functions $b(m)$ and $p(m, m^*)$, which describes size and mass distribution of baker's yeast (*S. cerevisiae*) [9]:

$$b(m) = \begin{cases} 0 & \text{for } m \leq m_t \\ \gamma e^{-\delta(m-m_d)^2} & \text{for } m_t < m < m_d \\ \gamma & \text{for } m \geq m_d \end{cases} \quad (3.99)$$

where m_t is the mass that has to be at least reached so that cell division is possible and m_d is the maximal mass. Parameter γ is the actual time constant and δ can change the progression of the e-function. Figure 3.17 shows the course of b for given parameter values.

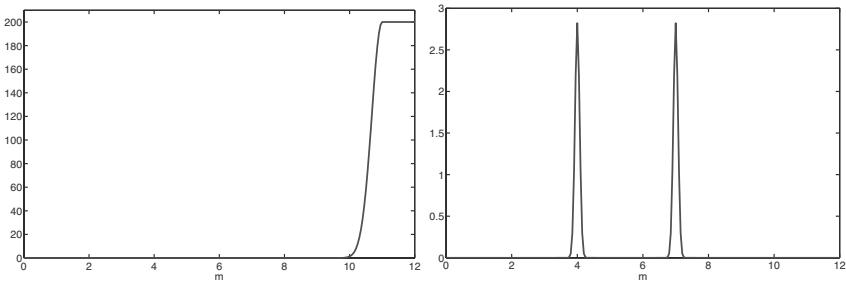


FIGURE 3.17: On the left side: Function $b(m)$ with the parameters $m_t = 7$, $m_d = 11$, $\gamma = 200$, $\delta = 5$. On the right side: Function $p(m, m^*)$ for fixed $m^* = 11$ with further parameters $\lambda = 2.82$, $\beta = 100$.

Function $p(m, m^*)$ has to assure that daughter and stem cells have a different size. This can be realized by the following function:

$$\begin{aligned} p(m, m^*) &= \lambda(e^{-\beta(m-m_t)^2} + e^{-\beta(m-m^*+m_t)^2}) \\ &\quad \text{for } m^* > m \wedge m^* > m_t \\ &= 0 \\ &\quad \text{for } m^* \leq m \wedge m^* \leq m_t \end{aligned} \quad (3.100)$$

where m^* is mass of the dividing stem cell. The right side of Figure 3.17 shows p 's course for the given parameter values. The left plot shows that from a threshold value m_t on, the probability for cell division increases. Once the value m_d has been exceeded, b stays constant. The right plot shows the division function. When a cell is divided, a large mother cell and a smaller daughter cell evolve. This can be seen in the two peaks. If the cell mass is smaller than m^* , there will be no cell division. Thus, the following is necessary for function p :

$$\int_{VE} p(m, m^*) dm = 1. \quad (3.101)$$

3.7 Data Driven Modeling

Until now, all approaches assumed that there is a connection between two components (or the connection does not exist). This interaction can be explained quantitatively in detail according to the biological knowledge. In the easiest case, an arrow is enough to show that a connection exists, though, in many cases it is not known if a connection exists or not. Now, it is the aim of data driven modeling to provide methods that allow to determine the existing connections within the cell based on experimental data. This approach is connected to the top-down approach, which integrates data on different levels and derives a network structure. The approaches again differ to each other in the following Figure 3.18.

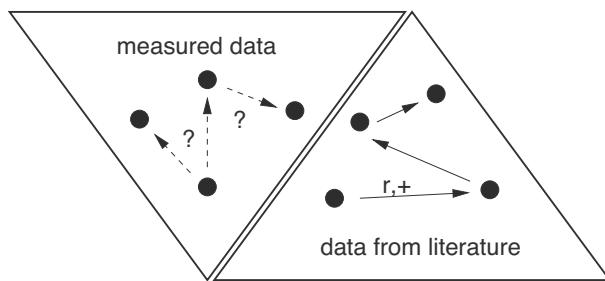


FIGURE 3.18: Data driven approach (left) tries to determine if there is a connection between the components of the network. The alternative approach (right) is usually based on data from literature, which lead to a first structure of the model.

The following methods are proposed in literature to reconstruct informations about networks (often also called reverse engineering):

- Usually, in a first step the large amount of data is reduced in dimension. If data for gene expression is available, the amount of data is already large ($\gg 4000$ entries for the bacterium *E. coli*) and one tries to group components with similar (dynamic and/or steady state) behavior. If time course data and data that were collected during different stimulations, nutritional limitations, stress situations or various other different environmental conditions are available, a cluster analysis can be used for determining which components show a comparable behavior and therefore can be grouped. If especially time course data are available, one cannot necessarily directly infer on a close relationship from a similar behavior of the components. Here, usually a correlation analysis comes into consideration. A further method to reduce the dimension of the data space is the Principal Component Analysis (PCA). Here, the characteristic is the variance of the objects under investigation. Objects

that are for example state variables can be compared by their variance. Those state variables that show a large variance are grouped together.

- If only a reduced data set is available, the challenge is to determine if there is an interaction between two components or not. In the easiest case, an algorithm assumes that there is a connection between all components and identifies if the data can be described if a connection is left out; if it is assumed vice versa that there first is no connection between the components, it is identified by the addition of links, if a connection is essential. To do so, regression techniques or simple linear dynamical models can be used. Furthermore, Bayesian networks can be set up and the interconnections are derived from statistical properties.
-

3.8 Thermodynamics

Until now, we have not considered thermodynamic aspects. However, they play an important role in many processes, since energy conversion is happening during biochemical reactions. The energy balance of a cell underlies, like all systems, the laws of thermodynamics, which shall briefly be explained first. Afterwards, reaction technical aspects shall be considered.

3.8.1 Fundamentals

In the laws of thermodynamics, fundamental characteristics are defined [10]:

1. Law of thermodynamics: During every physical or chemical change (reaction, etc.), the total amount of energy “in the universe” remains constant. Therefore, energy is only converted to other forms.

2. Law of thermodynamics: In all natural processes, the entropy (measure for the chaos, warmth) “in the universe” increases.

The first law of thermodynamics applies in the creation of mass balance equations. These were discussed above and consider that the mass of the system, the cell, can only change by substrate flow across the cell membrane or cell wall, not by internal production. Therefore, for all metabolites it can be assumed that the Kirchhoff's first law applies for the equilibrium state: The sum of all reactions that use or produce a component is zero (here, the dilution term due to growth is counted as rate).

The second law states that different forms of energy can not be converted into each other arbitrarily. This mainly plays a role in substance conversions in cycles (metabolic cycles), in which the Kirchhoff's first law is fulfilled, but in which the second law may not be fulfilled, as will be shown below.

Basis for the equations derived below is the concentration of a substance or the number of molecules n of a substance X . It is given in mol and 1 mole is equivalent to $\equiv 6.023 \cdot 10^{23}$ particles. For the concentration of a substance, the biomass or the volume of the cell can be used:

$$c = \frac{n}{V} = \frac{m}{mg \cdot V} = \frac{\rho}{mg} \quad (3.102)$$

with the molecular weight mg with the size $\frac{g}{mol}$ and the density ρ . Therefore, the concentration equals a density weighted with the molecular weight.

State variables describe the (dynamic) behavior of systems using mass balance equations. Next to the state variables of the last chapter, new ones are added. Examples are

Pressure	p	is changed with mechanical work
Temperature	T	is changed with energy input
Volume	V	is changed with inflows in gases with mechanical work
Number of molecules	n	is changed during reactions

Until now, state variables were coupled in differential equations, though there are several other connections which couple the state variables via so-called state functions. It must be kept in mind that these state functions only apply in thermodynamic equilibrium states. The best known state equation is the equation for ideal gases. It connects four state variables. Therefore three variables are sufficient for a complete description since the last variable can be calculated with the first three. The following correlation applies:

$$\frac{p \cdot V}{T} = R \cdot n \cdot mg = \mathbb{R} \cdot n \quad (3.103)$$

with the specific gas constant R and the universal gas constant \mathbb{R} . Under standard conditions ($T = 273.15 \text{ K}$, $p = 101.33 \text{ kPa} = 1 \text{ bar}$, $V = 22.41 \text{ l}$) it has the value $8.314 \frac{J}{mol \cdot K}$. In case $n = const.$, the following correlations apply:

Isobare	Isochore	Isotherme
$p = const.$	$V = const.$	$T = const.$
$\frac{V_1}{V_2} = \frac{T_1}{T_2}$	$\frac{p_1}{p_2} = \frac{T_1}{T_2}$	$\frac{p_1}{p_2} = \frac{V_2}{V_1}$

Mass balance equations that have been looked at so far have only considered the mass flow but not energy flow. In an equilibrium state, the inner energy U correlates to all other extensive state variables and therefore totally characterizes the system looked at. Therefore, the inner energy is called potential:

$$U = U(S, V, n_k) \quad (3.104)$$

whereby S describes the entropy of the system. Via transformation, other thermodynamic potentials can be derived from the inner energy U . The most important ones are enthalpy H and the Gibbs energy G . The enthalpy H describes the thermal-energetic balance of the system. It applies:

$$H = U + pV. \quad (3.105)$$

The enthalpy describes the total amount of thermal energy available. The Gibbs energy G also is a potential and describes the amount of energy that can do work and which is a function of temperature, pressure and the composition of the system. The equation is:

$$G(T, p, n_k) = H - TS = U + pV - TS. \quad (3.106)$$

Since in many cases it is assumed that temperature and pressure are constant inside the cell, the correlation simplifies. The first law describes the change of inner energy. This can be happen via thermal input δQ , mechanical work δW and chemical reaction work δW_c and one obtains:

$$dU = \delta Q + \delta W + \delta W_c = TdS - pdV + \sum_i \mu_i dn_i, \quad (3.107)$$

whereby the connection to entropy S is made for the second part of the equation, the mechanical work is described by the (negative) change of volume V and the chemical reaction is characterized by the change in the number of molecules; the chemical potential μ_i of the component i therefore describes the change in inner energy if the number of molecules changes. From Equation (3.106) the following equation can be determined via total differentiation and inserting Equation(3.107). It is called Gibbs fundamental equation:

$$dG = -SdT + Vdp + \sum_k \mu_k dn_k. \quad (3.108)$$

For constant temperature and pressure, the equation simplifies to:

$$dG = \sum_k \mu_k dn_k. \quad (3.109)$$

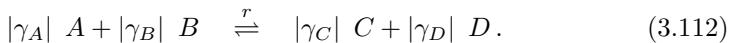
So, the Gibbs energy changes, as long as the reaction happens. Therefore, in an equilibrium, the Gibbs energy has reached an extremum (in a closed system), if the temperature and the pressure do not change. Now, one can show that this extreme represents a minimum. The chemical potential μ_i has a specific meaning here, since it indicates the direction in which the reaction actually occurs. The definition of the chemical potential μ_i arises from the equations above:

$$\mu_i = \left(\frac{\partial G}{\partial n_i} \right)_{T, p, n_j} = \left(\frac{\partial U}{\partial n_i} \right)_{S, V, n_j}, i \neq j. \quad (3.110)$$

Therefore, it describes the change of G/U through a change in the amount of the components. With help of the chemical potential, many molar state variables can be determined. The following approach is made for the chemical potential in dependence of the concentration of a component:

$$\mu = \mu^0 + \mathbb{R} T \ln\{c\}, \quad (3.111)$$

whereby μ^0 is a value at standard conditions and c is a concentration that is related to a reference value $c_{ref} = 1 \text{ mol}$. Therefore, the concentration c is without a dimension in the equation above. In biochemical reactions, the change in the Gibbs energy plays an important role. To make the change independent of the size of a system, it is based on the extent of reaction λ . For a reaction of a general kind:



The extent of reaction is defined:

$$d\lambda = \frac{dn_C}{\gamma_C} = \frac{dn_D}{\gamma_D} = \frac{dn_A}{\gamma_A} = \frac{dn_B}{\gamma_B}; \quad (3.113)$$

that is, the change of number of molecules divided by the stoichiometric coefficient is for all reaction partners the same. The derivation of the extent of reaction with respect to time $d\lambda/dt$ represents the speed of reaction r . With the extent of reaction, Equation (3.109) can be rewritten like this:

$$dG = \sum_k \mu_k \gamma_k d\lambda. \quad (3.114)$$

If the quotient $dG/d\lambda$ is formed and if it is defined as the change in the Gibbs energy, one gets:¹

$$\Delta G = \left(\frac{\partial G}{\partial \lambda} \right)_{T,p} = \sum_k \gamma_k \mu_k. \quad (3.115)$$

Based on the equation for the chemical potential, an equation for the Gibbs energy may be stated, if a (bio-)chemical reaction is looked at:

$$\Delta G = \Delta G^{0'} + \mathbb{R} T \sum_k \ln\{c_k^{\gamma_k}\}, \quad (3.116)$$

where $\Delta G^{0'}$ describes the reaction energy under normed conditions (the normed conditions ΔG^0 are chosen disadvantageously for biochemical conditions; therefore, usually a different normed condition applies, pH=7, and

¹ Actually a small letter should be chosen for the state variable related to the amount, like the molar Gibbs energy, to clearly differentiate to the potential state variable. The same applies for changes in these state variables, though the use of capital letters has been established in literature and is also used here. Attention must be paid to units: $[G]$ J, and of $[\Delta G]$ J/mol.

the variable is described using $\Delta G^{0'}$). The tailed expression describes the concentration related to the reference value. Since the same reference is taken for the substrates and products, it cancels out. The change of G shows quantitatively the difference of the system to the steady state. For (bio-)chemical reactions apply:

$\Delta G < 0$ exergonic, spontaneous reaction (the final state has less Gibbs energy than the starting state)

$\Delta G > 0$ endergonic, the reaction does not happen spontaneously .

All living things must gain energy via nutrition. The energy-delivering reactions are coupled with energy-consuming reactions. In the reaction equilibrium with the concentrations c_{Eq} , ΔG does not change anymore. The steady state constant K_{Eq} of the reaction can be determined as:

$$\begin{aligned} 0 &= \Delta G^{0'} + \mathbb{R} T \sum_k \ln\{c_{Eq}^{\gamma_k}\} \\ \longrightarrow \quad \Delta G^{0'} &= -\mathbb{R} T \ln K_{Eq}. \end{aligned} \quad (3.117)$$

Here we see that the equilibrium constant K_{Eq} is related to the Gibbs energy under standard conditions. This value can be obtained from databases or internet sources and therefore is an important parameter as we will see later on in enzyme kinetics.

3.8.2 Relation between ΔG and flux distributions

When looking at stationary flux distributions, it will be important to pay attention to avoid contradictions of the mathematical signs of the fluxes in Equation (3.115). This means that $-\Delta G$ and the stationary flux J_{st} must have the same mathematical sign: $-\Delta G J_{st} > 0$. Equation (3.115) can be set up for all reactions with help of a stoichiometric matrix N . One obtains:

$$\underline{\Delta G}^T = \underline{\mu}^T N \quad (3.118)$$

with the vector of all chemical potentials $\underline{\mu}$ and the stoichiometric matrix N .

EXAMPLE 3.9 Network with three reactions and one cycle.

The problem is now visualized in an example. In Figure 3.19 two flux distributions are shown, which each represent a steady state. The influx to A and the outflux from B , respectively C , are the same. Only the inner distribution of the fluxes differs. To realize the second scheme, the following requirements must apply, so the fluxes show the applicable direction, that is the respective ΔG is negative:

$$\begin{aligned} \text{Flux from A to B:} \quad \mu_A &> \mu_B \\ \text{Flux from B to C:} \quad \mu_B &> \mu_C \\ \text{Flux from C to A:} \quad \mu_C &> \mu_A. \end{aligned} \quad (3.119)$$

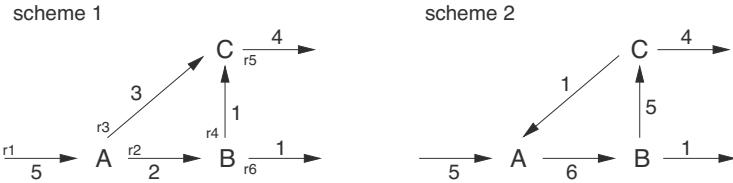


FIGURE 3.19: Network of the reaction with three components and six reactions: Two schemes with same in- and outflow but different internal flux distribution. Scheme 2 is not valid, even though the sum of all in- and outflows from all knots sum up to zero.

The last equation is a contradiction to the first two. The flux distribution in scheme 2 is not valid. For scheme 1, the following is determined:

$$\begin{aligned} \text{Flux from A to B: } & \mu_A > \mu_B \\ \text{Flux from B to C: } & \mu_B > \mu_C \\ \text{Flux from C to A: } & \mu_A > \mu_C, \end{aligned} \quad (3.120)$$

which does not lead to a contradiction, though it is crucial that no flux occurring in the cycles is allowed to be larger than zero. Conditions from above can be proved like shown in the following [11]. If the vector \underline{s} has the algebraic signs of a cycle in the network and the vector \underline{s}_J has the algebraic signs of the flux distribution, then the following must apply:

$$|\underline{s}^T \underline{s}_J| < \underline{s}^T \underline{s}. \quad (3.121)$$

Here, it is talked about orthogonality of the two sign vectors. This condition can also be interpreted like this: Sign vectors can be called orthogonal if at least one algebraic sign, but not all, is different. The network described above has one cycle containing the reactions $r_2, -r_3, r_4$ (the stoichiometric matrix is made up by the arrows following scheme 1). With the vectors according to scheme 1:

$$\underline{s} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \quad \underline{s}_J = \text{sign} \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad (3.122)$$

the condition stated above is fulfilled:

$$|\underline{s}^T \underline{s}_J| = 1 < \underline{s}^T \underline{s} = 3. \quad (3.123)$$

For the flux distribution from scheme 2 we get:

$$\underline{s}_J = \text{sign} \begin{pmatrix} 6 \\ -1 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \rightarrow |\underline{s}^T \underline{s}_J| = 3 = \underline{s}^T \underline{s} = 3, \quad (3.124)$$

and we see that the condition given above is not fulfilled. Note that if matrix \bar{N} contains only reactions inside the cell (transport reactions are not considered and the respective columns are deleted from N), null space \bar{K} of \bar{N} describes all cycles inside the system. A further important equation can be obtained when Equation (3.118) is extended by a multiplication with the null space: $\Delta G^T \bar{K} = 0$. This formula can be used for example to check if some of the reaction rates could have a value $r = 0$.

Transport against electric fields

So far, only the case that perturbations from a steady state lead to mass flow was looked at. In cellular systems, mass flow which is caused by electric currents, for example along a membrane, can occur. For reactions, which are influenced by an electric field, the following relationship for the Gibbs energy applies:

$$\Delta G = (-)z \mathbb{F} \Delta\Psi \quad (3.125)$$

with z , the charge of the ion, \mathbb{F} , the Faraday-constant $96.48 \frac{kJ}{Vmol}$ and $\Delta\Psi$, the electric trans-membrane potential. It has a positive value and represents the potential between the outer side and the inner side of the membrane: The outer side is charged positively whereas the inner side is charged negatively. The algebraic sign now depends on the direction of the transport. From this equation, it results that no transport against the field could happen without the concentration difference or coupling to another process (since ΔG is positive). The proton gradient along cell membranes plays an important role in the energy supply of cells. To keep up the gradient, protons are pumped to the outside against the gradient during various processes.

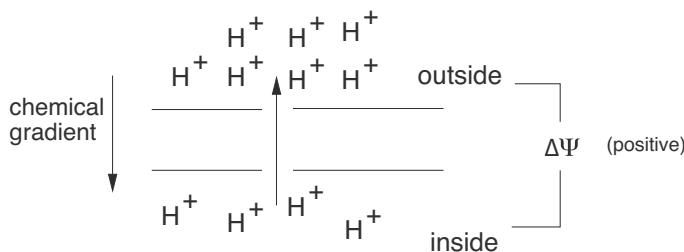


FIGURE 3.20: If protons are pumped towards the electrochemical potential, energy must be supplied from coupling with other processes.

Figure 3.20 shows the electrochemical gradient. The “power,” which is produced by the gradient, is referred to as a proton motive force. The following relation is used to calculate how much energy is needed — from coupling with another process — to transport a molecule against the force. From the equation above the following results with two concentrations from in- and

outside $c_{H_a^+}$, $c_{H_i^+}$ for ΔG_M (M stands for membrane):

$$\Delta G_M = \mathbb{R} T \ln \frac{c_{H_a^+}}{c_{H_i^+}} + \mathbb{F} \Delta \Psi. \quad (3.126)$$

If attention is paid to the definition of the pH,

$$pH = -\log c_{H^+} \quad (3.127)$$

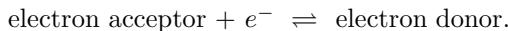
which describes the concentration of protons, a correlation between the natural logarithm and the 10-logarithm can be used to get:

$$\Delta G_M = -2.303 \mathbb{R} T \Delta pH + \mathbb{F} \Delta \Psi, \quad (3.128)$$

where the value for $\Delta pH = pH_a - pH_i$ results from the difference between outside and inside and therefore has a negative value.

Redox Reactions

A special class of reactions is the one where electrons are transferred. These are called redox reactions and can generally be written as:



Strongly reducing components serve as a source of electrons. Analogous to above, in this case, the electron affinity serves as a measure for the direction in which the reaction is going. The transfer of the electrons can occur in four different ways:

- Direct transfer of electrons
- Transfer as hydrogen: in a reaction $XH_2 \rightleftharpoons X + 2H^+ + 2e^-$ electrons are released by hydrogens
- Transfer in form of a hybrid-ion
- Direct reaction with oxygen

The standard reduction potential $E^{0'}$ of a reaction is determined under the same conditions as above and serves for calculating the reduction potential with other concentrations. The following correlation applies (Nernst equation):

$$E = E^{0'} + \frac{\mathbb{R} T}{n_E \mathbb{F}} \ln \frac{c_{EA}}{c_{ED}} \quad (3.129)$$

with the concentration of the electron acceptor c_{EA} and the concentration of the electron donor c_{ED} , as well as the number of transferred electrons n_E . The Gibbs energy G results in:

$$\Delta G = -n_E \mathbb{F} \Delta E. \quad (3.130)$$

The value for ΔE thereby results from the difference of the value of two partial reactions (usually, the reaction looked at can be split in two parts, for which the standard values are known).

EXAMPLE 3.10 *Transfer of a phosphate group onto glucose.*

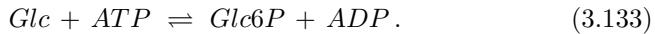
The determination of the Gibbs energy is explained for a reaction system with two substrates and two products:



Using Equation (3.116), the Gibbs energy can be determined from the concentrations and stoichiometric coefficients.

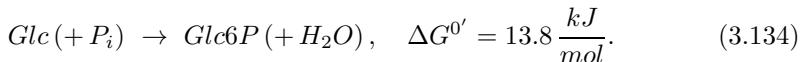
$$\Delta G = \Delta G^{0'} + \mathbb{R} T \ln \frac{c_C^{|\gamma_C|} c_D^{|\gamma_D|}}{c_A^{|\gamma_A|} c_B^{|\gamma_B|}} \quad (3.132)$$

with $\Delta G^{0'}$, the Gibbs energy under standard conditions ($\text{pH}=7$, $T=298$ K). Specifically, a reaction where ATP is needed is considered. The reaction is

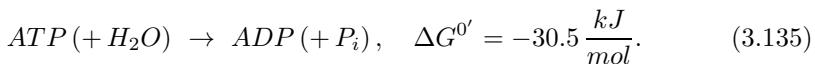


Now, the reaction can be split into two parts and the energy can be determined for each part.

Step 1.



Step 2.



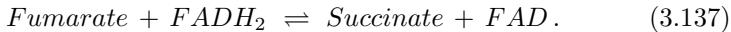
The sum is $\Delta G^0 = -16.7 \frac{\text{kJ}}{\text{mol}}$, so the reaction occurs under standard conditions using ATP (exergonic). If physiologic concentrations of the reaction partners involved are assumed (the following values can be used: ATP = 7.9 mM, ADP = 1 mM, Glc = Glc6P = 0.4 mM), one obtains:

$$\begin{aligned} \Delta G &= \Delta G^{0'} + \mathbb{R} T \ln \frac{c_{\text{ADP}} c_{\text{Glc6P}}}{c_{\text{ATP}} c_{\text{Glc}}} \\ &= -16.7 \frac{\text{kJ}}{\text{mol}} - 5.13 \frac{\text{kJ}}{\text{mol}} = -21.83 \frac{\text{kJ}}{\text{mol}} \end{aligned} \quad (3.136)$$

So, under physiological conditions, a higher value is obtained.

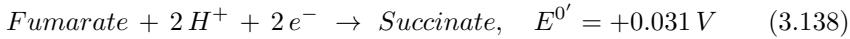
EXAMPLE 3.11 *Transfer of electrons (redox systems).*

Looking at the reaction

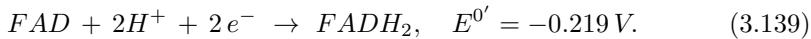


Again, the reaction can be split into two parts:

Step 1.



Step 2.



From the difference $E^{0'} = 0.25\text{ V}$, one obtains $\Delta G^{0'} = -48.3 \frac{\text{kJ}}{\text{mol}}$. This shows that the reaction occurs from fumarate to succinate under standard conditions.

EXAMPLE 3.12 *The reaction of the NADH-Dehydrogenase along the cell membrane.*

The NADH-Dehydrogenase transfers protons from NADH onto ubiquinone. At the same time, hydrogen is pumped to the outside against the concentration gradient, as can also be seen in Figure 3.21 [12].

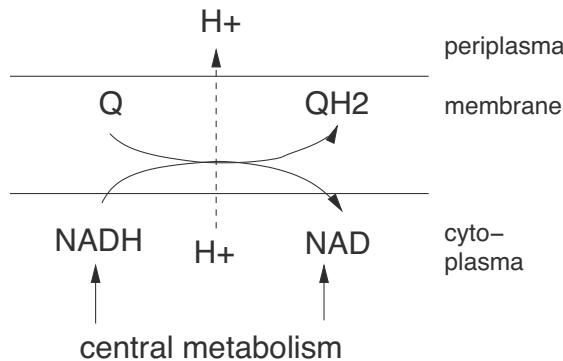


FIGURE 3.21: The NADH-Dehydrogenase-Reaction transfers protons from NADH onto ubiquinone Q. At the same time protons are pumped out of the cell.

NADH comes from the central metabolism, for example, from reactions of glycolysis and TCA. The reaction equation is:



whereby 2 electrons are transferred and 4 protons are pumped to the outer side of the cell (index o). The standard values for the reduction potentials of the two partial reactions can be determined the following way:

$$E_{Q/QH_2}^{0'} = 70 \text{ mV} \quad \text{and} \quad E_{NAD/NADH}^{0'} = -320 \text{ mV}.$$

One obtains for the partial reactions:

$$E_{Q/QH_2} = E_{Q/QH_2}^{0'} + \frac{\mathbb{R} T}{2F} \ln \frac{c_Q}{c_{QH_2}} \quad (3.141)$$

$$E_{NAD/NADH} = E_{NAD/NADH}^{0'} + \frac{\mathbb{R} T}{2F} \ln \frac{c_{NAD}}{c_{NADH}}. \quad (3.142)$$

The second partial reaction goes into the other direction:

$$\begin{aligned} \Delta E &= E_{Q/QH_2}^{0'} - E_{NAD/NADH}^{0'} + \frac{\mathbb{R} T}{2F} \left(\ln \frac{c_Q}{c_{QH_2}} - \ln \frac{c_{NAD}}{c_{NADH}} \right) \\ &= \Delta E^{0'} + \frac{\mathbb{R} T}{2F} \ln \frac{c_Q \cdot c_{NADH}}{c_{QH_2} \cdot c_{NAD}}. \end{aligned} \quad (3.143)$$

Therefore, the free reaction energy sums up to:

$$\Delta G = -2F \Delta E + 4\Delta G_M. \quad (3.144)$$

The algebraic sign of the total reaction results from both parts. If the redox potential is high enough, protons are pumped to the outside against the gradient.

Exercises

Since the correct set up of the equations is one of the most important tasks in systems biology a number of exercises are presented here to deepen the information presented in this chapter.

EXERCISE 3.1 Master equation.

Consider a simple model of DNA binding where an inactive binding site D_i or the active D_a form is available. The RNA polymerase can change the state from inactive to active. As the polymerase is not explicitly modeled, one obtains the following reaction scheme for which a master equation is to be determined and analyzed:



a For the process for which a component is increased and, respectively, decreased the driving force is the component itself. For the given reaction equation, determine the transitions for the variable D_i . Note that the total amount is constant: $n_T = D_i + D_a$. Complete the table for the corresponding driving forces and the kinetic parameter for the individual cases.

	Case 1	Case 2	Case 3	Case 4
State D_i	$\#D_i = n - 1 \rightarrow n$			
D_a	$\#D_a = m + 1 \rightarrow m$			
Driving force	$\#D_a = m + 1$			
Parameter	k_2			

Using the entries of the table, a master equation for $P_n(D_i)$ can be determined.

b Simplify the master equation for the case $n_T = 1$. Eliminate the corresponding terms from the master equation.

c Determine the solutions of the master equation for $P_0(D_i)$ with the initial conditions: (i) $P_0(D_i(t=0)) = 1$ and (ii) $P_0(D_i(t=0)) = 0$. Illustrate graphically the significance of initial conditions. Show that the solutions for a large time are stationary and independent of the initial conditions.

d Determine the mean value and variance of the steady state. Compare with the deterministic solution.

EXERCISE 3.2 Balance equation.

A biotechnological process is ongoing in a continuously running reactor with a volume V (substrate S_i will be added at a rate of q_i^{in} and a concentration S_i^{in}). The biochemical network of the biomass shall be considered, where for every extracellular substrate an intracellular metabolite exists (M_i) and a single transport rate (r_{ai}) is taken into account (stoichiometry N_a). In total there are n substrates/metabolites and q reactions in the intracellular network (r_i , stoichiometry N_i) that is taken into account. The metabolites M_i are precursors for the main components P_j of the biomass. The reaction equations for the synthesis are:



a Determine the dimensions of the matrices and vectors of the system.

b Determine the corresponding differential equations for the individual components (substrate, biomass, intracellular metabolites and main components of the biomass).

c Determine the steady state of the whole system. Under what condition can the rates r_i in the network be directly read from the measurement data (uptake rates and biomass composition)?

EXERCISE 3.3 Analysis of the growth rate.

The basic equation for the dynamics of intracellular metabolites reads:

$$\dot{X} = \sum r_j - \mu X, \quad (3.147)$$

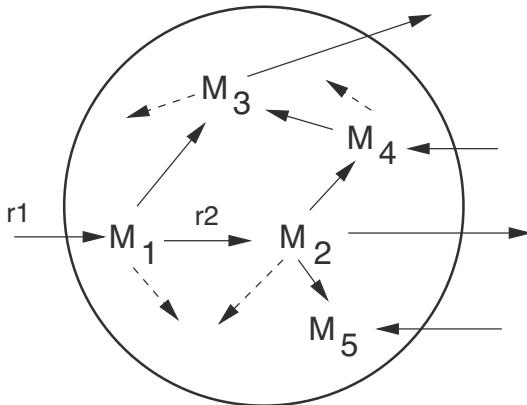


FIGURE 3.22: A network with uptake/production as well as intracellular rates (solid line) and drain (dashed line) to the main components of the biomass.

with a specific growth rate μ . The rates r_j are assumed to be dependent on the growth rate. This yields the equation:

$$\dot{X} = f(\mu) - \mu X, \quad (3.148)$$

where f is a function to be defined.

a Calculate and sketch the steady states of the system for the following functions f :

$$f_1 = k_1 \mu^2; \quad f_2 = k_2 \frac{\mu}{\mu + K}; \quad f_3 = k_3 \frac{\mu^n}{\mu^n + K}. \quad (3.149)$$

b Discuss in detail the course of the steady states at f_3 for $n = 2$.

EXERCISE 3.4 Two populations model.

A structured model divides the whole biomass m into the sub-populations B_1 and B_2 (unit: gB_q/gDM). Both sub-populations can forge into one another (transition) using the rate r_B (reaction speed, unit: $gB/gDMh$):



For every sub-population, an intracellular network is considered with stoichiometry:



The rates have units of g/gB_qh . Sought next is a general form of the balance equation for the intracellular metabolites c_X^i (unit: gX/gB_q) where i is the index of the population. Proceed as following:

a Set up the mass balance for the sub-population B_q . Take into the account the transition of the population as well as an individual growth rate μ_i for each population. Calculate the following term from the balance:

$$\frac{\dot{(B_q m)}}{B_q m} . \quad (3.152)$$

b For setting up the mass balance for metabolites c_X^i ($i = 1, 2$), the reactions of the intracellular stoichiometric network are to be taken into account. When a population transits into the other, all the metabolites do so as well with the same rate. For this case, provide the balance equation for the metabolites X in both sub-populations and interpret the results.

EXERCISE 3.5 Two compartment model.

The balance equations for an organism with two cellular compartments shall be determined. As can be seen in the figure, metabolite X is present in compartment 1 (X_1) and compartment (X_2). Provide the balance equations for both metabolites. Take into account that the rate r_1 based on the mass in compartment 1 (unit: $g/g m_1 h$) and the rate r_2 is based on the mass in compartment 2 (unit: $g/g m_2 h$). Both compartments grow independently at specific growth rates of μ_1 and μ_2 (this for instance is applicable for a cell and mitochondria).

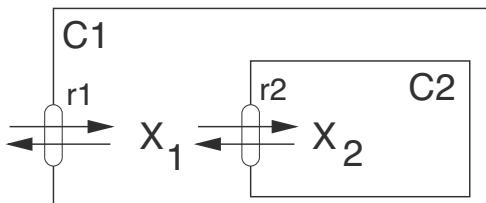


FIGURE 3.23: A network with two compartments.

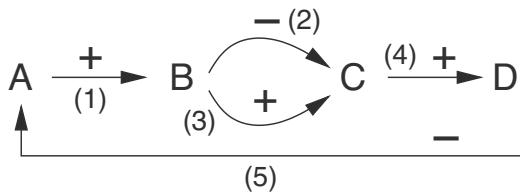
a What are the corresponding differential equations for the metabolite X_1 and X_2 ?

b Determine the stationary values of both concentrations for X_1 and X_2 .

c Experimentally, only the total amount of substance $X = X_1 + X_2$ can be measured. Provide a suitable relationship for this mean concentration. Investigate this relationship in connection with the ratio f of both masses where $f = m_1/m_2$.

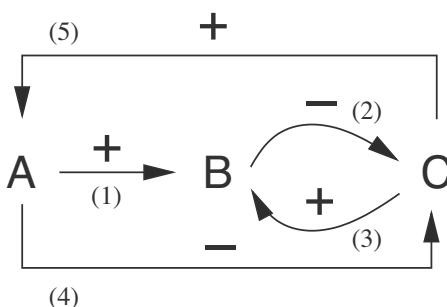
EXERCISE 3.6 Network analysis.

Given is the following graph:

**FIGURE 3.24:** A simple graph.

a Provide the adjacency and incidence matrices. Calculate the null space of the incidence matrix. What dimensions does the null space have? According to the definition, all entries in the null vector must be positive so that it represents a feedback loop. The second condition is that the vector is not allowed to be decomposable. How many feedback loops do you obtain? Provide the sign of the loops as well. Show that the vector $\underline{g} = \{2, 1, 1, 2, 2\}$ is a null space vector but does not represent a feedback loop.

b Consider the extended network:

**FIGURE 3.25:** A simple graph (2).

Provide here as well the adjacency and incidence matrices. What are the dimensions of the null space of the incidence matrix? Provide the null space. Provide the feedback loops with their corresponding signs. Why does not the vector $\underline{g} = \{2, 3, 1, 0, 2\}$ represent a feedback loop?

EXERCISE 3.7 Lactose metabolism.

The following graph is that of lactose metabolism:

The model can be refined. The extended version shall contain the β -galactosidase (LacZ) that splits the intracellular lactose. Also to be taken into account is that the permease (LacY) and β galactosidase (LacZ) do not act on other components yet they do influence the rate. For the model refinement, use a bipartite graph. This basically is the splitting of the nodes into

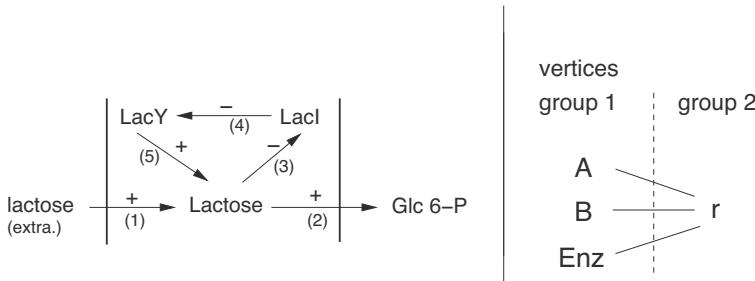


FIGURE 3.26: Left: A simple scheme of lactose metabolism. Right: A reaction r converts A to B and is catalyzed by the enzyme Enz . The conversion is represented by a bipartite graph.

two disjoint groups G_1 and G_2 . The edges can exist between an element from G_1 and an element from G_2 , but not between nodes of the same group. As an example (in Figure 3.26) is a conversion process $A \rightarrow B$ shown catalyzed by an enzyme. How can the lactose metabolism be represented in more detail? Define two types of nodes so as to yield a bipartite graph.

EXERCISE 3.8 Quantitative analysis of a reaction chain.

Consider a linear cascade with an arbitrary number of components X_i . The equations read:

$$\begin{aligned}\dot{X}_1 &= u - X \\ \dot{X}_l &= f(X_{l-1}) - X_l, \quad l \geq 2.\end{aligned}$$

Using f , an activation ($f(X_{l-1}) = X_{l-1}$) or an inhibition ($f(X_{l-1}) = \frac{1}{X_{l-1}}$) of the component l by component $l-1$ can be expressed. Show that for an even number of inhibitions, the equilibrium (steady state) relationship for the last element is $X_n = u$. What does this mean for long chains and circuit pathways?

EXERCISE 3.9 Reaction with ATP.

a Under standard conditions the ATP delivers a $\Delta G^0'$ of -30 kJ/mol . In retrograde, this energy is required for the synthesis of ATP. How large is the energy delivered by the proton's gradient when a voltage of $\Delta\Psi$ of 0.14 V and a pH difference of 0.05 is assumed, letting the temperature $T = 293.15\text{ K}$? What is the least number of protons required to produce a mole of ATP? How does the result change if the concentration of ATP and ADP in the cell is taken to be 7.9 mM and 1 mM , respectively?

b Consider the given reaction network on the left side of the following figure. The synthesis and consumption of A_1 and A_2 is not considered so that a steady state can occur. Provide all the null space vectors of the stoichiometric

matrix. Which condition must the concentrations of A_1 and A_2 fulfill in connection with the dependence on the equilibrium constants K_i of the reactions r_2 , r_3 and r_4 , such that a flow distribution according to the directions of the reactions (arrows) can be used?

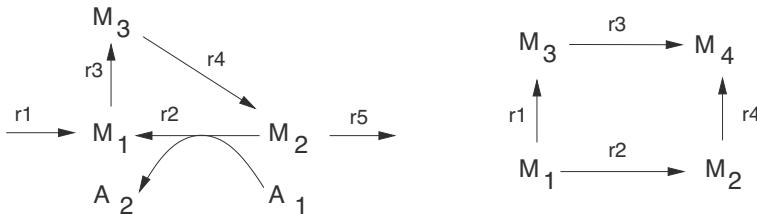


FIGURE 3.27: Left: A network with a cycle. Right: Reaction $M_1 \rightarrow M_4$ with two intermediate products.

c Consider a reaction $M_1 \rightarrow M_4$ with two intermediate products (Figure 3.27, right). Which relationship results for the equilibrium constants of the partial reactions?

EXERCISE 3.10 *Network analysis.*

a The following reaction network is given. The rates r_7 and r_8 are uptake and production rates and are set to be irreversible. The rates r_2 to r_6 characterize the intracellular network.

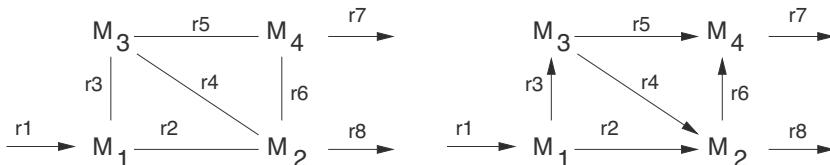


FIGURE 3.28: A network with 4 components and 8 reactions.

It is assumed that the reactions r_2 to r_6 are irreversible (consider the left plot, the direction is not fixed here!). Give examples for the following conditions by fixing the direction (sketch the arrows):

- The network has a flux distribution where the fluxes are balanced.
- The network has a flux distribution where the thermodynamic constraints are not fulfilled.
- The network has a flux distribution where fluxes are balanced and the thermodynamic constraints are satisfied. Provide the conditions for the chemical potential of the metabolites M_i involved.

b How many network structures in total can be investigated (the number of combinations when the direction is fixed)? How many of them fulfill the node's rule (are balanced) and satisfy the thermodynamic conditions?

c Provide the stoichiometric matrix N . Proceed from the illustration on the right. As the cycles are to be determined, a reduced matrix with the rates r_2 to r_6 can be used. The rank of the reduced matrix \hat{N} is 3. What is the dimension of the null space? Provide the null space vectors. Show that 3 cycles can be determined from the null space using linear combinations.

EXERCISE 3.11 Interrelation between reaction rate and Gibbs free energy.

a A correlation between the reaction rate of a reaction r and the Gibbs free energy ΔG of the reaction is to be derived when the reaction is near the equilibrium. Consider the reaction:

$$A \rightleftharpoons B \quad \text{with} \quad r = r_1 - r_2 = k_1 A - k_2 B. \quad (3.153)$$

Convert the equation for the chemical potential

$$\mu_i = \mu^0 + \mathbb{R} T \ln \frac{c_{Mi}}{c_{Mi}^r} \quad (3.154)$$

to the concentration for a metabolite and insert the expression into the reaction speed. For a reference concentration c_{Mi}^{Eq} , the concentration at equilibrium is chosen. Approximate the e-function for the equilibrium point (Taylor series).

b The previous result shall be verified using a different approach. As a first step, show that the variation of the Gibbs free energy can be described as follows:

$$\Delta G = -\mathbb{R} T \ln \frac{r_1}{r_2}. \quad (3.155)$$

Carry out the Taylor expansion as earlier at equilibrium and show that the result is consistent with the result from above. The relationship in Equation (3.155) shall now be used to deduce a relationship for reactions that are almost irreversible. $r_1 \gg r_2$ applies here. Show the net flow of the reaction in this case is proportional to r_2 .

Bibliography

- [1] W. Wiechert. Modeling and simulation: tools for metabolic engineering. *J. Biotech.* 94: 37-63, 2002.
- [2] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* 81(25): 2340-2361, 1977.

- [3] E. M. Ozbudak et al. Multistability in the lactose utilization network of *Escherichia coli*. *Nature* 427, 2004, p. 737.
- [4] O. Wolkenhauer et al. Modeling and simulation of intracellular dynamics: Choosing an appropriate framework. *IEEE Transactions on Nanobioscience* 3, 2004, 200-207.
- [5] R. Diestel. *Graph Theory*. Springer, 2010.
- [6] M. Ch. Körner and A. Schöbel (Editors). *Gene, Graphen, Organismen*. Shaker, 2010.
- [7] S. Klamt et al. A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics* 7: 56, 2006.
- [8] J. Nielsen, J. Villadsen and G. Liden. *Bioreaction Engineering Principles*. Kluwer Academic/Plenum Publisher, 2003
- [9] G.-Y. Zhu et al. Model predictive control of continuous yeast bioreactors using cell population balance models. *Chem. Eng. Science* 55: 6155-6167, 2000.
- [10] D. L. Nelson and M. M. Cox. *Lehninger Principles of Biochemistry*. Palgrave Macmillan, 2008.
- [11] D. A. Beard et al. Thermodynamic constraints for biochemical networks. *J. Theor. Biol.* 228: 327–333, 2004.
- [12] S. Klamt et al. Modeling the electron transport chain of purple non-sulfur bacteria. *Mol. Sys. Biol.*, 4: 156, 2008.

Chapter 4

Model Calibration and Experimental Design

Model calibration is a process of getting simulation calculations to match experimental results and data. In general, this can be done via an optimization procedure where the structure of the model is not changed but only the kinetic parameters. If the system under consideration is strongly non linear with respect to the parameters, advanced numerical methods must be used. In the following section, the focus will not be on optimization algorithms but rather on the accuracy and quality of the estimated parameters and the model. In systems biology the measured data are often available only with large uncertainties. This is what makes knowledge of the parameters' uncertainties that will be estimated even more important.

4.1 Regression

Regression refers to the determination of the parameters of a mathematical model with a state variable Y when measurement data Y_M is available. Next, two approximations are introduced, the Least Square method (LS) and the Maximum Likelihood (ML) approach.

4.1.1 Least square method & maximum likelihood estimation

For the Least Square method (LS), one adds the quadratic difference between the model and experiment for all measured time points L . In order to weight the differences, they are divided by the measurement error σ_k . This means that measurement points with large errors are weakly considered in comparison with those with small errors. This results in the following optimization problem:

$$\min_{\underline{p}} \Phi = \sum_{k=1}^L \frac{(Y_{Mk} - Y_k(\underline{p}))^2}{\sigma_k^2}, \quad (4.1)$$

where Y_M denotes the measurement data, Y the simulation data and \underline{p} represents the parameter vector with n_p elements.

The Maximum Likelihood (ML) approach follows a different way to look at the parameters. It starts with the question, what is the probability with which a parameter describes a set of the measured data? What is sought after is then to determine the parameter set with the highest probability. This is shown in the left part of Figure 4.1.

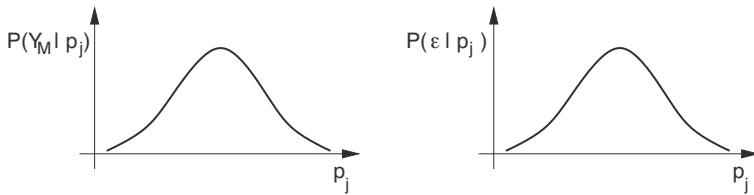


FIGURE 4.1: Probability densities for Y_M and ϵ as a function of an individual parameter p_j .

In Figure 4.1, $P(Y_M|p_j)$ is the conditional probability of the measurement Y_M as a function of the parameter p_j . The parameter is therefore considered as a random variable. Because the individual measurement points Y_{Mk} cannot be considered independent of each other, instead, the quantity $P(\epsilon = Y_M - Y|p_j)$ is used which is more suitable. The differences between the model and the experiment ϵ should be purely random and average to zero. Therefore, it is expected that it is a normally distributed quantity. What is sought now is the parameter p_j which maximizes this probability P . Here, all the values of ϵ pairs can be assumed to be independent and this results in the following product (all elementary events have to occur) for the probability $P(\epsilon|p_j)$:

$$P(\epsilon|p_j) = P(\epsilon_1|p_j) \cdot P(\epsilon_2|p_j) \cdot \dots \cdot P(\epsilon_L|p_j). \quad (4.2)$$

Let the individual ϵ_k be normally distributed, then the following is applicable:

$$P(\epsilon_k|p_j) = \frac{1}{\sqrt{2\pi}\sigma_k} \cdot e^{-\frac{1}{2} \cdot \frac{\epsilon_k^2}{\sigma_k^2}}. \quad (4.3)$$

Substituting the above equation in the product we obtain the Likelihood function L :

$$L = P(\epsilon|p_j) = \prod_{k=1}^L \frac{1}{\sqrt{2\pi}\sigma_k} \cdot e^{-\frac{1}{2} \cdot \frac{\epsilon_k^2}{\sigma_k^2}}, \quad (4.4)$$

which is to be maximized. Instead of the function L , as an alternative $\ln L$ can be considered which is easier to calculate. One assumes the same measurement

uncertainties σ for all time points; this results in:

$$\ln L = \ln \frac{1}{\sigma}^L + \ln \frac{1}{\sqrt{2\pi}}^L - \frac{1}{2\sigma^2} \sum_{k=1}^L e_k^2. \quad (4.5)$$

This relationship is equivalent to the LS method as when differentiating the first two terms here makes them vanish, and one assumes that the measurement uncertainty σ is known. The ML method itself presents a general approach, which can be used for other distributions as well.

4.1.2 Linear models

Cellular networks distinguish themselves by their strong nonlinearity. However, many of the parametric analysis methods are based on linear approximations of the mathematical model with respect to the parameters. For this reason, the most important relationships shall be summarized first. A model with a single output quantity is considered, where the parameters p_j take a linear form. Those parameters can be written as a column vector \underline{p} , which along with the vector \underline{x} of the independent variables (normally a selected state variable or time t) lead to the following representation:

$$Y = \underline{x}^T \underline{p} \quad (4.6)$$

where the row vector can contain an arbitrary function of the independent variable x , shown in the following example.

EXAMPLE 4.1 Second order polynomial.

The following model is used:

$$\begin{aligned} Y &= p_1 x^2 + p_2 x + p_3 \\ &= (x^2 \ x \ 1) \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}. \end{aligned} \quad (4.7)$$

The model is linear in terms of the parameters p_j , but non linear in terms of the regression variable x . For the method introduced, it is useful to express all equations as vectors and matrices. The column vector of the measured data therefore reads:

$$\underline{Y}_M = \begin{pmatrix} Y_{M1} \\ Y_{M2} \\ \dots \\ Y_{ML} \end{pmatrix}. \quad (4.8)$$

The row vector \underline{x}^T can be written for all measurements as a matrix X with L rows and n_p columns:

$$X = \begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ \dots & \dots & \dots \\ x_N^2 & x_N & 1 \end{bmatrix}. \quad (4.9)$$

The objective function can be written using the LS approximation:

$$\Phi = \sum_{k=1}^L \frac{(Y_{Mk} - Y_k)^2}{\sigma_k^2} = (\underline{Y}_M - X \underline{p})^T \Sigma^{-1} (\underline{Y}_M - X \underline{p}) \quad (4.10)$$

where Σ is a matrix whose main diagonal represents the variance of the measurement error σ_k^2 . Differentiating with respect to the parameters results in a solution of the optimization problem for the desired parameters vector:

$$\hat{\underline{p}} = A^{-1} \underline{b}, \quad (4.11)$$

where matrix A and vector \underline{b} are given as:

$$A = X^T \Sigma^{-1} X, \quad \underline{b} = X^T \Sigma^{-1} \underline{Y}_M. \quad (4.12)$$

Matrix A is a $n_P \times n_p$ matrix; vector \underline{b} has n_p rows.

4.2 Model and Parameter Accuracy

To ensure the quality of the results, statistical tests are carried out. Those tests are applicable to the model and the parameters. First, the quality of the model will be assessed followed by the accuracy of the parameters. The previously defined functional Φ can be considered as a random variable for the comparison of the model's calculations and the measured data:

$$\Phi = \sum_{k=1}^L \left(\frac{\epsilon_k}{\sigma_k} \right)^2 = \underline{\epsilon}^T \Sigma^{-1} \underline{\epsilon} \quad (4.13)$$

where ϵ_k is the deviation between the model's calculation and simulation (residues). Therefore, the random variable Φ follows a χ^2 distribution with $df = L - n_p$ degrees of freedom (see Appendix). Note that the number of parameters has to be subtracted from the number of data points. This can be explained as follows: Assume the model is a straight line that can be described with two parameters. Then, two measured data points are enough to calculate — and not to estimate — the parameters. However, more data points are

available in general and the number of data points that exceed the minimum number of data points is called the degree of freedom. Usually, the domain of χ^2 with a 95% confidence interval is calculated.

If Φ generates values larger than what the distribution function allows, then (i) the model is not correct and must be modified or (ii) the measurement error was set too low. As it is assumed here that the deviations ϵ_k are normally distributed, every summand should — on average — contribute a value of 1 to the sum given above. Therefore, as a rule of thumb Φ is:

$$\Phi \approx L. \quad (4.14)$$

The numerical values can be determined concretely from known tables (or corresponding software). The parameters of a straight line ($n_p = 2$) and $N = 20$ measurement points shall be determined, and one obtains for Φ the following minimal and maximal values ($df = 18$): 8.23 and 31.53.

The residues ϵ_k can be examined for their correlation. One first calculates the covariances between a value $\epsilon_k(t_i)$ and l time steps shifted value $\epsilon_k(t_{i-l})$. The reason is as follows: In a dynamical system, state variables may be correlated not directly but indirectly via other components and processes. This results in time delays and effects will be visible first some time steps later on. By shifting the residues' correlations that occur not at the same time, points can be detected. If a correlation is present then this suggests the presence of model or measurement uncertainties.

$$Cov_\epsilon(l) = \frac{1}{L} \sum_{i=1+l}^L \epsilon_k(t_i) \epsilon_k(t_{i-l}). \quad (4.15)$$

The covariance is then related to the variance of ϵ_k (can be represented as $Cov_\epsilon(0)$) and summed up. Thus, one has a measure, which in turn can be tested against a χ^2 distribution with l degrees of freedom as above:

$$R(l) = \frac{\sum_l Cov_\epsilon(l)^2}{Cov_\epsilon(0)^2}. \quad (4.16)$$

The accuracy of the parameters, i.e., their confidence interval, can be determined with the help of the matrix A . Therefore we calculate the mean value (expectation value) and the variance of the parameters. For the expectation value of the parameters one finds:

$$\begin{aligned} E[\hat{p}] &= E\left[\left(X^T \Sigma^{-1} X\right)^{-1} \cdot X^T \Sigma^{-1} \underline{Y}_M\right] \\ &= E\left[\left(X^T \Sigma^{-1} X\right)^{-1} \cdot X^T \Sigma^{-1} (X \underline{p})\right] = \hat{p}. \end{aligned} \quad (4.17)$$

Hence it is an unbiased estimation. The following is then applicable to the

variance:

$$\text{Var}[\hat{p}] = \text{Var}\left[\left(X^T \Sigma^{-1} X\right)^{-1} \cdot X^T \Sigma^{-1} \underline{Y}_M\right]. \quad (4.18)$$

Using the calculation rule for variances (see Appendix) one obtains:

$$\begin{aligned} \text{Var}[\hat{p}] &= (X^T \Sigma^{-1} X)^{-1} \cdot X^T \Sigma^{-1} \text{Var}[\underline{Y}_M] \cdot \\ &\quad \left((X^T \Sigma^{-1} X)^{-1} \cdot X^T \Sigma^{-1}\right)^T; \end{aligned} \quad (4.19)$$

inserting the measurement uncertainties σ_k for $\text{Var}[\underline{Y}_M]$, resolving the last parenthesis, that is, taking the transpose, and assorting, one obtains:

$$\begin{aligned} \text{Var}[\hat{p}] &= (X^T \Sigma^{-1} X)^{-1} \cdot X^T \cdot \left(\Sigma^{-1T} X \left((X^T \Sigma^{-1} X)^{-1}\right)^T\right) \\ &= (X^T \Sigma^{-1} X)^{-1} = A^{-1}. \end{aligned} \quad (4.20)$$

The inverse of the matrix A is therefore the variance-covariance matrix of the parameter uncertainties and one obtains for every single parameter the variances σ_{p_j} as well as the covariances $\text{cov}_{p_{lm}}$ between the parameters:

$$\sigma_{p_j} = \sqrt{(A^{-1})_{jj}}, \quad \text{cov}_{p_{lm}} = (A^{-1})_{lm} \quad l \neq m. \quad (4.21)$$

Geometric Representation

The result can also be illustrated in another way. For this, the behavior of the system in the proximity of the optimum parameters vector is considered via a small deviation Δp_j of the parameters. Using a Taylor approximation (see the Appendix), this deviation can be calculated:

$$\Delta Y = \frac{\partial Y}{\partial p_1} \cdot \Delta p_1 + \dots + \frac{\partial Y}{\partial p_n} \cdot \Delta p_n. \quad (4.22)$$

As a linear model is considered and along with the vector of the regression variables \underline{x} , the variation Y' of the model's output is:

$$Y' = Y + \Delta Y = Y + \underline{x} \Delta \underline{p}. \quad (4.23)$$

Inserting the objective function, which is now being varied, $\Delta \Phi$, and using the matrix notation we obtain:

$$\Phi + \Delta \Phi = (\underline{Y}_M - (\underline{Y} + X \Delta \underline{p}))^T \Sigma^{-1} (\underline{Y}_M - (\underline{Y} + X \Delta \underline{p})). \quad (4.24)$$

With $\underline{Y}_M \approx \underline{Y}$ (one assumes that the measured values are reproduced very well) this gives:

$$\begin{aligned} \Delta \Phi &= (X \Delta \underline{p})^T \Sigma^{-1} (X \Delta \underline{p}) \\ &= \Delta \underline{p}^T X^T \Sigma^{-1} X \Delta \underline{p} = \Delta \underline{p}^T A \Delta \underline{p}. \end{aligned} \quad (4.25)$$

In this representation, the covariance matrix A of the parameters shows up again. To better understand the above equation, a coordinate transformation is introduced. Matrix A is then decomposed into $A = T \Lambda T^{-1}$, where Λ is a diagonal matrix whose entries are the eigenvalues of A . Inserting this decomposition the following relationship is obtained:

$$\boxed{\Delta\Phi = \Delta p^T T \Lambda T^{-1} \Delta p = \Delta p^{*T} \Lambda \Delta p^*} \quad (4.26)$$

which is an ellipse for the case of two parameters. A question that arises is, which values of $\Delta\Phi$ are to be used. The ellipse represents the area which defines the allowed values of the parameters' deviations Δp_1 and Δp_2 in order to tolerate large values of the objective function. By fixing one parameter (degrees of freedom = 1), one determines for a confidence interval the standard deviation (confidence interval of 68.28%) corresponding to the value $\Delta\Phi = 1$. Here Φ is a χ^2 distributed quantity and this applies to $\Delta\Phi$ as well. The position and the form of the ellipse allow drawing the following conclusions:

- λ_i is large \rightarrow small main axis \rightarrow small parameters' uncertainty
 λ_i is small \rightarrow large main axis \rightarrow large parameters' uncertainty
- The ellipse goes from the bottom left to the top right; the parameters are possessing sensitivities with different signs.
- If the ellipse is very flat, the parameters are then correlated. If it is closer to a perfect circle, the parameters are (almost) uncorrelated.

A strong correlation of the parameters is the biggest problem of parameters' estimation, where the parameters can not be determined independently of each other.

Figure 4.2 illustrates the different cases that can arise. The curves shown in each plot represent contours with values of the objective function. Points outside the ellipse have large values of the objective function. In **A**, both sensitivities have the same sign with respect to the objective function: After a perturbation in the direction of Δp_2 (increasing the parameter), and then to return to the curve, the system must move in the negative Δp_1 direction. Case **B** is just the opposite: Both parameters are increased and return to the same value of the objective function. The case in **C** shows two correlated parameters. If one of the parameters is fixed then this leaves the other with little freedom. Case **D** suggests only a small uncertainty for the parameter p_2 . If p_2 is fixed, a statement regarding p_1 can hardly be made. If p_1 has been fixed, then there is less variability in parameter p_2 since the area of the ellipse is very small.

The standard deviation of the parameters can be determined using a projection on the axes. Because the degree of freedom = 1 is selected, the projection of the tangents on the ellipse results in the σ_{p_j} values of the parameters. The confidence interval can therefore be given:

$$p_1 - 1.96 \sigma_{p_1} < \bar{p}_1 < p_1 + 1.96 \sigma_{p_1}. \quad (4.27)$$

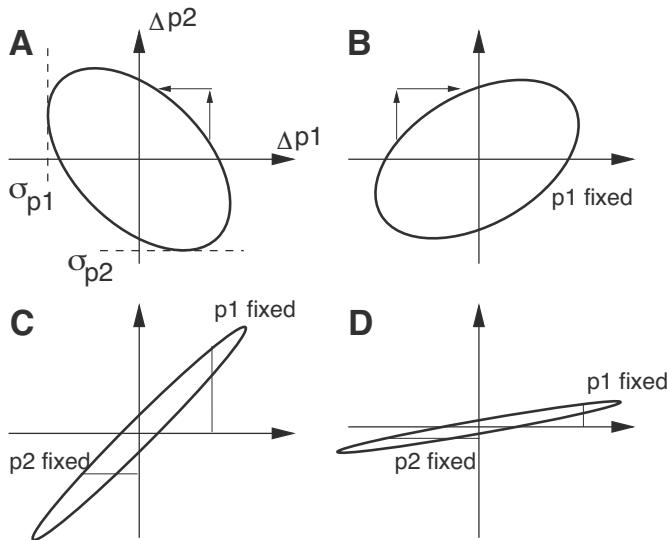


FIGURE 4.2: **A** The parameters have the same sign and sensitivities. The standard deviation can be found by searching the intersection point of the maximal values with the respective axis. **B** The parameters have different signs and sensitivities. **C** Both parameters are correlated. Fixing p_1 leads to fewer possibilities for p_2 , i.e., p_2 has less freedom within the area. **D** Worst case scenario. p_1 has large uncertainty and is correlated to p_2 .

The value 1.96 results from the normal distribution, if one wants to cover 95% of the corresponding data. The following consideration can help in determining the parameter with the largest influence on the objective function: one looks for a combination of the parameters that lead to the maximal variance of $\Delta\Phi$. Therefore:

$$\max \Delta\Phi = \underline{\Delta p}^T A \underline{\Delta p} \quad \text{with the constraint: } \underline{\Delta p}^T \underline{\Delta p} = 1.$$

The constraint reduces the possible parameters' combinations and allows solving the problem analytically. The calculation leads to the optimization problem (the constraint is integrated within the objective function):

$$\max \underline{\Delta p}^T A \underline{\Delta p} - \lambda \cdot (\underline{\Delta p}^T \underline{\Delta p} - 1), \quad (4.28)$$

which can be solved as following:

$$\frac{d\Delta\Phi}{d\underline{\Delta p}} = 2A \underline{\Delta p} - 2\lambda \underline{\Delta p} = 0 \rightarrow A \underline{\Delta p} = \lambda \underline{\Delta p}. \quad (4.29)$$

This leads to an eigenvalue problem, where the eigenvector of the maximal eigenvalue indicates the direction of the optimal deviation $\underline{\Delta p}$.

If the variance of the measurement is not known, the following approach is applicable. The deviation $\Delta\Phi$ is:

$$\Delta\Phi = \underline{\Delta p}^T X^T X \underline{\Delta p} = \underline{\Delta p}^T A^* \underline{\Delta p}. \quad (4.30)$$

The estimated value of the uncertainty taking into account the objective function Φ is:

$$\hat{\sigma}^2 = \frac{\Phi}{L - n_p}, \quad (4.31)$$

where n_p is the number of parameters and L is the number of the measurement points which were used to determine the variance. Expanding the above equation for $\Delta\Phi$ leads to:

$$\begin{aligned} \underline{\Delta p}^T A^* \underline{\Delta p} &= \Delta\Phi \frac{\hat{\sigma}^2}{\Phi/(L - n_p)} \\ \rightarrow \underline{\Delta p}^T \frac{A^*}{\hat{\sigma}^2} \underline{\Delta p} &= \frac{\Delta\Phi/1}{\Phi/(L - n_p)}. \end{aligned} \quad (4.32)$$

The right side shows the ratio of two χ^2 distributions with

$$\Delta\Phi \sim \chi_1^2 \quad (4.33)$$

$$\Phi \sim \chi_{L-n_p}^2. \quad (4.34)$$

The behavior of two χ^2 distributions is a F distribution:

$$\underline{\Delta p}^T \frac{A^*}{\hat{\sigma}^2} \underline{\Delta p} \sim F_{1, L-n_p}, \quad (4.35)$$

where now two degrees of freedom are present. As the measurement error $\hat{\sigma}$ is not certain, the F distribution leads to larger ellipses. Figure 4.3 presents an example. Here two curves are shown representing different levels of uncertainties. In the first case (solid line) the standard case with one degree of freedom and 68.3% certainty is shown for fitting a straight line (these data are not given explicitly). In the second case, the standard deviation is not known but estimated based on 5 measurements. Therefore, the ellipse is larger since the uncertainties in the measurements lead to additional uncertainties in the parameters.

An important relationship allows the calculation of the uncertainty of the measurement quantities \underline{Y}_M , if a prediction of the following form is made with new values X_0 :

$$\hat{\underline{Y}}_{M0} = X_0 \hat{\underline{p}}. \quad (4.36)$$

If one seeks the variance of the estimation $\hat{\underline{Y}}_{M0}$ given here, the following quantity is considered next:

$$\Psi = \underline{Y}_{M0} - \hat{\underline{Y}}_{M0}. \quad (4.37)$$

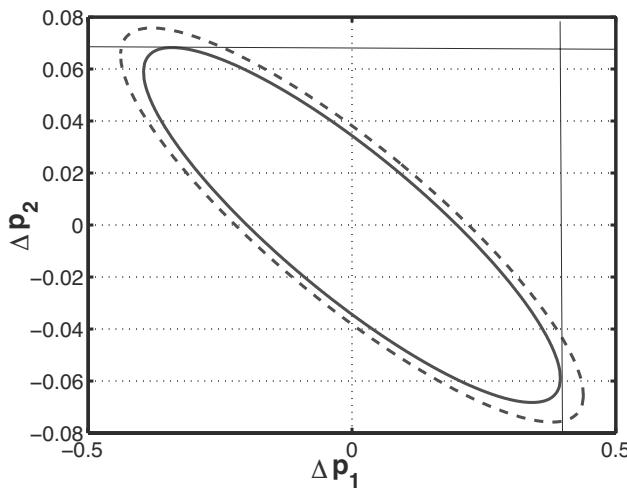


FIGURE 4.3: Solid curve: χ^2 with a single degree of freedom. The standard deviation of the parameters can be found with the intersection of the two lines with the respective axis. Dashed: F distribution with one and five degrees of freedom.

For the variance of Ψ one obtains:

$$\text{Var}[\Psi] = \text{Var}[\underline{Y}_{M0}] + \text{Var}[\hat{Y}_{M0}]. \quad (4.38)$$

Both quantities here are uncorrelated when considered statistically. Calculating both components leads to:

$$\text{Var}[\Psi] = \Sigma + \text{Var}[X_0 \hat{p}] - 2 \text{Cov} [\underline{Y}_{M0}, X_0 \hat{p}]. \quad (4.39)$$

As no covariance arises here between both quantities, this simplifies the equation:

$$\text{Var}[\Psi] = \Sigma + X_0^T \text{Var}[\hat{p}] X_0 = \Sigma + X_0^T (X^T \Sigma^{-1} X)^{-1} X_0. \quad (4.40)$$

The confidence interval (95%) for the prediction is then given:

$$\begin{aligned} \hat{Y}_{M0} - 1.96 \sqrt{\Sigma + X_0^T (X^T \Sigma^{-1} X)^{-1} X_0} &< \underline{Y}_{M0} \\ &< \hat{Y}_{M0} + 1.96 \sqrt{\Sigma + X_0^T (X^T \Sigma^{-1} X)^{-1} X_0}. \end{aligned} \quad (4.41)$$

4.2.1 Nonlinear models

For models whose parameters are non linear, there exists no general theory for determining the parameters' uncertainties. Also, some progress was made

in the last few years. Two approaches will be followed which will enable making some conclusions.

4.2.1.1 Linearization

One of the approaches for non linear equations is their linearization about a set point Y^* . One then obtains for the model's equation:

$$Y = Y^* + \frac{\partial Y}{\partial p_1} \Delta p_1 + \frac{\partial Y}{\partial p_2} \Delta p_2 \dots \quad (4.42)$$

In this case the derivatives

$$\frac{\partial Y}{\partial p_j}|_{\underline{p}} \quad (4.43)$$

are dependent on the selected parameter's vector (set point) \underline{p} . Otherwise, one proceeds just as in the linear case earlier, using the matrix of the derivatives $W^T = \frac{\partial Y}{\partial \underline{p}}$:

$$\Delta \Phi = \Delta \underline{p}^T W^T \Sigma^{-1} W \Delta \underline{p} = \Delta \underline{p}^T F \Delta \underline{p}. \quad (4.44)$$

In the nonlinear case, the term $F = W^T \Sigma^{-1} W$ is known as the Fisher Information Matrix (FIM). The following relationship can be shown:

$$\sigma_{p_j} \geq \sqrt{(F^{-1})_{jj}}, \quad (4.45)$$

that is, only a lower limit for the parameter's confidence interval can be given; the actual but unknown uncertainties limits are larger in any case.

4.2.1.2 Bootstrapping

One alternative approach is the Bootstrap method [1]. This is a statistical method to determine parameters' uncertainties. There idea here, based on the set of measurement data \mathbf{D} , is to estimate the parameters $\hat{\underline{p}}$ not only once, but more often. The measured data will be randomly adjusted in the range of their measurement uncertainty and a series of new data $\mathbf{D}_1^*, \mathbf{D}_2^*$, etc. is generated. For every data set one obtains a different set of the estimated parameters \hat{p}_1^*, \hat{p}_2^* . The determination of the parameters' uncertainties is done via the evaluation of parametric histograms. The data are sorted and plotted according to frequency. A confidence interval of 95% can then be directly read out from the histogram.

EXAMPLE 4.2 Comparison between Fisher information matrix and Bootstrapping approximations.

Consider the following non linear function in terms of the parameter b :

$$Y = \frac{1}{b+x}$$

Figure 4.4 shows exemplary histograms, when the range of x contains ($0 \leq x \leq 25$) 1000 (left) or 25 (right) measurement points. For a large number of measurement points a normal distribution can be detected.

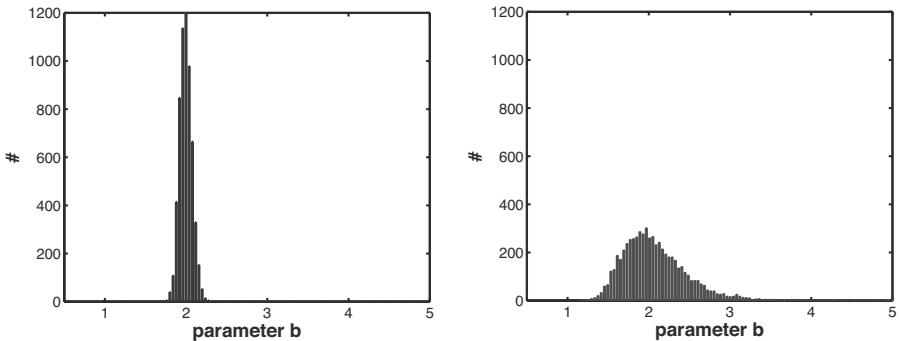


FIGURE 4.4: Histogram for 6000 Bootstrapping runs. For 1000 measurement points one obtains a normal distribution (left) while for 25 measurement points (right), a distortion can be seen. The chosen value for b is $b = 2$. The Fisher information matrix gives smaller intervals in the second case.

For a small number of measurement points a notable distortion in the histogram is visible; however, the average value of both distributions is almost identical. The Fisher information matrix gives smaller intervals in the second case in comparison with the bootstrap method (in the range of 5%).

The difference between the linear and non linear models shall now be illustrated geometrically. It is assumed that two measurement points Y_{M1} and Y_{M2} are available. The Figure 4.5 (left) shows the measurement points and a linear as well as a non linear model, which are specified by the measurement points. The data can now be differently plotted: Consider the space Y_M of the measurement points (here: two dimensional). Now, the associated x_k values are held fixed and in the model's equations the parameter is varied (here: b). The curves $Y(b)$ denote the solution locus and lie in the proximity of the measurement points. A vertical projection of the measurement points on the model Y results in the best parameter \hat{b} . For the non linear case (right), one sees that values of $Y(b)$ of the model are distributed equidistantly on the curve; therefore, the range of the uncertainties of the parameters can only be given badly. In the linear case (bottom), the values of $Y(b)$ are equidistant and one can directly read out the uncertainty of the parameter.

In the linear case, one finds out from the figure how large the parameter uncertainty is: The estimated value \hat{b} must be increased by Δb , so that a circle can be marked which includes the (uncertain) measurement point (σ):

$$(\hat{b} + \Delta b) \underline{x} = b \underline{x} + \underline{\sigma}. \quad (4.46)$$

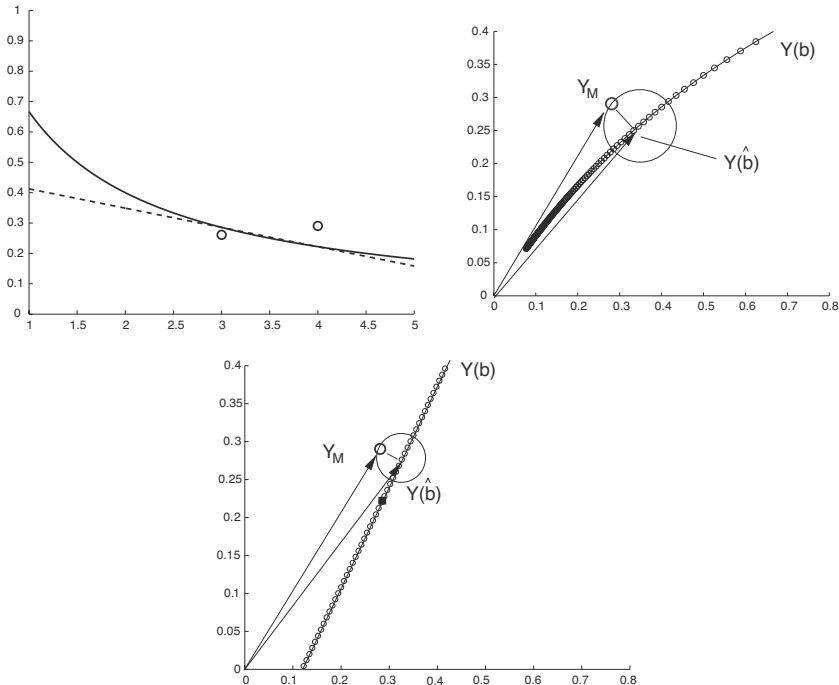


FIGURE 4.5: Geometric representation of the regression. Left: The curves result from the estimation from the two measurement points. The solid lines represent a non linear model while the dashed curve shows a linear model. Right: Representation in a Y_{M1}, Y_{M2} diagram. Shown is the solution locus for a non linear model. The estimated \hat{b} value is as close as possible to the solution locus. Bottom: Representation in a Y_{M1}, Y_{M2} diagram. Shown is the solution locus for a linear model. The circle gives for each case the corresponding uncertainty of the parameters. This uncertainty must be at least large enough that it covers all the measurements within the context of the measurement accuracy. Y_M lies directly on the boundary.

If it is assumed that the estimation is totally good $\hat{b} \approx b$, one obtains: $\Delta b \underline{x} = \underline{\sigma}$, which corresponds formally to the above determined relation for the parameter uncertainty Δb .

4.2.2 Experimental design

One focus in experimental design is to optimize input functions u and measurement time points t_k so as to minimize the parameters' uncertainties. In general, an objective function is used which contains the parameters' uncertainty, e.g., the matrix A or the Fisher information matrix F . One distinguishes the following criteria:

- A—Optimality: $\max \text{Trace}(A) = \max (\lambda_1 + \lambda_2)$
- D—Optimality: $\max \det(A) = \max (\lambda_1 \cdot \lambda_2)$
- E—Optimality: $\max \lambda_{\min}$ or $\min \frac{\lambda_{\max}}{\lambda_{\min}}$.

In the last criterion, there is an attempt to minimize the largest uncertainty. This criterion is often used.

EXAMPLE 4.3 Straight line.

The range of x is sought in which the measurements must be carried in order to obtain a specific standard deviation. The number of the measurement points is confined to be 5 and the measurement variance $\sigma = 2$. The model reads:

$$Y = p_1 + p_2 x = \begin{pmatrix} 1 & x \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}. \quad (4.47)$$

This enables the calculation of the matrix A .

$$A = \frac{1}{\sigma^2} \sum_{i=1}^N \begin{pmatrix} 1 \\ x_i \end{pmatrix} \begin{pmatrix} 1 & x_i \end{pmatrix} = \frac{1}{\sigma^2} \sum_{i=1}^N \begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix}. \quad (4.48)$$

The system has been simulated. Figure 4.6 (left) shows the ellipse for matrix A for measurement points $x \leq 4$ and $x \leq 32$. The representation of matrix A as an ellipse is substantially smaller in the second case (dashed ellipse). On the right plot the progression of the standard deviation over x is shown. Using measurement points in the range $x > 20$, the relative standard deviation of 1% is reached for parameter p_2 (slope). Note that the accuracy of the intercept is not affected by increasing x . This is intuitive since higher numbers of x allow a better estimation of the slope while additional information to estimate the intercept is not given.

If distribution functions are available for parameters that provide a statement concerning the confidence intervals, the above criteria can be expanded accordingly [2]:

- ED optimality: $\max E[\det(A)]$
- EID optimality: $\min E\left[\frac{1}{\det(A)}\right]$
- ELD optimality: $\max E[\ln \det(A)]$

where the expectation value $E[\cdot]$ is calculated.

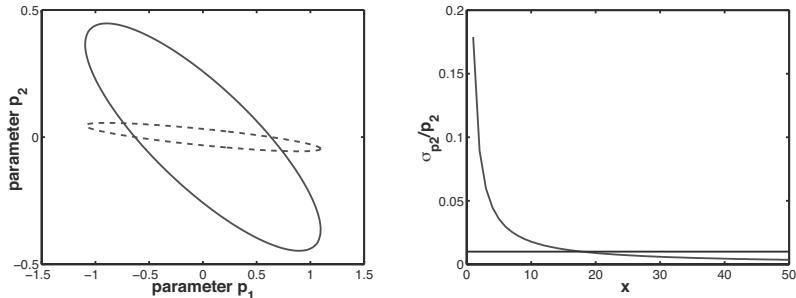


FIGURE 4.6: Left: Ellipse for the range till $x = 4$ (continuous) and for the range $x = 32$ (dashed). Right: Progression of the relative standard deviation of parameters p_2 .

4.3 Dynamic Systems

The procedure for dynamic systems proceeds in analogy to the previous section. In the general case, the derivatives are to be calculated, which are time dependent. Because the systems are non linear in the parameters, the FIM approach provides only a lower boundary. The following simple system can be analytically treated.

EXAMPLE 4.4 *A differential equation of 1st order with a single parameter.*

The following applies here:

$$\dot{x} = b - x \quad \text{mit} \quad x(0) = 0 \quad (4.49)$$

where the measurement error is σ . There are two possibilities for a solution here, as the differential equation can be explicitly solved.

(i) The solution of the differential equation reads:

$$x(t) = b \cdot (1 - e^{-t}) \quad (4.50)$$

Here the solution is a linear function of b . Hence:

$$A = \sum^L \frac{(1 - e^{-t_k})^2}{\sigma^2} \rightarrow \sigma_b = \sqrt{\frac{1}{A}}. \quad (4.51)$$

(ii) Alternatively, the derivative $w = \frac{\partial x}{\partial b}$ can be calculated from the given differential equation. One obtains (see chapter on sensitivities¹):

$$\left(\frac{\partial x}{\partial b} \right) = \dot{w} = \frac{\partial(b - x)}{\partial x} + \frac{\partial(b - x)}{\partial b} = 1 - w. \quad (4.52)$$

¹Exceptionally, here one is pointed to a later chapter.

Here w can be explicitly calculated as well:

$$w = (1 - e^{-t}) \quad (4.53)$$

and, as previously, one obtains for A :

$$A = \frac{1}{\sigma^2} \sum_{k=1}^L w^2 = \sum_{k=1}^L \frac{(1 - e^{-t_k})^2}{\sigma^2}. \quad (4.54)$$

EXAMPLE 4.5 Experimental design: 1st order differential equation.

For an experimental design one can use the given criteria. For the dynamic system:

$$\dot{y} = -a y \quad \text{with} \quad y(0) = b. \quad (4.55)$$

Optimal measurement time points t_1 and t_2 ($t_2 > t_1$) have to be determined, in order to get both parameters a, b with the largest possible accuracy. The D-criterion is used here. The solution of the system reads:

$$y = b e^{-at}. \quad (4.56)$$

For the calculation of the FIM, the derivatives are required. One obtains:

$$\frac{\partial y}{\partial a} = -t b e^{-at}; \quad \frac{\partial y}{\partial b} = e^{-at}. \quad (4.57)$$

The FIM is given by:

$$F = \frac{1}{\sigma^2} \sum_{t_1}^{t_2} t e^{-2at} \begin{bmatrix} t^2 b^2 & -t b \\ -t b & 1 \end{bmatrix}. \quad (4.58)$$

One then inserts both time points and calculates the determinant of the matrix. After some calculations one gets:

$$\det(F) = \frac{1}{\sigma^4} b^2 (t_2 - t_1)^2 e^{-2a(t_1+t_2)}. \quad (4.59)$$

If the system is simulated (Figure 4.7), one obtains the maximum at $t_1 = 0$ and $t_2 = \frac{1}{a}$.

EXAMPLE 4.6 Experimental design: 1st order differential equation (continuation).

Here, the distribution of parameter a is known from a histogram. This information can be used to apply the criteria described above. For the EID-criterion one calculates the term $\frac{1}{\det F}$ to obtain:

$$\frac{1}{\det F} = \frac{\sigma^4}{b^2 (t_2 - t_1)^2} e^{2a(t_1+t_2)}. \quad (4.60)$$

The criterion then minimizes the expectation value that is determined using the integral with the specified distribution π_a :

$$\min \frac{\sigma^4}{b^2 (t_2 - t_1)^2} \int e^{2a(t_1+t_2)} \pi_a da. \quad (4.61)$$

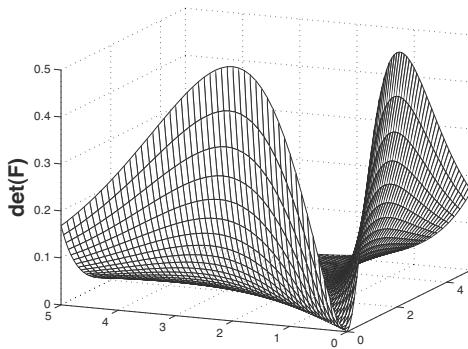


FIGURE 4.7: Simulation study for $\sigma = 1$, $b = 1$ and $a = 1/2$. Note that the function for $\det(F)$ is symmetric and that there is only one maximum, since $t_2 > t_1$ is valid.

4.4 Identifiability of Dynamic Systems

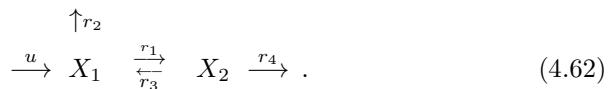
A question that is often overlooked is whether the parameters of a model can be explicitly determined from the available measurements. This problem is formally formulated as follows: A model is identifiable if

$$\text{Model}(\underline{p}_1) = \text{Model}(\underline{p}_2) \Rightarrow \underline{p}_1 = \underline{p}_2.$$

This signifies that when two models exhibit the same behavior, then their parameters must be the same. The following two examples highlight models that are not identifiable.

EXAMPLE 4.7 *A network with two components; only X_1 can be measured.*

The network is given as follows:



One obtains the following system of 1st order differential equations when the reactions are assumed to be of 1st order:

$$\begin{aligned} \dot{X}_1 &= -(k_1 + k_2) X_1 + k_3 X_2 + u \\ \dot{X}_2 &= k_2 X_1 - (k_3 + k_4) X_2. \end{aligned} \quad (4.63)$$

As the differential equations are linear, they can also be written in matrix form:

$$\begin{pmatrix} \dot{X}_1 \\ X_2 \end{pmatrix} = \begin{bmatrix} -a_{11} & a_{12} \\ a_{21} & -a_{22} \end{bmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} u \quad (4.64)$$

with the corresponding coefficients in the system matrix:

$$\begin{aligned} a_{11} &= -(k_1 + k_2); & a_{12} &= k_3 \\ a_{21} &= k_2; & a_{22} &= -(k_3 + k_4). \end{aligned} \quad (4.65)$$

Since only X_1 can be measured, the system is rewritten in terms of a second order differential equation for X_1 . Differentiating with respect to time the differential equation of X_1 leads to:

$$\begin{aligned} \ddot{X}_1 &= -a_{11} \dot{X}_1 + a_{12} \dot{X}_2 + u \\ &= -a_{11} \dot{X}_1 + a_{12} (a_{21} X_1 - a_{22} X_2) + \dot{u}. \end{aligned} \quad (4.66)$$

The quantity X_2 is obtained by algebraic manipulations of the first differential equation from above:

$$X_2 = \dot{X}_1 + a_{11} X_1 - u. \quad (4.67)$$

Inserting the expression one obtains:

$$\begin{aligned} \ddot{X}_1 &= -a_{11} \dot{X}_1 + a_{12} \left(a_{21} X_1 - \frac{a_{22}}{a_{12}} (\dot{X}_1 + a_{11} X_1 - u) \right) + \dot{u} \\ \rightarrow \quad \ddot{X}_1 &+ (a_{11} + a_{22}) \dot{X}_1 + (-a_{12} a_{21} + a_{22} a_{11}) X_1 \\ &= \dot{u} + a_{22} u. \end{aligned} \quad (4.68)$$

The parameters can be collapsed and result in:

$$\ddot{X}_1 + p_1 \dot{X}_1 + p_2 X_1 = \dot{u} + p_3 u. \quad (4.69)$$

Using system identification allows estimating 3 parameters from the measurement of X_1 . There is no direct reverse transformation possible of the 4 k_i parameters from the original system. The system is therefore not identifiable.

EXAMPLE 4.8 A network with two components; only X_2 can be measured.

The network structure looks like the following:



The following system of 1st order differential equations is obtained:

$$\begin{aligned}\dot{X}_1 &= -(k_1 + k_2) X_1 + k_3 X_2 + u \\ \dot{X}_2 &= k_1 X_1 - k_3 X_2.\end{aligned}\quad (4.71)$$

The investigation this time shall proceed in the Laplace domain (see Appendix). One determines for the first differential equation:

$$\begin{aligned}s X_1 &= -(k_1 + k_2) X_1 + k_3 X_2 + U \\ \longrightarrow (s+k_1+k_2) X_1 &= k_3 X_2 + U \\ \longrightarrow X_1 &= \frac{k_3 X_2 + U}{s+k_1+k_2}.\end{aligned}\quad (4.72)$$

One obtains for the second differential equation:

$$\begin{aligned}(s+k_3) X_2 &= k_1 X_1 = k_1 \cdot \frac{k_3 X_2 + U}{s+k_1+k_2} \\ \longrightarrow ((s+k_1+k_2) \cdot (s+k_3) - k_1 k_3) \cdot X_2 &= k_1 U.\end{aligned}\quad (4.73)$$

The transfer function G can then be given as follows:

$$G = \frac{X_2}{U} = \frac{k_1}{s^2 + s (k_1 + k_2 + k_3) + k_2 k_3}.\quad (4.74)$$

The model is not identifiable since two sets of the parameters can be found which lead to the same transfer behavior:

$$\begin{array}{ll}\text{Model 1 } (\hat{k}_1, \hat{k}_2, \hat{k}_3) & \hat{k}_1 = k_1^* \\ \text{Model 2 } (k_1^*, k_2^*, k_3^*) & \hat{k}_2 + \hat{k}_3 = k_2^* + k_3^* \\ & \hat{k}_2 \cdot \hat{k}_3 = k_2^* k_3^*\end{array}$$

The models $M_1(1, 2, 3)$ and $M_2(1, 3, 2)$ with differing parameter values lead to the same results.

The situation must be particularly observed in the presence of a closed loop. The parameters then can not be possibly identified as the next example shows.

EXAMPLE 4.9 Closed loop.

Consider a model with

$$\dot{x} = -ax + bu\quad (4.75)$$

and comparing it with a system which includes a closed loop, where a feedback with

$$u = K(w - x) \quad (4.76)$$

is used, the equation of the modified model then reads:

$$\dot{x}^* = -(a + bK)x^* + bKw = -a^*x^* + b^*w. \quad (4.77)$$

Both models have the same parameters in terms of structure and are therefore indistinguishable.

Exercises

EXERCISE 4.1 Linear regression.

In Equation (4.11) the solution of the optimization problem

$$\max_{\underline{p}} \Phi = \sum_{k=1}^N \frac{(Y_{Mk} - Y_k)^2}{\sigma_k^2} \quad (4.78)$$

was given:

$$\hat{\underline{p}} = A^{-1} \underline{b}. \quad (4.79)$$

By applying the rules for derivatives with respect to vectors, try to verify the calculation for $\sigma_k = 1$.

SOLUTION:

First, the objective function is rewritten in matrix form and the model equation is inserted:

$$\Phi = (\underline{Y}_M - X \underline{p})^T (\underline{Y}_M - X \underline{p}). \quad (4.80)$$

Note that the first term is a row vector and the second one a column vector. We start by calculating the transpose of the first term:

$$\Phi = (\underline{Y}_M^T - \underline{p}^T X^T) (\underline{Y}_M - X \underline{p}) \quad (4.81)$$

and then we expand the product:

$$\Phi = \underline{Y}_M^T \underline{Y}_M - \underline{p}^T X^T \underline{Y}_M - \underline{Y}_M^T X \underline{p} + \underline{p}^T X^T X \underline{p}. \quad (4.82)$$

Now we apply the rules given in the Appendix (the first summand doesn't depend on the parameters):

$$\begin{aligned}
 \frac{d\Phi}{d\underline{p}} &= -X^T \underline{Y}_M - (\underline{Y}_M^T X)^T + 2 X^T X \underline{p} \\
 &= -X^T \underline{Y}_M - X^T \underline{Y}_M + 2 X^T X \underline{p} = 0 \\
 \rightarrow \quad X^T X \underline{p} &= X^T \underline{Y}_M \\
 \rightarrow \quad \underline{p} &= (X^T X)^{-1} X^T \underline{Y}_M. \tag{4.83}
 \end{aligned}$$

EXERCISE 4.2 Parameter accuracy.

Generate data for a straight line with parameter $p_1 = 1$ and $p_2 = 10$. In the chapter, it was shown that the accuracy for p_1 couldn't be improved by increasing the maximal value of x . Try to increase the number of data points N . How large must be N to get $\sigma_{p1}/p1 \approx 30\%$ while the maximal $x = 32$?

Bibliography

- [1] M. Joshi et al. Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems. *Metab. Eng.* 8, 2006, 447-455.
- [2] E. Walter and L. Pronzato. *Identification of Parametric Models from Experimental Data*, Springer-Verlag, 1997.

This page intentionally left blank

Part II

Modeling of Cellular Processes

This page intentionally left blank

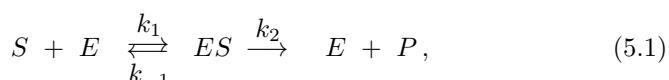
Chapter 5

Enzymatic Conversion

The conversion of substrates into products by means of enzyme is one of the most important intracellular processes. We start by explaining the fundamentals using a simple reaction mechanism that converts a single substrate molecule into a single product molecule. Here, we learn that assumptions are necessary to simplify the rate laws in such a way that they can be used later on in more complex networks.

5.1 Fundamentals of Enzyme Kinetics

Enzymes are specific proteins with catalytic features. Therefore, the steric formation of the amino acids is of great importance. Usually, enzymes are built up by several domains. One distinguishes between catalytic and allosteric centers (centers, which have an affinity for substances, which act regulatory and are not converted). Effectors can bind to such allosteric centers: Inhibitors and activators. They only have an influence on the reaction rate of the enzymes, not on the equilibrium state of the reaction. This is only determined by the converted substrate. Enzymes have a high specificity for substances; therefore a specific enzyme is needed for almost every single process inside the cell. Furthermore, the reaction rate is dependent on the temperature, the pH, metallic ions and the concentration of salt inside the cell. Usually a high number of steps is necessary to convert a substrate. A simple mechanical idea is shown in Figure 5.1. The reaction mechanism of the conversion can formally be written as:



whereupon E is the free and ES is the enzyme substrate complex. The constants of the reaction rates have the following units: $[k_1] \frac{l}{mol\cdot h}$, $[k_{-1}] \frac{1}{h}$, $[k_2] \frac{1}{h}$. An expression of the rate r is demanded, which represents the speed with

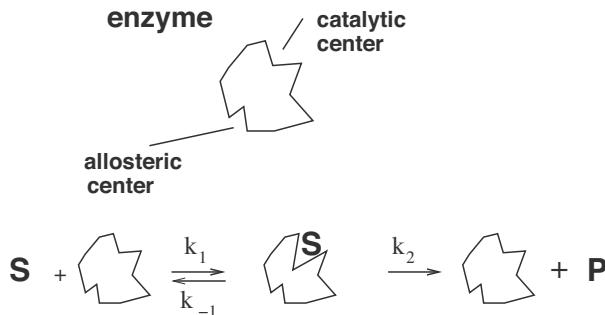


FIGURE 5.1: Simple illustration of the conversion of substrates to products by enzymes.

which a substrate is used and converted into product:¹

$$\frac{dS}{dt} = -r \quad \text{and} \quad \frac{dP}{dt} = r. \quad (5.2)$$

This rate should be as simple as possible to avoid problems if larger networks are considered; however, the experimental observations must be met. One typical observation is that the rate shows not a linear behavior as was discussed above, but shows a hyperbolic course, that is, for larger values of the substrate, the rate goes into saturation. Therefore it will be necessary to use some simplifications to come to such reduced and simplified relations. If one writes down differential equations for all components of the system given above we get:

$$\begin{aligned} \frac{dS}{dt} &= -k_1 E S + k_{-1} E S \\ \frac{dE}{dt} &= -k_1 E S + (k_{-1} + k_2) E S \\ \frac{dES}{dt} &= k_1 E S - (k_{-1} + k_2) E S \\ \frac{dP}{dt} &= k_2 E S, \end{aligned} \quad (5.3)$$

that is, a system of non linear differential equations which can only be solved by numerical integration for more complex systems.

Simplified assumptions can be made by close inspection of the time course of the solution shown in Figure 5.2. One can see that for a large time span

¹Concentrations of components are from now on only referred to with the formula letter of the component ($c_S \rightarrow S$, $c_P \rightarrow P$).

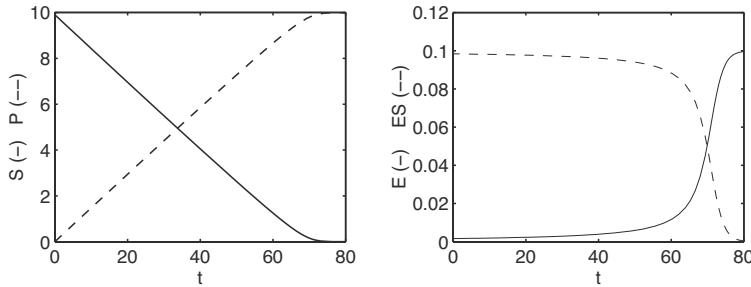


FIGURE 5.2: Simulation of the whole system of differential Equations (5.3) with the initial conditions $S_0 = 10 \gg E_0 = 0.1$. Because of the high affinity of the substrate, the free enzyme quickly changes into the complex and the concentrations of E and ES stay constant for a great period of time. It applies $E(t = 0) = E_0$, although the initial dynamic can not be seen in the illustration.

the concentrations of E and ES are constant and do only temporally change in the beginning and in the end of the simulation experiment. The following simplification is suggested: E and ES act quasi-stationary (steady state, quasi-steady state response (QSSR)). Since the general amount of E_0 does not change, the examination can be focussed on ES . This results in:

$$\boxed{\text{QSSR: } \frac{dES}{dt} = k_1 E S - (k_{-1} + k_2) ES \stackrel{!}{=} 0.} \quad (5.4)$$

Therefore, ES can be reorganized to:

$$ES = \frac{k_1 E S}{k_{-1} + k_2}. \quad (5.5)$$

The above equation can now be inserted into the conservation equation for the total amount of the enzyme E_0 :

$$E_0 = E + ES. \quad (5.6)$$

With $K_M = \frac{k_{-1} + k_2}{k_1}$ it results in:

$$E_0 = E \left(1 + \frac{S}{K_M} \right) \longrightarrow E = \frac{E_0}{1 + \frac{S}{K_M}}. \quad (5.7)$$

In combination with the expressions for the product formation, shown above, it follows:

$$\frac{dP}{dt} = k_2 ES = \frac{k_2 S}{K_M} \frac{E_0}{1 + \frac{S}{K_M}} = r_{max} \frac{S}{K_M + S} \quad (5.8)$$

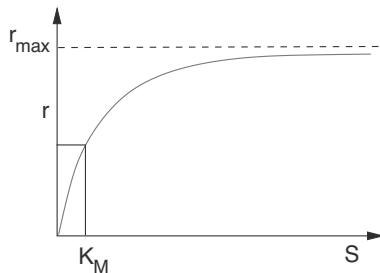


FIGURE 5.3: Graph of the Michaelis-Menten kinetics $r(S)$. The value of K_M can be taken from the x-axis. It applies: $r(S = K_M) = \frac{r_{max}}{2}$.

with $r_{max} = k_2 E_0$. These kinetics are called Michaelis-Menten kinetics. Drawing the graph $r(S)$, one gains the curve which is shown in Figure 5.3.

By looking at Figure 5.4, which demonstrates the simulated system Equation (5.3), with $S_0 \approx E_0$, one can see that the assumption, made above, does not always apply. The trend of the substrate shows that at the very beginning of the simulation process a part of the substrate is quickly bound by the enzyme but the product is not yet formed in the same amount. Thus the substrate is kept by the enzyme. This suggests another simplification of the model: It can be assumed that the reversible reaction of substrate binding remains in the equilibrium state, which is called rapid equilibrium (RE).

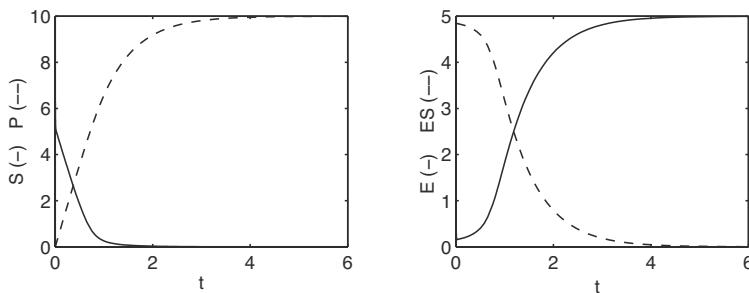


FIGURE 5.4: Simulation of the whole differential equation system (5.3) with the initial conditions $S_0 = 10 \approx E_0 = 5$. In the first time step, a part of the substrate is bound by the enzyme. Then saturation of the enzyme is reached.

The equilibrium reaction in question is: $S + E \rightleftharpoons ES$. The reaction speed for r_1 in the equilibrium state is zero:

$$r_1 = k_1 E S - k_{-1} ES \stackrel{!}{=} 0. \quad (5.9)$$

The examination of the equation leads to a correlation for ES :

$$k_1 E S = k_{-1} ES \quad \longrightarrow \quad ES = \frac{k_1}{k_{-1}} S E. \quad (5.10)$$

Similar to above, one inserts the expression into the equation for r :

$$\frac{dP}{dt} = k_2 ES = k_2 \cdot \underbrace{\frac{k_1}{k_{-1}}}_{1/K_S} S E, \quad (5.11)$$

with the binding constant K_S . Note that the binding constant is the inverse of the equilibrium constant of the reaction. The examination of the conservation equation for the enzyme and the equation for the product formation leads to:

$$E_0 = E + ES = E \left(1 + \frac{S}{K_S} \right) \quad \rightarrow \quad E = \frac{E_0}{1 + \frac{S}{K_S}}; \quad (5.12)$$

for the product formation it follows:

$$\frac{dP}{dt} = r = k_2 E_0 \frac{S}{S + K_S} = r_{max} \frac{S}{S + K_S} \quad (5.13)$$

with $r_{max} = k_2 E_0$. Both simplification methods of the model lead to the same structure of the equation. The difference lies in the definition of the half saturation parameter K . A comparison reveals:

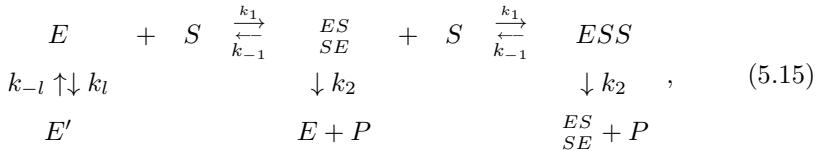
$$K_M = \frac{k_{-1} + k_2}{k_1} \quad k_2 \xrightarrow{k_{-1}} \quad K_S = \frac{k_{-1}}{k_1}. \quad (5.14)$$

This means that the reaction of product formation is very slow in comparison to the degradation of the ES -complex. Without knowledge of the individual kinetic parameter k_j , one cannot predict from the kinetics $r(S)$ which assumption is valid.

5.2 Models for Allosteric Enzymes

The prior introduced basic model shall now be extended to enzymes, which have several substrate binding sites. The term "allosterism" describes the realization that there are more centers for effectors than just the catalytic center, as well as the fact that the intensity of the bonding depends on the number of already bound substrate molecules. For a first model variant, we assume that the enzyme can be in an active state E (substrates can bind)

and in an inactive state E' (substrates cannot bind). If the enzyme has two binding sites and exists in both conformations E and E' , it follows:



whereupon ES and SE are two different kinds of molecules: In case of ES , the substrate binds to the first binding site, in SE to the second. Assuming a fast equilibrium for the reversible reactions, analyzing the conservation equation and evaluating the product formation yields in detail:

$$ES = \frac{E S}{K_S}; \quad ESS = \frac{E S S}{K_S} = \frac{E S^2}{K_S^2}; \quad SE = \frac{E S}{K_S}; \quad E' = \frac{E}{K_L}; \quad (5.16)$$

the conservation equation is:

$$\begin{aligned}
 E_0 &= E + ES + SE + ESS + E' = E \left(1 + 2 \frac{S}{K_S} + \frac{S^2}{K_S^2} \right) + \frac{E}{K_L} \\
 &= E \left(\left(1 + \frac{S}{K_S} \right)^2 + \frac{1}{K_L} \right);
 \end{aligned} \quad (5.17)$$

and product formation is described by:

$$\begin{aligned}
 r &= k_2 ES + k_2 SE + 2 k_2 ESS = 2 k_2 \frac{E S}{K_S} + 2 k_2 \frac{E \cdot S^2}{K_S^2} \\
 &= 2 k_2 \left(\frac{S}{K_S} + \frac{S^2}{K_S^2} \right) E = \frac{2 k_2 E_0 \cdot \frac{S}{K_S} \left(1 + \frac{S}{K_S} \right)}{\left(1 + \frac{S}{K_S} \right)^2 + \frac{1}{K_L}}.
 \end{aligned} \quad (5.18)$$

Comparing the model above with the simple Michaelis-Menten kinetics, one can see the influence of the constant K_L , which also alters the kinetics a lot in its dynamic in the r, S -diagram. Figure 5.5 on the left shows a sigmoid dynamic for different values of K_L by contrast to the hyperbolic dynamic shown above.

For further comparison, a model with cooperativity is considered. It describes an improved binding mechanism of substrate molecules and enzyme, while one or several substrate molecules are already bound. The following mechanism shall be considered, whereupon the ES -complex already covers both variations mentioned above (ES and SE). This becomes clearer by using modified parameters (the factor 2 on the arrows indicates that two substrates

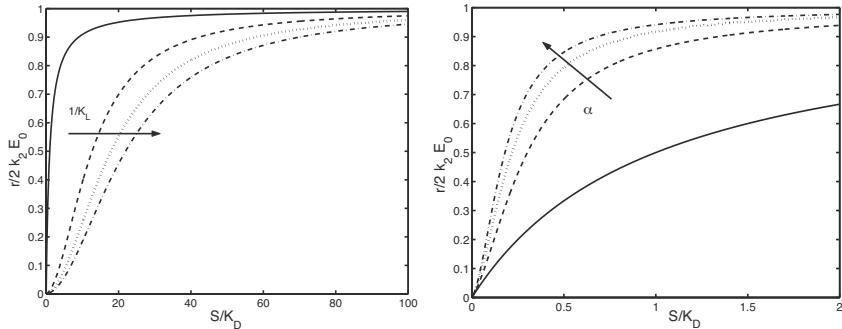
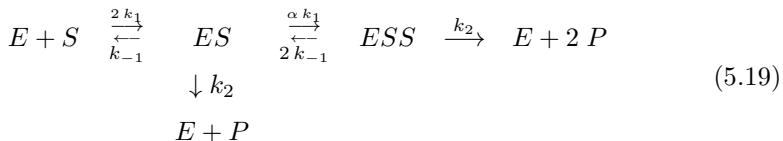


FIGURE 5.5: Simulation studies. On the left: Model with two binding sites and active/inactive conformation. On the right: Model with two binding sites and positive cooperativity. The y axis is scaled to the maximal rate.

can bind/get released):



with $\alpha > 1$ (positive cooperativity). The factor α has an effect when the second substrate molecule binds to the enzyme. If a molecule is already bound, it facilitates the binding of further molecules. The reaction rate increases by the factor α . Applying the conditions for a fast equilibrium again, the single conformations can be described with:

$$ES = \frac{2 E S}{K_S}; ESS = \frac{\alpha E S S}{2 K_S} = \frac{\alpha E S^2}{K_S^2}. \quad (5.20)$$

The conservation equation is:

$$E_0 = E + ES + ESS = E \left(1 + \frac{2S}{K_S} + \frac{\alpha S^2}{K_S^2} \right) \quad (5.21)$$

and product formation is given by

$$\begin{aligned}
 r &= k_2 ES + 2 k_2 ESS = 2 k_2 E \left(\frac{S}{K_S} + \frac{\alpha S^2}{K_S^2} \right) \\
 &= \frac{2 k_2 E_0 \frac{S}{K_S} \left(1 + \frac{\alpha S}{K_S} \right)}{1 + \frac{2S}{K_S} + \frac{\alpha S^2}{K_S^2}}.
 \end{aligned} \quad (5.22)$$

The right part of Figure 5.5 shows the dynamic for different values of α . Again the curve exhibits a sigmoid shape.

If enzymes have more than two binding sites for a substrate, a huge number of intermediate complexes occur and combinatorial methods have to be applied. The choice of the parameter for the reaction rate in Equation (5.19), for instance, anticipates that in ES there are actually two intermediate complexes represented. This has an impact on Equation (5.21) if the factor 2 appears in front of a summand. If the enzyme has three binding sites, the complex ES can appear in three different versions. The complex ESS also has three different manifestations and the complex $ESSS$ appears in only one conformation. Thus, the total amount of the enzyme E_0 can be written in the following way:

$$\begin{aligned} E_0 &= E \left(1 + 3 \frac{S}{K_S} + 3 \left(\frac{S}{K_S} \right)^2 + \left(\frac{S}{K_S} \right)^3 \right) = E \left(1 + \frac{S}{K_S} \right)^3 \\ &= E \sum_{k=0}^3 \binom{3}{k} \left(\frac{S}{K_S} \right)^k \end{aligned} \quad (5.23)$$

whereupon a binomial series is used and the binomial coefficient is defined as:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \quad n \geq k \geq 0. \quad (5.24)$$

Concretely, Equation (5.23) implies that the number of summands depends on the number of binding sites. Thus the binomial coefficient indicates the number of combinatorial possibilities to arrange the substrate molecules among the centers. A more detailed example follows below. A generalization to n binding sites results in the following conformations, which have to be considered:

- Number with only one molecule bound at a time $\binom{n}{1} = n$
- Number with two bounded molecules at a time $\binom{n}{2}$
- Number with n bounded molecules at a time $\binom{n}{n} = 1$

By assuming a fast equilibrium for all partial reactions and if no cooperativity occurs, it follows for the k th intermediate complex:

$$ES^k = \frac{E S^k}{K_S^k}. \quad (5.25)$$

The summation of the different versions of the intermediate complexes results

in the conservation equation:

$$E_0 = E \left(1 + \sum_{k=1}^n \binom{n}{k} \frac{S^k}{K_S^k} \right) = E \left(\sum_{k=0}^n \binom{n}{k} \frac{S^k}{K_S^k} \right) = E \left(1 + \frac{S}{K_S} \right)^k.$$
(5.26)

5.3 Influence of Effectors

Effectors have an influence on the speed of the reaction without being transformed. Activators increase the reaction rate while inhibitors slow it down. The following two schemas shall be analyzed: In the case of the competitive inhibition, the inhibitor competes with the substrate for the enzyme. In the essential activation, a substrate conversion can only occur if an activator is actually present.

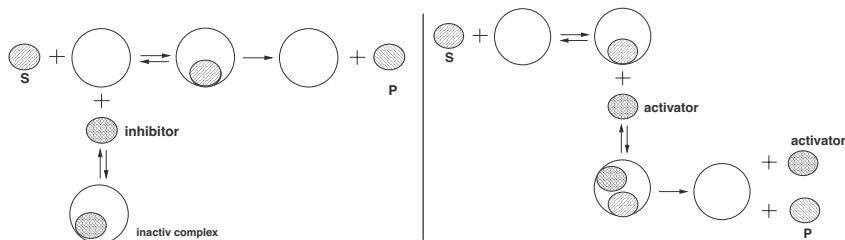
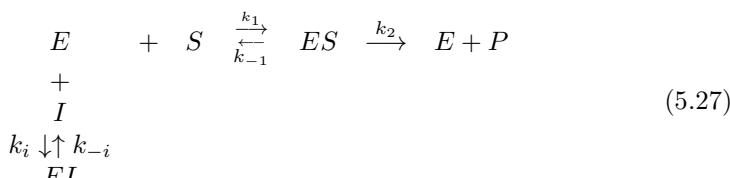


FIGURE 5.6: On the left: Competitive inhibition. The inhibitor competes with the substrate for the enzyme. On the right: Essential activation. The binding of an activator is needed in the case of the conversion of a substrate molecule.

5.3.1 Competitive inhibition

The schema above permits the deduction of the following reaction mechanism:



Assuming a fast equilibrium for the two reversible reactions results in combination with the conservation equation for enzymes in the following equations

for rate r . Analysis of the equilibrium conditions leads to:

$$\begin{aligned} k_1 E S &= k_{-1} ES \rightarrow ES = \frac{k_1}{k_{-1}} E S = \frac{E S}{K_S} \\ k_i E I &= k_{-i} EI \rightarrow EI = \frac{k_i}{k_{-i}} E \cdot I = \frac{E I}{K_I}; \end{aligned} \quad (5.28)$$

the conservation equation is:

$$\begin{aligned} E_0 &= E + ES + EI \\ &= E \left(1 + \frac{S}{K_S} + \frac{I}{K_I} \right) \rightarrow E = \frac{E_0}{1 + \frac{S}{K_S} + \frac{I}{K_I}}; \end{aligned} \quad (5.29)$$

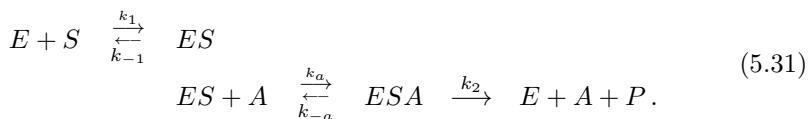
and product formation is given by:

$$\begin{aligned} r &= k_2 ES = \frac{k_2 \frac{S}{K_S} E_0}{1 + \frac{S}{K_S} + \frac{I}{K_I}} \\ &= r_{max} \frac{S}{K_S + S + \frac{K_S}{K_I} I} = \frac{r_{max} S}{K_S \left(1 + \frac{I}{K_I} \right) + S}. \end{aligned} \quad (5.30)$$

Comparing the graph for the standard kinetics with the competitive inhibition, one can only see the influence on the half saturation value (Figure 5.7). The half saturation parameter is a function of the inhibitor concentration. For larger value of the inhibitor the half saturation parameter increases. The maximal rate $r_{max} = k_2 E_0$ is reached again in case of high substrate values.

5.3.2 Essential activation

In the case of the essential activation it is assumed that the binding of an activator to the enzyme-substrate-complex is indispensable for the product formation.



If we assume a fast equilibrium state of both reversible reactions, we get

$$ES = \frac{E S}{K_S}, \quad ESA = \frac{ES A}{K_A} = \frac{E S A}{K_A K_S}, \quad (5.32)$$

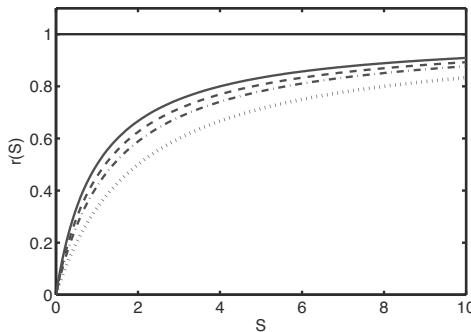


FIGURE 5.7: Influence of an inhibitor I on the rate r . The solid curve shows the Michaelis-Menten kinetics.

and with the conservation equation for enzymes:

$$\begin{aligned} E_0 &= E + ES + ESA = E \left(1 + \frac{S}{K_S} + \frac{SA}{K_S K_A} \right) \\ \rightarrow E &= \frac{E_0}{1 + \frac{S}{K_S} \left(1 + \frac{A}{K_A} \right)} \end{aligned} \quad (5.33)$$

it results, in combination with $r_{max} = k_2 E_0$, in the following relation for the rate r :

$$\begin{aligned} \frac{dP}{dt} &= k_2 EAS = \frac{k_2 \frac{SA}{K_S K_A} E_0}{1 + \frac{S}{K_S} \left(1 + \frac{A}{K_A} \right)} = r_{max} \frac{S \frac{A}{K_A}}{K_S + S \left(1 + \frac{A}{K_A} \right)} \\ &= \frac{\frac{r_{max} A}{K_A} S}{\left(1 + \frac{A}{K_A} \right) \left(\frac{K_S}{1 + \frac{A}{K_A}} + S \right)} = \frac{r'_{max} S}{K'_S + S}. \end{aligned} \quad (5.34)$$

Equation (5.34) is the result of smart transformations. In comparison with the standard kinetics it is possible to determine a maximum rate and a value

for the half saturation again:

$$\begin{aligned} r'_{max} &= \frac{r_{max} \frac{A}{K_A}}{1 + \frac{A}{K_A}} = \frac{r_{max} A}{K_A + A} \\ K'_S &= \frac{K_S}{1 + \frac{A}{K_A}} = \frac{K_S K_A}{K_A + A}. \end{aligned} \quad (5.35)$$

In the case of $A \gg K_A$, r'_{max} approximates the value r_{max} of the standard kinetics, while $K'_S < K_S$ applies (Figure 5.8). Therefore, the activation only shows itself in a low half saturation value. The maximum rate is lower since substrate molecules are bound to the additional enzyme-substrate-activator-complex.

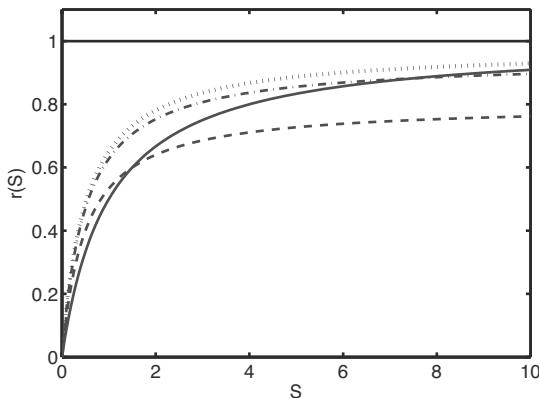
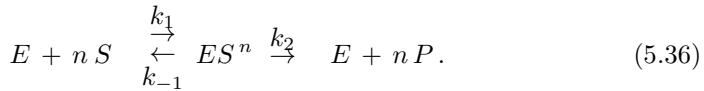


FIGURE 5.8: Influence of activator A on the reaction rate r : The half saturation value as well as the maximal rate are influenced. The solid line indicates the Michaelis-Menten kinetics.

5.4 The Hill Equation

In order to gain the Hill equation, one assumes a semi-empirical reaction approach for enzymes with several binding sites. This equation is often used

in case of empirical studies; there the particular mechanism is unknown:



In the same way as above, one gets:

$$\begin{aligned} ES^n &= \frac{E S^n}{K} \\ \rightarrow r &= n k_2 E_0 \frac{S^n}{K + S^n} = r_{max} \frac{S^n}{K_H^n + S^n}, \end{aligned} \quad (5.37)$$

whereupon the parameter K_H is introduced to ensure a better analysis. In case that one has experimental results, it is possible to estimate the Hill coefficient n and K_H by means of the following rearrangement:

$$\begin{aligned} r &= r_{max} \frac{S^n}{K_H^n + S^n} \rightarrow r K_H^n = (r_{max} - r) S^n \\ \frac{r}{r_{max} - r} &= \left(\frac{S}{K_H} \right)^n \rightarrow \log \frac{r}{r_{max} - r} = n (\log S - \log K_H). \end{aligned} \quad (5.38)$$

Figure 5.9 shows the course of the kinetics and the determination of the Hill coefficient as well as K_H in the case that the data are plotted according to the formula.

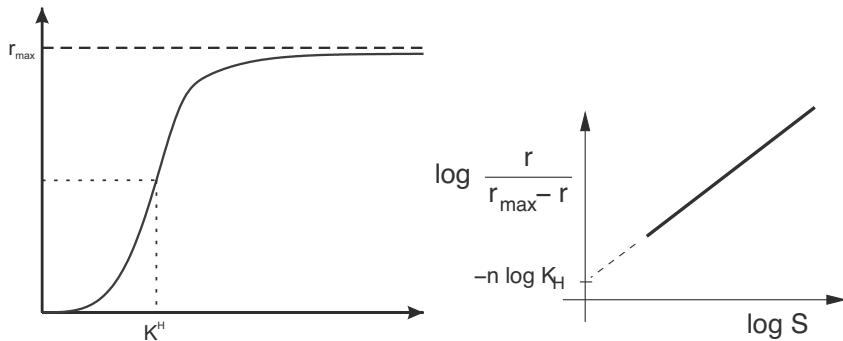


FIGURE 5.9: On the left: Course of the Hill equation as a function of S . On the right: Determination of the Hill coefficient n and K_H from the linear Equation (5.38).

5.5 Multi-substrate Kinetics

Many enzymes need cofactors or convert two substrates (often transfer of specific functional groups of molecules e.g., phosphate groups). Therefore, in literature, many possible mechanisms are described. Three variants are introduced in Figure 5.10:

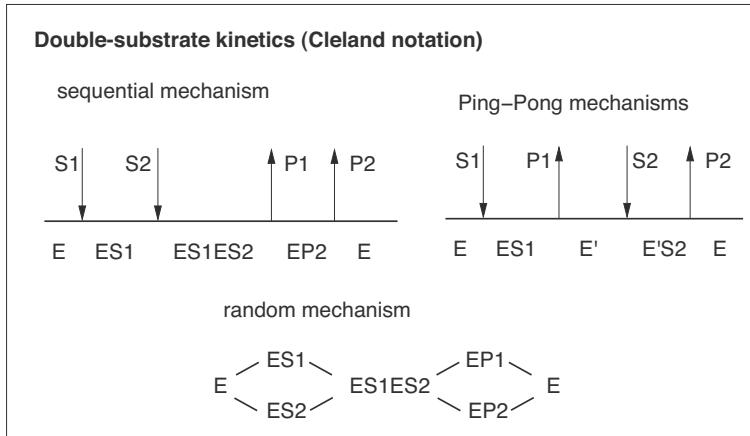


FIGURE 5.10: Three versions of double-substrate kinetics.

The derivation of the velocity-relations in this case is complex, because one usually acts on the assumption of a steady state of the enzyme complexes and it follows a complex equation system. Refer to the corresponding literature [1]. Double-substrate mechanisms can also be much more complicated as the following example reveals.

EXAMPLE 5.1 Double-substrate kinetics.

An enzyme with n centers is considered where substrate 1 and substrate 2 must be bound to the same center in case of product formation. Figure 5.11 shows a selection of possible intermediate complexes, which bind one substrate molecule of each kind if two centers are existent.

To derive a general equation for the determination of the product formation rate r , all complexes $R_{j,i}^{(p)}$ must be summed up whereupon p indicates that p of n centers are occupied by a substrate molecule 1 and a substrate molecule 2. It applies:

$$r = k \sum_{j=1}^n \sum_{i=1}^n \sum_{p=1}^i p R_{j,i}^{(p)}. \quad (5.39)$$

As usual, the Equation (5.39) can be extended by an expression for E_0 and

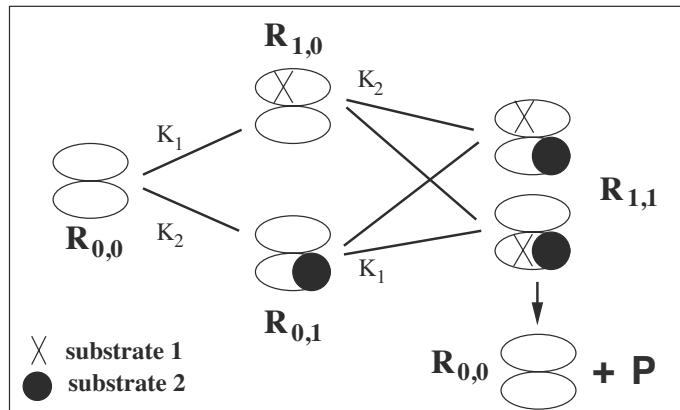


FIGURE 5.11: Intermediate complexes each with one substrate molecule with $n = 2$ centers. The indices are of the size $R_{j,i}$. The binding constants of the substrates are K_1 and K_2 .

one receives:

$$r = k E_0 \frac{\sum_{j=1}^n \sum_{i=1}^n \sum_{p=1}^i p R_{j,i}^{(p)}}{\sum_{j=0}^n \sum_{i=0}^n R_{j,i}}. \quad (5.40)$$

In combination with the expressions above, the numerator looks like:

$$R_{j,i}^{(p)} = \binom{n}{j} \binom{j}{p} \binom{n-j}{i-p} s_1^j s_2^i, \quad (5.41)$$

with $s_1 = \frac{S_1}{K_1}$, $s_2 = \frac{S_2}{K_2}$, whereupon the single binomial coefficients have the following meaning:

$\binom{n}{j}$ Arrangement of j molecules of substrate 1 on n places

$\binom{j}{p}$ Arrangement of p molecules of substrate 2 on j places that are covered with substrate 1

$\binom{n-j}{i-p}$ Distribution of all $i - p$ molecules left of substrate 2 on the remaining $n - j$ places, which are not covered with substrate 1.

The first binomial coefficient is independent from p . It is rearranged like this:

$$\begin{aligned} & \sum_{j=1}^n \sum_{i=1}^n \sum_{p=1}^i p \binom{n}{j} \binom{j}{p} \binom{n-j}{i-p} s_1^j s_2^i \\ &= \sum_{j=1}^n \sum_{i=1}^n \binom{n}{j} \left[\sum_{p=1}^i p \binom{j}{p} \binom{n-j}{i-p} \right] s_1^j s_2^i. \end{aligned} \quad (5.42)$$

For the expression in the brackets apply:

$$\begin{aligned} \sum_{p=1}^i p \binom{j}{p} \binom{n-j}{i-p} &= \sum_{p=1}^i \frac{j}{p} p \frac{j!}{(j-p)! p!} \binom{n-j}{i-p} = \\ j \sum_{p=1}^i \frac{(j-1)!}{(j-p)! (p-1)!} \binom{n-j}{i-p} &= j \sum_{p=1}^i \binom{j-1}{p-1} \binom{n-j}{i-p} = \\ j \left[\binom{j-1}{0} \binom{n-j}{i-1} + \dots + \binom{j-1}{i-1} \binom{n-j}{0} \right] &= \\ j \binom{n-1}{i-1} &= j \frac{(n-1)!}{(n-i)! (i-1)!}. \end{aligned} \quad (5.43)$$

The last line results due to the application of the summation theorem (see Equation (A.31) in the Appendix). Through inserting equation (5.43), one gets:

$$\begin{aligned} & \sum_{j=1}^n \sum_{i=1}^n \binom{n}{j} \left[\sum_{p=1}^i p \binom{j}{p} \binom{n-j}{i-p} \right] s_1^j s_2^i = \\ & \sum_{j=1}^n \sum_{i=1}^n \frac{n!}{(n-j)! j!} j \frac{(n-1)!}{(n-i)! (i-1)!} s_1^j s_2^i. \end{aligned} \quad (5.44)$$

Since j can be expressed as $j = \frac{j!}{(j-1)!}$ and $n! = n (n-1)!$, the last equation yields:

$$\sum_{j=1}^n \sum_{i=1}^n \frac{n (n-1)!}{(n-j)! (j-1)!} \frac{(n-1)!}{(n-i)! (i-1)!} s_1^j s_2^i. \quad (5.45)$$

To solve the sums, the limits must be changed: The sums have to start at

$j = 0$ and $i = 0$. This causes:

$$\begin{aligned}
 & \sum_{j=0}^{n-1} \sum_{i=0}^{n-1} n \frac{(n-1)!}{(n-j-1)! j!} \frac{(n-1)!}{(n-i-1)! i!} s_1^{j+1} s_2^{i+1} \\
 = & n s_1 s_2 \sum_{j=0}^{n-1} \sum_{i=0}^{n-1} \binom{n-1}{j} \binom{n-1}{i} s_1^j s_2^i \\
 = & n s_1 s_2 (1+s_1)^{n-1} (1+s_2)^{n-1}. \tag{5.46}
 \end{aligned}$$

The denominator can be calculated more easily:

$$R_{j,i} = \binom{n}{j} \binom{n}{i} R_{0,0} s_1^j s_2^i \Rightarrow \sum_{j=0}^n \sum_{i=0}^n R_{j,i} = (1+s_1)^n (1+s_2)^n. \tag{5.47}$$

Therefore the kinetics appear as follows:

$$\begin{aligned}
 r &= k \sum_{j=1}^n \sum_{i=1}^n R_{j,i}^{(p)} = r_{max} \frac{n s_1 s_2 (1+s_1)^{n-1} (1+s_2)^{n-1}}{(1+s_1)^n (1+s_2)^n} \\
 &= \frac{n r_{max} s_1 s_2}{(1+s_1) (1+s_2)}. \tag{5.48}
 \end{aligned}$$

This form of the equation conforms with the case of the random-mechanism, if one also considers the number of centers.

5.6 Transport Processes

Cells must transport nutrients and other important substances across the cell membrane. It is built up by a lipid bilayer and contains proteins with a sensor- or transportation function. In case of bacterial cells, the cell membrane is additionally surrounded by the cell wall. Small molecules can easily cross the membrane to enter or leave the cell; larger molecules have more difficulties. We distinguish between the following transportation mechanisms:

- Free diffusion
- Facilitated diffusion
- Active transport
- Group translocation

The first two mechanisms are driven by the concentration gradient between the inner and the outer space of the cell; energy (in form of negative ΔG values) is indispensable for active transport of components into the cell. In the case of group translocation, molecular groups are transferred onto the entering substrate, which is, for example, phosphorylated. The fundament of the calculation is the first law of Fick, which states that the flow of material is proportional to the negative gradient (the change of the concentration c refers to the distance x):

$$r' \sim -\frac{dc}{dx}. \quad (5.49)$$

If the equation is being integrated for constant diffusion coefficients, the rate with which a substance is transported is calculated in the following way:

$$r' = P (c_a - c_i), \quad (5.50)$$

whereupon P is the permeability coefficient (a function of the diffusion coefficient and the thickness of the membrane), c_a the concentration outside of the cell and c_i the concentration inside of the cell. It must be considered that the rate is usually stated in relation to the size of the cellular membrane: $[r'] \text{ mol/m}^2 \text{ h}$. The permeability coefficient therefore has the dimension m/h which matches its velocity. The rate r' is multiplied with the specific surface of the cellular membrane A_Z (\equiv surface of the cell) to gain dimensions, which allow us to set them off against the other processes. One gets:

$$r = P A_Z (c_a - c_i). \quad (5.51)$$

The specific surface of the cell must then be estimated. Following the detailed description given in the signal transduction chapter, a partial differential equation can be derived through a mass balance, which describes the spatial and temporal variations of the components in case that diffusional processes must be considered. In case of active transport processes, extended kinetic models of enzyme kinetics are normally used. Here, the enzyme can appear in several conformations on its outer or inner side and a complex formation occurs consequently (Figure 5.12).

5.7 The Wegscheider Condition

We obtain an equilibrium constant K_{Eq} for the fast equilibrium in each partial reaction. However, these aren't independent in the case of more complex mechanisms. The Wegscheider condition is able to illustrate this. If we consider the mechanism in Figure 5.13, we can see that four equilibrium constants K_i arise from the reactions.

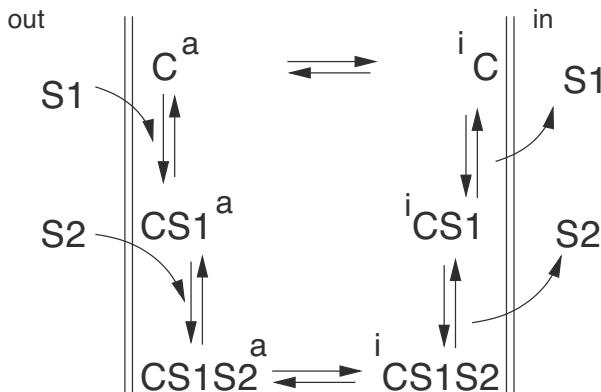


FIGURE 5.12: Reaction scheme of a carrier transporting two substrates S_1 and S_2 into the cell. The order of the binding is sequential.

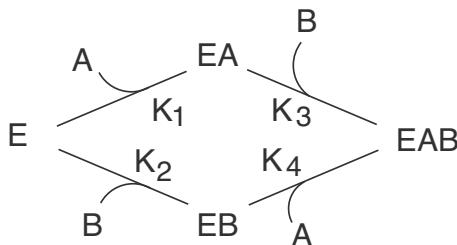


FIGURE 5.13: Illustration of the Wegscheider condition.

The following relations apply for the calculation of the intermediate complexes EA and EB :

$$EA = \frac{E A}{K_1}, \quad EB = \frac{E B}{K_2}. \quad (5.52)$$

The intermediate complex EAB can now be formed via two different pathways: Either via EA through a combination with B or via EB by the combination with A . The following equilibrium state relation applies for both pathways:

$$EAB = \frac{E A B}{K_1 K_3}, \quad EAB = \frac{E A B}{K_2 K_4}. \quad (5.53)$$

Both expressions have to be equal and the following necessity applies: $K_1 K_3 = K_2 K_4$. This means that the knowledge of three parameters determines the fourth one (see also the section on Thermodynamics).

5.8 Alternative Kinetic Approaches

The kinetic approaches, which have been introduced above, have the disadvantage of leading to systems which show a complicated mathematical structure and thus are difficult to handle. In the following, we consider classes of alternative descriptions of reaction kinetics, whereupon the following aspects are important for the analysis [2]:

- The rate is proportional to the amount of enzyme.
- A saturation behavior must be detectable in high substrate concentrations.
- The number of kinetic parameters must be minimal.
- The analytical solutions, for example for the calculation of steady states of large systems, should be as simple as possible.

In the publication cited above, different approaches were summarized and discussed, wherein the kinetics do not fulfill all listed criteria.

Linearization with respect to reference point.

If a Michaelis-Menten kinetic in the form of:

$$r = \frac{k E_0 S}{K + S} \quad (5.54)$$

exists, the function is approximated as a Taylor series (see Appendix). If the Taylor series is interrupted after the first element, one obtains:

$$r = r^R + \left. \frac{dr}{dE_0} \right|_R (E_0 - E_0^R) + \left. \frac{dr}{dS} \right|_R (S - S^R), \quad (5.55)$$

whereupon R signifies that the applicable operational point, which can be a steady state of the system, is inserted. If one calculates the derivations and inserts the steady state, one obtains:

$$\left. \frac{dr}{dE_0} \right|_R = \frac{k S^R}{K + S^R}; \quad \left. \frac{dr}{dS} \right|_R = \frac{k E_0^R K}{(K + S^R)^2}. \quad (5.56)$$

If this is inserted in the equations above, one obtains:

$$\frac{r}{r^R} - 1 = \frac{E_0}{E_0^R} - 1 + \frac{K}{K + S^R} \left(\frac{S}{S^R} - 1 \right). \quad (5.57)$$

The term in front of the brackets on the right side represents the normed derivation of the rate with respect to the substrate and is abbreviated with w

or ϵ . For a small range of enzymes and substrates, one also uses the simplification $y - 1 \approx \ln y$ and we get:

$$\frac{r}{r^R} - 1 = \ln \frac{E_0}{E_0^R} + \epsilon^R \ln \frac{S}{S^R}. \quad (5.58)$$

***n*th-order kinetics.**

If it is assumed that the rate is proportional to the concentration with exponent n , one gets:

$$r = k E_0 S^n \quad (5.59)$$

with n , the ordinal number/stoichiometry of the reaction. Via transposition and referring to the operational point (r^R), one obtains:

$$\frac{r}{r^R} = \frac{E_0}{E_0^R} \left(\frac{S}{S^R} \right)^{\epsilon^R}, \quad (5.60)$$

whereupon the scaled derivation $\epsilon = \epsilon^R$ correlates with the ordinal number n . Using the logarithm, one obtains:

$$\ln \frac{r}{r^R} = \ln \frac{E_0}{E_0^R} + \epsilon^R \ln \frac{S}{S^R}. \quad (5.61)$$

Kinetics based on thermodynamic considerations.

During the considerations with respect to thermodynamics, we introduced the chemical potential μ of a component. In analogy with the calculation of μ , the following form is suggested:

$$r = k E_0 (b + a \ln S). \quad (5.62)$$

Based on the operational point r^R and rearrangement results in the following relationship:

$$\frac{r}{r^R} = \frac{E_0}{E_0^R} \left(1 + \epsilon^R \ln \frac{S}{S^R} \right) \quad (5.63)$$

in combination with the definition for ϵ^R :

$$\epsilon^R = \left. \frac{dr}{dS} \frac{S}{r} \right|_R. \quad (5.64)$$

The kinetic approach is also called Lin-Log approach and is often used in literature because it fulfills the criteria mentioned above in the best way. Figure 5.14 shows the results from calculations, which consider different kinetic approaches from above. It shows the dynamic of Michaelis-Menten kinetics, its linearized form, the n th-order kinetics and the Lin-Log kinetics.

Logarithmized linearization

The kinetic equation of the reaction rate is logarithmized and linearized about a specific value S_0 . One gets:

$$\begin{aligned} r &= r(S, p_j) \quad \rightarrow \quad \log r = \log r(S, p_j) \\ \log r &= \log r_0 + \left. \frac{d \log r}{d \log S} \right|_{\log S = \log S^R} (\log S - \log S^R) \\ &= \log a + b \log S \quad \rightarrow \quad r = a S^b \end{aligned} \quad (5.65)$$

whereupon the two coefficients a and b can be calculated in the following way:

$$b = \left. \frac{d \log r}{d \log S} \right|_{\log S = \log S^R} = \left. \frac{dr}{dS} \frac{S}{r} \right|_{S=S^R} \quad (5.66)$$

$$\log a = \log(r(S^R, p_j)) - \left. \frac{d \log r}{d \log S} \right|_{\log S = \log S^R} \log S^R \rightarrow a = r_0 (S^R)^{-b} \quad (5.67)$$

where parameter b can be interpreted as a measure for the order of the reaction and a as a measure for the rate constant. The equation matches the structure of a kinetic rate with order n .

EXAMPLE 5.2 Michaelis-Menten kinetics

We consider the following equation again:

$$r = r_{max} \frac{S}{K_S + S}$$

and the applicable coefficients lead to:

$$\begin{aligned} b &= \left. \frac{dr}{dS} \frac{S}{r} \right|_{S=S^R} = \frac{K_S}{K_S + S_0} \\ a &= \frac{r_{max}}{K_S + S^R} S^R e^{-\frac{K_S}{K_S + S^R}}. \end{aligned} \quad (5.68)$$

5.9 Thermodynamic of a Single Reaction

We have already seen that the equilibrium constant of a reversible reaction is related to the Gibbs energy under standard conditions (Equation (3.117)). These values can be found in literature or can be calculated based on the

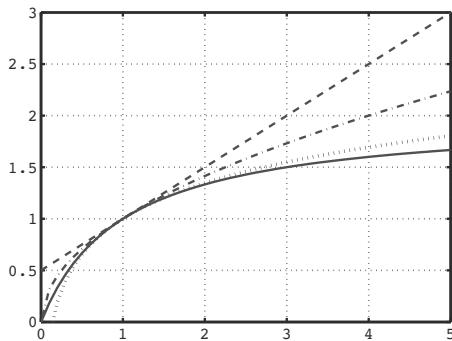


FIGURE 5.14: Michaelis-Menten (solid line), its linearized form (dashed line), kinetics of order n (dash dotted line) and the Lin-Log kinetics (dotted line).

chemical potential of the individual components. Considering a reaction of first order for a reaction



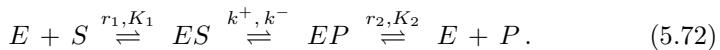
with

$$r = k^+ A - k^- B, \quad (5.70)$$

parameters k^+ and k^- are not independent but related by the equilibrium constant K_{eq} . Therefore, a better representation of the reaction rate would be

$$r = k^+ \left(A - \frac{1}{K_{eq}} B \right). \quad (5.71)$$

This concept can be generalized also for enzyme catalyzed reactions. Let's consider a reversible Michaelis-Menten kinetics:



If we assume as above that reactions r_1 and r_2 are fast and already in equilibrium, the kinetics that describe the conversion from S into P is in standard form:

$$r = E_0 \frac{k^+ \frac{S}{K_1} - k^- \frac{P}{K_2}}{1 + \frac{S}{K_1} + \frac{P}{K_2}}. \quad (5.73)$$

In equilibrium $r = 0$ the equilibrium constant K_{eq} can be obtained:

$$K_{eq} = \left. \frac{P}{S} \right|_{Eq} = \frac{k^+ K_P}{k^- K_S}. \quad (5.74)$$

Here, again we see that the parameters cannot be chosen free (or independently

estimated) but are related. There are several approaches to take into account these considerations. We follow [3] (an alternative approach is described in [4]). A new reaction velocity parameter k^* is defined by:

$$k^* = \sqrt{k^+ k^-}. \quad (5.75)$$

Inserting the expression for the equilibrium constant one finds for the two velocities:

$$k^- = k^* \sqrt{\frac{K_P}{K_S K_{Eq}}} \quad \text{and} \quad k^+ = k^* \sqrt{\frac{K_S K_{Eq}}{K_P}}. \quad (5.76)$$

In this way rate r can be written as:

$$r = E_0 k^* \frac{\sqrt{\frac{K_S K_{Eq}}{K_P}} \frac{S}{K_1} - \sqrt{\frac{K_P}{K_S K_{Eq}}} \frac{P}{K_2}}{1 + \frac{S}{K_1} + \frac{P}{K_2}} \quad (5.77)$$

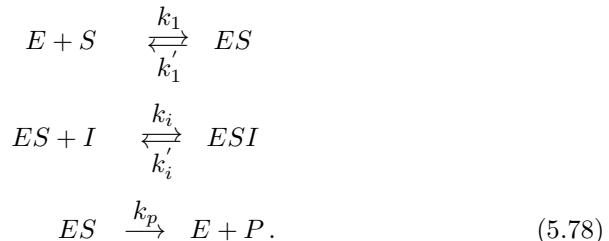
and the thermodynamic constraint given by the equilibrium relation above is satisfied. The approach is valid also for other kinetics and is called convenience kinetics.

Exercises

EXERCISE 5.1 Inhibition kinetics.

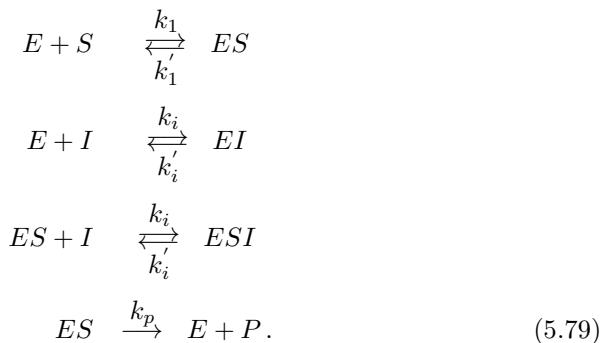
The mechanisms of uncompetitive inhibition and noncompetitive inhibition are to be investigated.

a For uncompetitive inhibition, the inhibitor binds to an ES complex resulting in the following reaction system:



Derive the equation of the product formation for this mechanism, assuming equilibrium of the reversible reactions.

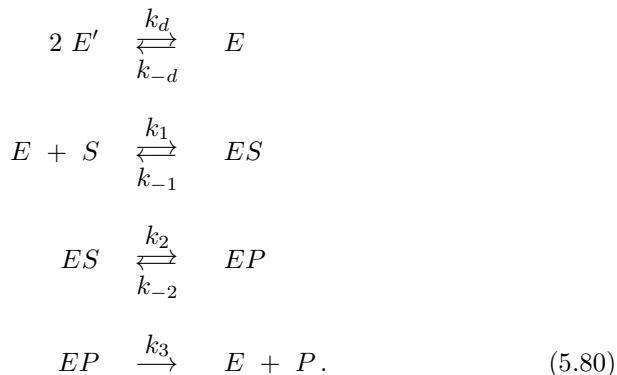
b For the noncompetitive inhibition, the inhibitor binds to a free enzyme and to an ES complex leading to the following reaction system:



Derive the equation of product formation for this mechanism as well, assuming equilibrium of the reversible reactions. What differences arise from comparing this mechanism to the previous one?

EXERCISE 5.2 Enzyme builds a dimer.

Consider the following reaction mechanism that is catalyzed by enzyme E (the reaction volume V is constant). However, the enzyme must dimerize in order to be active (first reaction equation). The monomers are denoted by E' . For the dimer E , two monomers are required (put differently, the active enzyme consists of two molecules E'). The mechanism converts the substrate S into the product P .



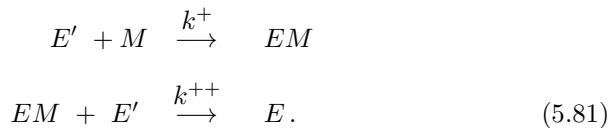
a The reversible reactions are in equilibrium. Give the equilibrium relation for those reactions. Use the following parameters: $K_D = \frac{k_{-d}}{k_d}$, $K_1 = \frac{k_{-1}}{k_1}$ and $K_2 = \frac{k_{-2}}{k_2}$. Give the concentration of E , ES and EP as function of E' , S and parameters.

b Give the conservation equation of the total enzyme amount. Test your result as following: set up the balances equations for the various forms of the enzyme and sum up the equations. Which value must be obtained when the total enzyme amount is constant?

c Using the results so far, provide a relationship for the product formation speed $r_P = \frac{dP}{dt}$ in terms of the total enzyme amount E_0 and S . Assume that K_S is very small (i.e., $E' = 0$, but $\frac{E'^2}{K_S} \neq 0$). What is the significance of having K_S be very small?

d Compare the kinetics with the standard Michaelis-Menten kinetics with the parameters r_{max} and K (half saturation value). How does the half saturation value vary for your version of the kinetics?

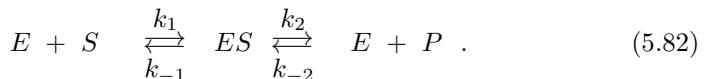
e In the following part of the exercise, the dimerization step will be examined more closely. The following simplified mechanism describes the dimerization by means of a co-factor M :



Provide the differential equations that describe the dynamics of all components (let the reaction volume be constant). The quasi-stationary approximation applies for complex EM . For the components E' and E , give their temporal course $E'(t)$ and $E(t)$ given $E'(t=0) = 1$, $E(t=0) = 0$ and $M = const = M_0$. Plot the graph of $E(t)$ for $k^+ = 1$ and $M_0 = 2$.

EXERCISE 5.3 Reversible reaction.

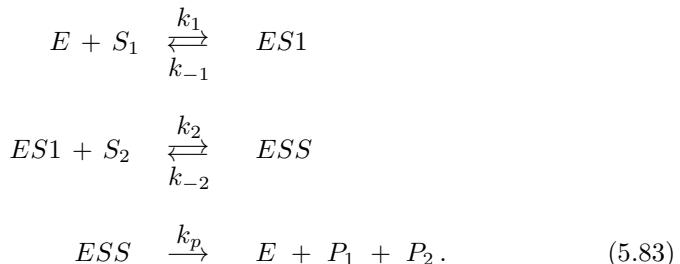
Given the following reactions mechanism:



A relationship for the product formation rate is sought, under the assumption, the enzyme substrate complex ES is in quasi-steady state.

EXERCISE 5.4 Double-substrate kinetics.

The conversion of two substrates S_1 and S_2 proceeds sequentially; thus, the ordering of the connection is an important factor. For the case of an irreversible conversion, provide reaction's rate of the product formation, when the following mechanism is taken as a basis:



It is assumed that both reversible reactions are in equilibrium.

Bibliography

- [1] H. Bisswanger. *Enzyme Kinetics: Principles and Methods*. Wiley-VCH, 2008.
- [2] J.J. Heijnen. Approximative kinetic formats used in metabolic network modeling. *Biotech. & Bioeng.* 91, 2005, 534-545.
- [3] W. Liebermeister and E. Klipp. Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theoretical Biology and Medical Modelling* 3, 2006, 41.
- [4] M. Ederer and E.D. Gilles. Thermodynamically feasible kinetic models of reaction networks. *Biophys. J.* 92, 2007, 1846-1857.

This page intentionally left blank

Chapter 6

Polymerization Processes

Polymerization plays an important role in cellular processes. The synthesis of almost all macromolecules happens by linking monomers in long chains. These processes are made possible by the activity of specific enzymes. The most important processes are the DNA replication, mRNA synthesis (transcription) and protein synthesis (translation). The biological basis of these processes has been explained in Biological Fundamentals. In order to have a basic understanding, the chapter will start with a macroscopic view on transcription. The following sections will introduce a detailed reaction mechanism of the coupling between transcription and translation, which also allows estimating the approximate number of active RNA-polymerases or ribosomes.

6.1 Macroscopic View

Figure 6.1 shows an example of transcription and illustrates the subprocesses and variables that are involved in it. The different phases are (i) Initiation: Docking of the RNA polymerase to the promoter and formation of a complex, which can move along the information strand. (ii) Elongation: Reading of the information stored on the strand and linking of the monomers into a polymer. (iii) Termination: Ending of the process after a stop sequence has been reached. Activators and inhibitors can accelerate and slow down the processes at each stage. All variables are summarized in Table 6.1.

TABLE 6.1: Summary of variables used for a macroscopic view on transcription.

v	Speed of the polymerase <i>Nucleotide/Polymerases</i>
n	Number of polymerases <i>Number Polymerases</i>
l	Length template in nucleotides per transcript <i>Nu/particles mRNA</i>
d	Distance between polymerases Nu

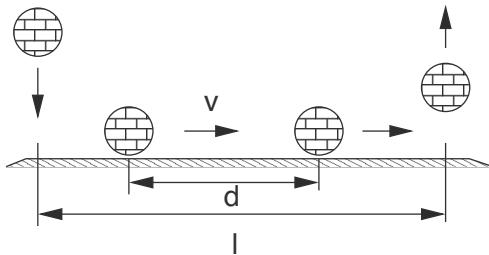


FIGURE 6.1: Schematic illustration of all subprocesses involved in transcription.

The rate r with which one transcript of mRNA is synthesized can be calculated in a few simple steps. First, the distance between two RNA-polymerases can be determined by the ratio $\frac{l}{n} = d$. Since the sought rate r is proportional to v and n while inversely proportional to length l of the template we get:

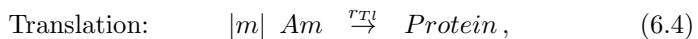
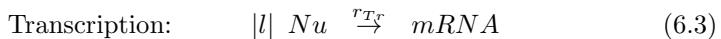
$$r = \frac{v n}{l} = \frac{v}{d}. \quad (6.1)$$

The ratio of speed to distance is also described as frequency f . In an undisturbed process, the rate can be determined by the docking frequency of the RNA polymerase. This frequency can be deduced from a reaction mechanism describing the interaction of the RNA polymerase with the promoter binding site, as shown below. Once the frequency is known, the number of active RNA polymerases can be estimated with

$$n = \frac{l}{d} = \frac{l r}{v}. \quad (6.2)$$

6.2 Microscopic View

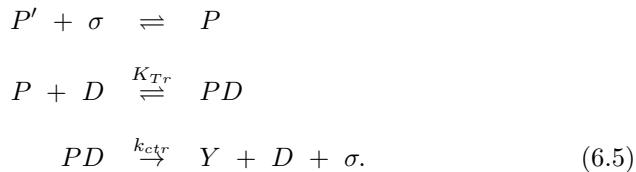
The following detailed reaction mechanism describes transcription as well as translation [1]. Since both processes are linked, it is almost impossible to write down each single step of the reaction. Figure 6.1 does not only show all the sub-processes but is also a quasi-stationary snapshot, since — under idealized conditions — a steady state is reached after the first reading cycle of the RNA-polymerase which results in a constant rate of synthesis. The aim is now to deduce such a rate. Looked at formally, it is based on stoichiometry:



where the dependencies of the monomers (nucleotides Nu and amino acids Am) are not considered here in determining the rate. It is assumed that there are enough nucleotides and amino acids in the cell so the processes are not limited by them.

6.2.1 Calculation of the transcription rate

The initiation of transcription involves the docking of the complex formed by the free RNA polymerase P' and a sigma-factor σ . The sigma-factor recognizes the according binding site D (promoter) of a gene on the DNA. Note, that for the previous examples in the modeling chapter, the DNA was represented in a closed form D_c and an open form D_o that allow transcription. This scheme is now extended to include action of the RNA polymerase in more detail. The following simple mechanism can be observed:



The first step describes the formation of the complex consisting of an RNA polymerase and a sigma-factor σ ; the second step describes the docking of this complex to the promoter (complex PD). During the third step, the PD of the polymerase releases the sigma factor and adopts the conformation of a mobile complex Y that will proceed along the DNA strand. Thereby, the promoter is released and a new initiation can occur. Figure 6.2 shows these processes.

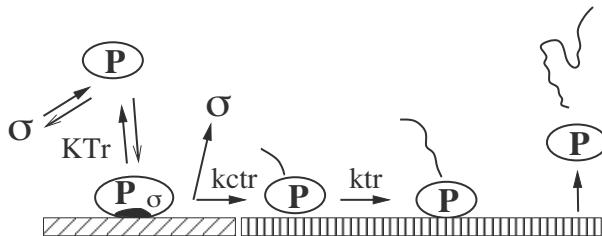
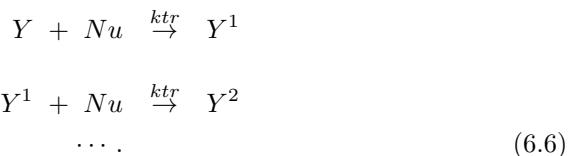


FIGURE 6.2: Detailed mechanism of transcription.

During elongation, the complex Y moves along the DNA and assembles nucleotides. The index of Y^i indicates the number of attached nucleotides:



The process runs until the whole DNA is read and a RNA molecule is generated (termination). The last step to synthesize one molecule mRNA is:



thereby, the RNA polymerase dissociates again. The mRNA is a relatively unstable molecule and undergoes degradation:



To simplify the kinetics, it is assumed that the reversible binding reaction is happening rather fast in comparison to the real polymerization. Thus, a quick equilibrium is assumed for the interaction. Moreover, it can be assumed that a sufficient amount of RNA-polymerase molecules and sigma-factors are present in the cell so they are not limiting: $P_0 \gg D_0 \rightarrow P_0 \approx P$. The following relation results for the equilibrium:

$$PD = \frac{P \cdot D}{K_{Tr}}. \quad (6.9)$$

Applying this equation to all available binding sites D_0 results in:

$$D_0 = D + PD = D + \frac{P \cdot D}{K_{Tr}} \quad (6.10)$$

$$\rightarrow D = \frac{D_0}{1 + \frac{P}{K_{Tr}}} \quad (6.11)$$

$$\rightarrow PD = \frac{P}{K_{Tr}} \cdot \frac{D_0}{1 + \frac{P}{K_{Tr}}} = \frac{P}{K_{Tr} + P} \cdot D_0. \quad (6.12)$$

The binding and dissociation rates of the Y^i complex are very similar. Therefore it can be assumed that they are in a steady state:

$$Y^i = Y^{i-1} = Y \quad \text{and} \quad Y = \frac{k_{ctr}}{k_{tr}} \cdot PD, \quad (6.13)$$

so the production rate for mRNA is finally:

$$r_{Tr} = k_{tr} \cdot Y^{l-1} = k_{tr} \cdot Y = k_{ctr} \cdot PD = k_{ctr} \cdot \frac{P}{K_{Tr} + P} \cdot D_0. \quad (6.14)$$

With this, an equation for the production rate of mRNA is provided. It reflects the frequency with which the polymerase docks to the promoter. The frequency basically depends on the promoter-clearance k_{ctr} and its affinity

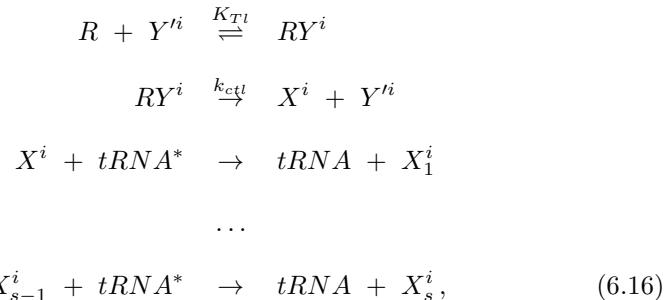
K_{Tr} . To verify the previous analysis, the sum of all active polymerases is determined. They split up among several Y^i complexes:

$$P_{activ} = Y + Y^1 + \dots Y^{l-1} = l \cdot Y = \frac{l}{k_{tr}} \cdot k_{ctr} \cdot PD. \quad (6.15)$$

It results in an equation analogous to the one above ($n = \frac{l}{v} \cdot r$).

6.2.2 Calculation of the translation rate

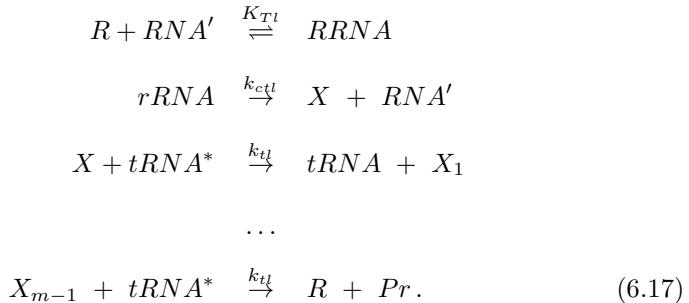
The calculation of translation rates is similar to that shown in the previous section. The subprocesses are analogous to those of transcription with ribosomes R taking over the role of polymerases and attaching to the complexes Y^i (we denote the free site with Y'^i). It must be kept in mind that a protein can only be synthesized if the complete mRNA exists. The load of the tRNA with amino acids is not included in the model but can easily be added. In the reaction scheme, $tRNA^*$ represents the loaded tRNAs and $tRNA$ the unloaded RNAs. The following reaction equation applies for the attachment of the ribosomes R to the complexes Y'^i :



where the length of the protein is $s = \frac{i}{3}$ since three nucleotides code for one amino acid (see biological fundamentals) and a polypeptide can only be as long as the DNA information that has been read so far (index i for the Y complex). The first stage describes the attachment of the free ribosome to a RNA polymerase that is freely moving. The docking of single amino acids can now happen to this complex.

Ribosomes can also attach to a finished mRNA and can read it through. Here, the free RNA binding site is called RNA' and the one blocked by a

ribosome *RRNA*.



The protein *Pr* is only ready after the last amino acid has linked to the complete mRNA. Afterwards the ribosome detaches itself. Analyzing the equation analogous to above, one gets the following equations:

$$\begin{aligned}
 RY^i &= \frac{R}{K_{Tl} + R} \cdot Y^i, \quad RRNA = \frac{R}{K_{Tl} + R} \cdot RNA \\
 X^i &= \frac{k_{ctl}}{k_{tl}} \cdot RY^i, \quad X = \frac{k_{ctl}}{k_{tl}} \cdot rRNA.
 \end{aligned} \tag{6.18}$$

One gets for the rate of synthesis, r_{Tl} :

$$r_{Tl} = k_{tl} \cdot X = k_{ctl} \cdot \frac{R}{K_{Tl} + R} \cdot RNA. \tag{6.19}$$

Again, the rate results as a frequency that is depending on parameters k_{ctl} and K_{Tl} . The sum of all active ribosomes is obtained by considering all strands which are occupied by ribosomes:

- On the finished mRNA:

$$R_{active} = m \cdot X = m \cdot \frac{k_{ctl}}{k_{tl}} \cdot rRNA \tag{6.20}$$

- On the Y^i , and finally for all Y^i complexes:

$$\begin{aligned}
 R_{active}^i &= s \cdot X^i = i \cdot \underbrace{\frac{m}{l} \cdot \frac{k_{ctl}}{k_{tl}}}_{\frac{1}{3}} \cdot RY^i \\
 \sum R_{aktiv}^i &= \frac{m}{l} \cdot \frac{k_{ctl}}{k_{tl}} \cdot \sum i \cdot RY^i \\
 &= \frac{m}{2}(l-1) \cdot \frac{k_{ctl}}{k_{tl}} \cdot \frac{R}{K_{Tl} + R} \cdot Y.
 \end{aligned} \tag{6.21}$$

The number of ribosomes that occupy unfinished mRNA molecules can represent a large percentage of the total active ribosomes.

6.2.3 Combination of equations

The most important equation for the processes mentioned above is summarized again. For the dynamics of mRNA synthesis, the equation

$$R\dot{N}A = k_{ctr} \cdot \frac{P}{K_{Tr} + P} \cdot D_0 - (k_z + \mu) \cdot RNA \quad (6.22)$$

applies that results in

$$\boxed{R\dot{N}A = k_{ctr} \cdot \psi_{tr} \cdot D_0 - (k_z + \mu) \cdot RNA}, \quad (6.23)$$

whereas the transcription efficiency ψ_{tr} can be taken as an indication of the number of promoters that are occupied by RNA polymerases:

$$\boxed{\psi_{tr} = \frac{P}{K_{Tr} + P}}. \quad (6.24)$$

For the dynamics of protein synthesis one applies:

$$\begin{aligned} \dot{Pr} &= k_{ctl} \cdot \frac{R}{K_{Tl} + R} \cdot RNA - (k_{ab} + \mu) P_{rot} \\ &= k_{ctl} \cdot \psi_{tl} \cdot RNA - (k_{ab} + \mu) P_{rot} \end{aligned} \quad (6.25)$$

with the translation efficiency ψ_{tl} , which is defined analogously to the above mentioned transcription efficiency

$$\boxed{\psi_{tl} = \frac{R}{K_{Tl} + R}} \quad (6.26)$$

as the ratio of binding sites that are occupied by ribosomes. Since the RNA usually shows faster dynamics than protein biosynthesis, quasi-stationarity can be assumed for the mRNA and a differential equation for protein biosynthesis is obtained:

$$RNA = \frac{k_{ctr} \cdot \frac{P}{K_{Tr} + P} \cdot D_0}{k_z + \mu} \quad (6.27)$$

$$\dot{Pr} = \frac{k_{ctl} \cdot k_{ctr} \cdot \frac{P}{K_{Tr} + P} \cdot \frac{R}{K_{Tl} + R} \cdot D_0}{k_z + \mu} - (k_{ab} + \mu) Pr \quad (6.28)$$

and finally:

$$\boxed{\dot{Pr} = k' \cdot \eta_{ex}(P, R) \cdot D_0 - (k_{ab} + \mu) Pr} \quad (6.29)$$

with the expression efficiency η_{ex} .

6.3 Influence of Regulatory Proteins (Transcription Factors, Repressors)

The previously presented model can be taken as a standard case, which can be suitably extended by the addition of regulatory effects. Controlling the docking of RNA polymerase is important for the economy of the cell, since it prevents the expression of unnecessary genes and the subsequent waste of resources. In the following, some such extensions of transcription initiation as given in reaction Equation system (6.5) will be presented. Two widely spread regulatory schemes are induction and repression (see Biological Fundamentals). The procedure of deducing the equations can be summarized in the following steps, which generally correspond to the procedure of deducing reaction rates for enzymatic reactions.

- Setting up the reaction network:
 - Usually reversible reactions
 - Considering the number of binding sites
 - Considering the cooperative effects
- Assuming that the reaction happens much faster than the process of elongation:
 - Reversible reactions are in an equilibrium state
 - Setting up the conservation equation for all components (exception: components present in excess)
 - Solving the algebraic system
- The rate of mRNA synthesis expressed as transcription efficiency is set proportionally to the number of RNA-polymerases that are bound to promoters.

6.3.1 Induction model

Induction and repression are the best known mechanisms to describe gene regulation. A rough scheme is shown Figure 6.3. Inducible proteins are only synthesized when they are needed; for example, degradation pathways for specific carbohydrates are only synthesized if their substrate is present in the medium. For the characterization of induction, the following mechanism shall

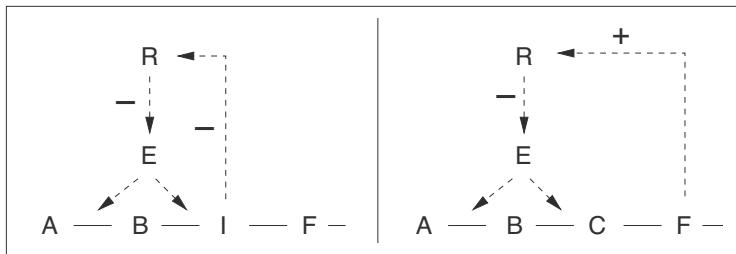
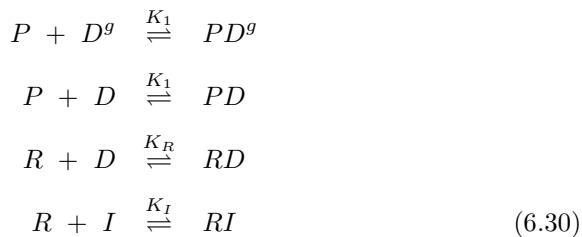


FIGURE 6.3: Left: Induction; an inducer I deactivates a repressor R (transcription factor), the enzyme E is synthesized. Right: Repression; a metabolites D activates a repressor R that blocks transcription of enzyme E .

be considered:



where two groups of genes are included: Group D^g comprises genes that are not affected by repressor R ; group D shows those to which the repressor can attach. The inducer I can deactivate the repressor. The complex RI cannot attach to the binding site. In this way the scheme is more general and can take into account the distribution of the RNA-polymerase on the whole DNA. The following conservation equations must be valid:

$$\begin{aligned}
 P_0 &= P + PD^g + PD \\
 D_0^g &= D^g + PD^g \\
 D_0 &= D + PD + RD \\
 R_0 &= R + RD + RI \\
 I_0 &= I + RI.
 \end{aligned} \tag{6.31}$$

The transcription efficiency adds up to: $\psi_{tr} = \frac{PD}{D_0}$. It is assumed that the RNA polymerase and the inductor are in excess, $P \approx P_0, I \approx I_0$, which

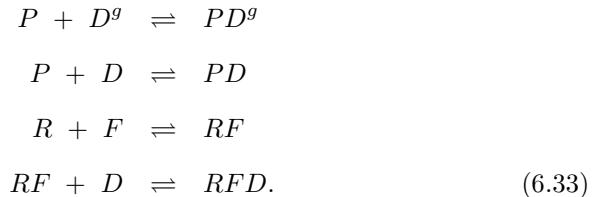
simplifies the system down to:

$$\begin{aligned} D_0 &= D + \frac{P_0 \cdot D}{K_1} + \frac{R \cdot D}{K_R} = D \cdot \left(1 + \frac{P_0}{K_1} + \frac{R}{K_R} \right) \\ R_0 &= R + \frac{R \cdot D}{K_R} + \frac{R \cdot I_0}{K_I} = R \cdot \left(1 + \frac{D}{K_R} + \frac{I_0}{K_I} \right). \end{aligned} \quad (6.32)$$

The rearrangement of the last equation leads to a quadratic equation in R .

6.3.2 Repression model

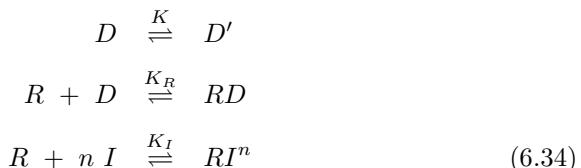
The synthesis of repressible enzymes is affected by the end-product of their metabolic pathway (F) via feedback (Figure 6.3). Analogously to the previous case, the mechanism can be described as follows:



Here, the repressor R cannot attach to the RNA binding site but must be activated through the component F first. The ternary complex RFD then blocks the transcription of the DNA. The transcription efficiency adds up to $\psi_{tr} = \frac{PD}{D_0}$ in this case.

6.3.3 Extended induction model

Solving the above equations shows that expression of genes responds hyperbolically to the concentration of its inducer. However, for some systems, a sigmoidal response is observed. In the following, two alternative mechanisms are analyzed that can result in such behavior. The first model is based on the assumption that the regulator R has several binding sites n for the inducer I . As before, it is assumed that some promoters are already allocated by the RNA polymerase. This can simply be illustrated as a equilibrium reaction between the two states D and D' of the promoter. This leads to three reaction equations:



and three equilibrium conditions:

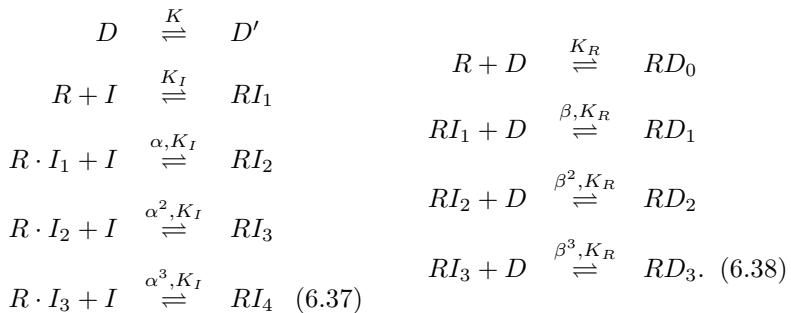
$$D' = \frac{D}{K}, \quad RD = \frac{R \cdot D}{K_R}, \quad RI^n = \frac{R \cdot I^n}{K_I} = R \cdot \left(\frac{I}{K'_I} \right)^n. \quad (6.35)$$

With this, two algebraic equations can be set up:

$$\begin{aligned} D_0 &= D \cdot \left(1 + \frac{1}{K} + \frac{R}{K_R} \right) \\ R_0 &= R \cdot \left(1 + \frac{D}{K_R} + \left(\frac{I}{K'_I} \right)^n \right). \end{aligned} \quad (6.36)$$

The system of equations can be solved numerically to obtain the influence of the inductor for different values of n , as is shown in Figure 6.4. The transcription efficiency is given here by $\psi_{tr} = \frac{D'}{D_0}$.

The second model assumes a successive binding of the molecules to form the complexes RI_1, RI_2, RI_3 , which have decreasing affinities for the binding site D . Finally the repressor protein with all its sites blocked by the inductor, RI_4 , cannot bind at all. The reaction on the left column describes the docking of the inducer (cooperative binding with parameter $\alpha > 1$) to the repressor. The reactions on the right column describe the interaction with the binding site (parameter $\beta < 1$):



The following equations result for the binding site and the regulatory protein:

$$\begin{aligned} D_0 &= D + D' + \sum RD_i \\ R_0 &= R + \sum_{i=0}^3 RD_i + \sum_{j=1}^4 RI_j. \end{aligned} \quad (6.39)$$

Here, combinatorics must again be taken into account. For complexes RI_1 and

RI_2 we get:

$$\begin{aligned} 4 \cdot k_i \cdot R \cdot I &= k_{-i} \cdot RI_1 \longrightarrow RI_1 = \frac{4 \cdot R \cdot I}{K_I} \\ \alpha \cdot 3 \cdot k_i \cdot RI_1 \cdot I &= 2k_{-i} \cdot RI_2 \longrightarrow RI_2 = \frac{6\alpha \cdot I^2 \cdot R}{K_I^2}, \end{aligned} \quad (6.40)$$

and for the remaining complexes:

$$RI_3 = 4\alpha^3 \cdot I^3 \cdot \frac{R}{K_I^3}, \quad RI_4 = \alpha^6 \cdot I^4 \cdot \frac{R}{K_I^4}. \quad (6.41)$$

The RD_j complexes can be calculated in the same way. The following equation results for the particular case $\alpha = 1, \beta = 0$:

$$\begin{aligned} R_0 &= R + \frac{4R \cdot I}{K_I} + \frac{6R \cdot I^2}{K_I^2} + \frac{4R \cdot I^3}{K_I^3} + \frac{R \cdot I^4}{K_I^4} + RD_0 \\ &= R \cdot \left(1 + \frac{I}{K_I}\right)^4 + RD_0, \end{aligned} \quad (6.42)$$

where a binomial series occurs again.

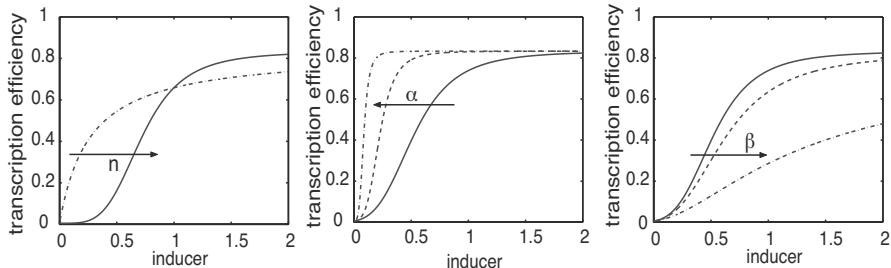


FIGURE 6.4: On the left: Transcription efficiency calculated using the first variant of the model with Equation (6.36). In the middle and on the right: Transcription efficiency calculated using the second variant and Equation (6.39).

Figure 6.4 shows the influence of parameter α and β on the transcription efficiency. The simulation was carried out in the same range of the inducer concentration. The saturation value was already reached for small inducer concentration in the middle plot.

6.4 Interaction of Several Regulators

In bacteria, transcription of a given gene is often subject to regulation by two or more proteins. In such cases, two different regulators can have separate or overlapping binding sites. Furthermore, the number of molecules and the total number of binding sites also plays a role in the model. In the following section, two simple cases will be considered first. Finally, a hierarchical connection is described.

6.4.1 AND – OR gates

Figure 6.5 shows two cases, one where binding sites do not overlap (left) and one where they do overlap (right). For overlapping sites, only one regulator can bind at a time — both regulators are competing for the binding site. In the first case both binding sites are far enough from each other, so the proteins do not interfere and can attach at the same time.

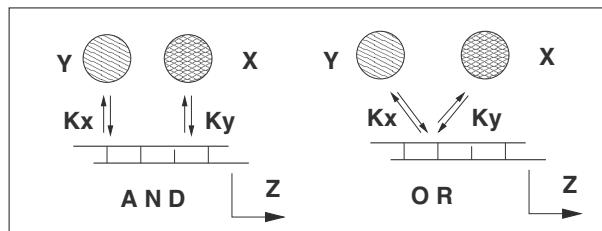
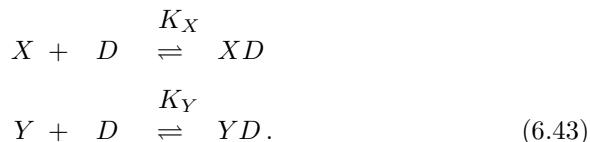


FIGURE 6.5: Left: AND gate. Both transcription factors can attach at the same time. Right: OR gate. The binding sites overlap.

For the diagram on the left, the equation system for the reaction looks like the following:



Since the binding sites are independent, it is possible that both regulators attach at the same time and the following reaction needs to be added:



where XYD is the ternary complex of both proteins and the binding site. For

in the following calculation it may be assumed that for protein synthesis, both regulators must be attached. Therefore the transcription efficiency leads to:

$$\psi = \frac{XYD}{D_0}. \quad (6.45)$$

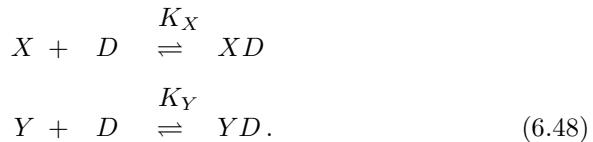
The single terms can usually be calculated with the assumption that the reaction equilibrium is quickly attained. Substituting all terms results in:

$$\psi = \frac{\frac{X}{K_X} \frac{Y}{K_Y}}{1 + \frac{X}{K_X} + \frac{Y}{K_Y} + \frac{X}{K_X} \frac{Y}{K_Y}}. \quad (6.46)$$

This expression can be rearranged and then leads to:

$$\psi = \frac{\frac{X}{K_X}}{1 + \frac{X}{K_X}} \cdot \frac{\frac{Y}{K_Y}}{1 + \frac{Y}{K_Y}} = \psi_x \psi_Y. \quad (6.47)$$

Thus the product of the single activities is obtained under the assumption that the other regulator is not present. This is called an AND gate. For the diagram on the right, the equation system looks like the following:



Since the binding sites can only be blocked by one regulator, only two reactions are to be considered. Analogously to the approach above, the transcription efficiency is:

$$\psi = \frac{XD + YD}{D_0}. \quad (6.49)$$

Assuming that gene expression proceeds if at least one protein has attached, substituting the terms again:

$$\psi = \frac{\frac{X}{K_X}}{1 + \frac{X}{K_X} + \frac{Y}{K_Y}} + \frac{\frac{Y}{K_Y}}{1 + \frac{X}{K_X} + \frac{Y}{K_Y}} = \psi_x + \psi_Y. \quad (6.50)$$

Thus we get the sum of both activities. It is noteworthy that the mixed term presented in the first case does not appear in the denominator of the single activities. Here, each regulator represents an inhibitor for the other. This is called an OR gate.

6.4.2 Hierarchical approach

Regulators are often organized hierarchically. This means that some regulators are very specific for a small group of genes (active/inactive at a specific

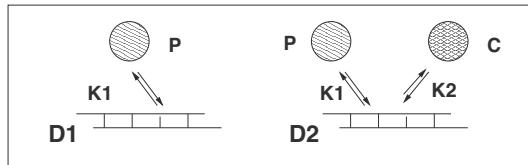


FIGURE 6.6: Schematic illustration of the interaction of two proteins with their respective DNA binding sites. Protein P is in interaction with both binding sites; protein Crp C only with binding site 2.

stimulus), whereas global regulators have an influence on a high number of genes (general starvation signal; stress signal which intensifies the strength of a specific regulator). These hierarchies can be of use in mathematical modeling because there are methods that enable simplifying such networks without introducing too high errors. There are few published methods describing a cellular response in a multigene system. We will now shortly discuss one such method [2]. The following assumptions are used to calculate the transcription efficiency: transcriptional efficiency is proportional to the occupancy of binding sites $\psi_1 = \frac{c_{P,D}}{c_{P,D}^0}$ by RNA polymerases. The influence of a protein, which represses transcription (occupancy rate ψ_2), results from the factor $1 - \psi_2$ (number of unoccupied sites). The influence of an activator with occupancy rate ψ_3 and the activating factor α appears in the equation as $(1 + \alpha \psi_3)$. Therefore the rate of synthesis is proportional to

$$r_{syn} \sim \psi_1 (1 - \psi_2) (1 + \alpha \psi_3). \quad (6.51)$$

Using this method, the fraction of occupied sites ψ_i is calculated while regarding the binding sites as independent and assuming that there are no interactions; just the fractional occupancy of a specific binding site is needed. Upon multiplication of the first two factors of the equation above, the term $\psi_1 \psi_2$ is obtained; that means that the calculation allows the simultaneous binding of both proteins to their binding sites. This would not be the case of the *lac* operon, because it involves an OR gate. Both binding sites overlap and can only be occupied either by the RNA polymerase or the regulating protein *LacI*. A simple example, describing the interaction of two transcription factors, shall be used to introduce a second method [3], that is based on a hierarchical approach. Figure 6.6 shows as an example two types of binding sites D_1 and D_2 .

Comprehensive model

The proteins RNA polymerase P and Crp C interact with their specific control sequence D_2 . Crp is a global regulator in *E. coli* and is involved in the regulation of uptake systems as well as in central metabolism. To be active, Crp must build a complex with a second messenger, cAMP. The structuring genes of D_2 code for an enzyme whose synthesis shall be modelled. Next to

the control sequence D_2 , the protein P interacts with several other binding sites D_1 . The RNA polymerase P' is only active in the σ -factor activated form P .

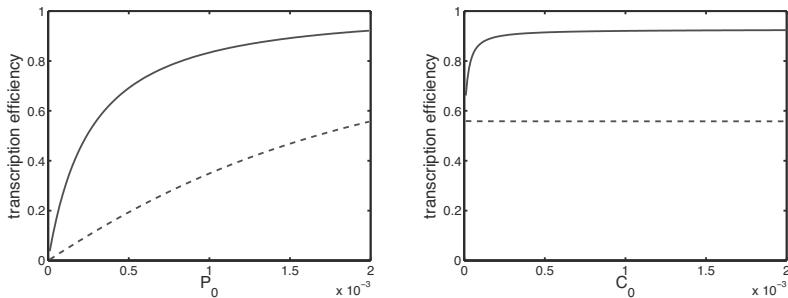
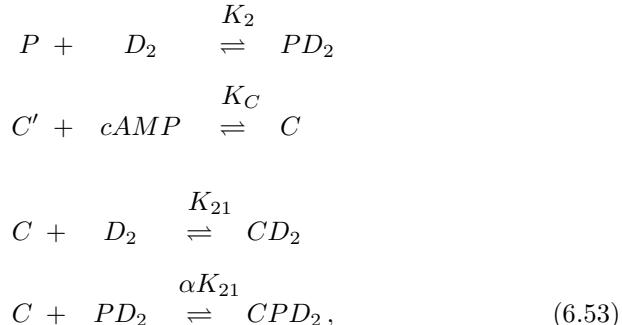


FIGURE 6.7: On the left: Influence of P on both binding sites with fixed C (D_2 solid line, D_1 dotted line). On the right: Influence of C on both binding sites with fixed P (D_2 solid line , D_1 dotted line).

Protein C is activated by metabolite $cAMP$ and can only bind to DNA in its activated form. The following reaction equations apply for the control sequence D_2 with the interaction parameter α :



whereby C and P are positively interacting with each other ($\alpha > 1$). The variables PD_2 and CPD_2 describe the initiation complex if P has linked alone and P and C have linked together to the controlling sequence. The interaction of the controlling sequence D_1 with P can be described with the reaction:



under the assumption of an averaged affinity K_1 of the RNA polymerase to

all promoters D_1 . Usually the affinity K_2 is weaker because the activation happens with transcription factor C . Figure 6.7 shows the behavior of the system to varying amounts of P_0 and C_0 . The polymerase influences both promoters. The transcription efficiency increases for growing amounts of P_0 . Protein C has only an influence on the second binding site (solid line); however, because of the interaction, molecules of P are also needed, which are not available for the first binding site anymore. However, this modification can barely be seen in the plot (dashed line in the right plot). This observation is used for model reduction.

The reduced model

In the top of Figure 6.8 the transformation of a comprehensive model into a block diagram is shown. This is structured into two layers in the lower part of the figure.

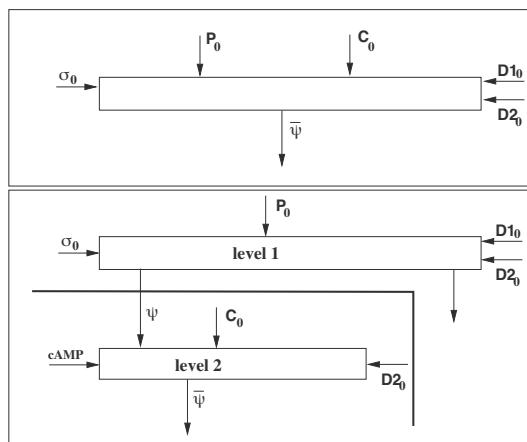
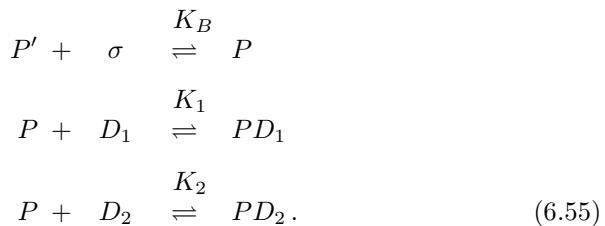


FIGURE 6.8: Exemplary system with two binding sites and two regulatory proteins. Top: Block diagram of the complete model. Below: A single level is assigned to each regulator for easier calculation. Both levels are built up hierarchically and are only connected with each other in one direction.

The reaction network (6.52)–(6.54) is decomposed into these levels by letting protein P represent the upper level and protein C the lower one. The upper level considers the interaction of P with the binding sites 1 and 2 and

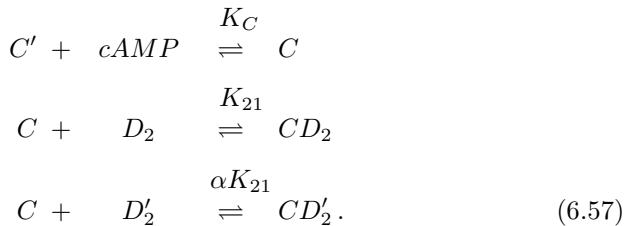
the σ -factor:



The second level represents the activation by *cAMP* and the binding to the free DNA-binding site. The occupation of binding site 2 with the RNA polymerase is described in the upper level and the occupancy rate $\psi = PD_1/D_{10}$ is used as an input for the lower level. Therefore the reaction on the lower level can easily be described with the following reaction, where the polymerase-occupied conformation is referred to as D'_2 :



This sums up the following reaction scheme for the lower level:



Here, CD'_2 represents the ternary complex of P , C and the binding site 2. The simulation of the reduced model in two steps leads to the same results shown in the figure above. There is barely any difference. Therefore, the method enables simplification and modeling of big networks. The complexity of solving the algebraic equations and getting a numeric result can thus be significantly reduced.

6.5 Replication

Bacterial DNA is a circular, double stranded molecule made of four deoxyribonucleotides, dATP, dTTP, dGTP and dCTP. The two opposing strands are always linked through hydrogen bonds, where only the base combinations of adenine-thymine and guanine-cytosine are allowed. The genome of *E. coli* consists of approx. 4,720,000 base pairs. The order of the base pair combinations represents the genetic information. A specific number of base pairs, a

gene, represents the blueprint for a protein. The genes do not all lie in a row but are split by noncoding regions (e.g., promoter, operator), which encode further information (see Figure 6.9).

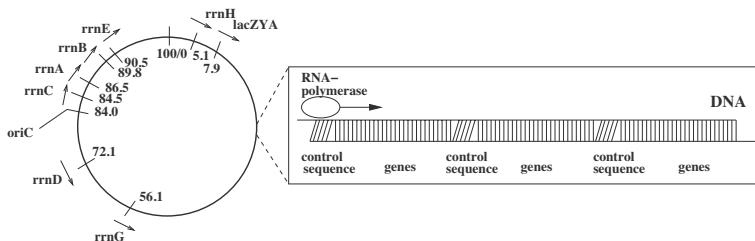


FIGURE 6.9: Genetic card of *E. coli*. The origin of the DNA replication (origin, *oriC*) is at position 84. The structural genes are divided by sections of control sequences where each structural gene is controlled by at least one control sequence. In exceptional cases, the control sequences may also be within the structural gene.

The period of replication comprehends the time of DNA synthesis (time *C*) and the time after the synthesis until cell division (time *D*). The genetic information must be available in double amounts before it can be distributed between the daughter cells. In experiments, a replication time of only 20 minutes can be observed and it must be assumed that the process of replication already restarts before the running replication phase has finished since the minimum time *D* for replication has been established at around 20 minutes. That means that the phases C and D overlap (see Figure 6.10).

The process of replication always starts at a fixed point, the origin, and is catalyzed by a protein called DNA-polymerase. The molecule is split up in 100 minutes. The initial point of a gene can be exactly quantified. Starting from the origin, the new strand is synthesized in both directions. The cell division can occur when two copies of the DNA are available. In cells with high replication rates, the genes close to the origin are available in numerous copies, as can be seen in Figure 6.10. The *rrn* genes, coding for ribosomal RNA and which are encoded close to the origin, are available in almost 8 times the usual amount at a replication rate of $\mu = 2.08 \text{ l/h}$. The following relation by Cooper [4] was stated for determining the number of DNA copies n_{DNA} at a growth rate μ in connection with the time *C* and *D*.

$$n_{DNA} = \frac{1}{C\mu} \left(2^{\frac{(C+D)}{\ln 2}\mu} - 2^{\frac{D\mu}{\ln 2}} \right). \quad (6.58)$$

Regarding the position *x* of a specific gene on the DNA, one gets the following equation, again by Cooper, for determining the number of templates n_x of a specific gene:

$$n_x = 2^{\frac{(D+C(1-x))\mu}{\ln 2}} \quad 0 \leq x \leq 1, \quad (6.59)$$

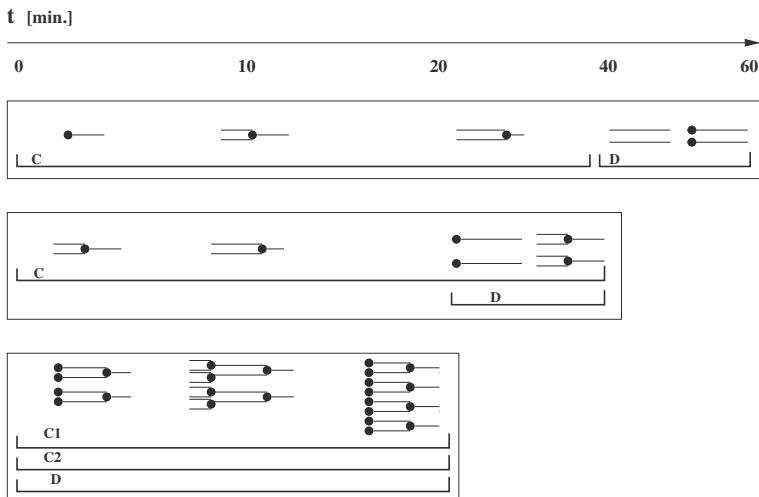


FIGURE 6.10: Illustration of the number of DNA molecules with different doubling times τ . Line 1 shows the temporal development of the process of replication for the doubling time $\tau = 60$ minutes. For lines 2 and 3, $\tau = 40$ minutes and $\tau = 20$ minutes. The black dot indicates the movement of the DNA-polymerase along the DNA. After replication, the cell needs at least 20 minutes (time D) till separation is completed. This is the smallest possible time of replication τ . Time C for duplicating DNA takes between 40 to 70 minutes. As can be seen in line 2, the cell already has duplicated a part of its DNA at $t = 0$ and starts a new replication after 20 minutes. Time C and D overlaps. In a replication time of 20 minutes, the cell needs two generations to read the DNA once (time C_1 and C_2). Parts of the DNA that are close to the origin are available in about eight times the amount (this figure was adapted from S. Cooper: *Bacterial Growth and Division*; Copyright Elsevier, 1991).

whereby $x = 0$ is the site of origin and $x = 1$ the terminal site. The numerical value on the gene map must be converted adequately. Figure 6.11 shows the development of the average number of copies of the DNA as well as those of two specific genes. The gene *rrnC* is close to the origin; gene *lac* a bit further down (see Figure 6.9).

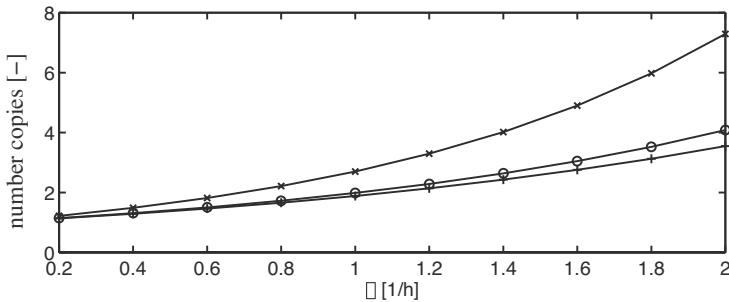
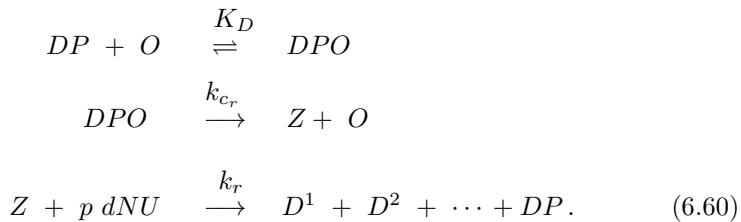


FIGURE 6.11: Number of copies of DNA (o), gene *rrnC* (x) on position 84.5 and of gene *lac* (+) on position 7.9 in dependence from the growth rate μ . Genes near the origin appear in higher numbers inside the cell.

When modeling the process, the same approach as above is used. One can act on the assumption that the DNA-polymerase *DP* interacts with the origin *O* (complexes *DPO* and *Z*) and the synthesis of DNA and therefore new genes D^i follow this scheme:



K_D is the affinity of the DNA-polymerase for the origin; k_{cr} and k_r are the rate constants. For DNA synthesis the reaction rate r_{Re} can be stated analogous to the above like this:

$$r_{Re} = k_r c_Z. \quad (6.61)$$

The rate k_r can be estimated from the time C for DNA-synthesis (see above). However, this depends on the growth rate μ .

Exercises

EXERCISE 6.1 *Repression.*

Write down all algebraic equations for the repression model according to Figure 6.3. Assume hereby that RNA polymerase P and metabolite F are in excess.

EXERCISE 6.2 *Hierarchical network.*

Many genes are controlled by multiple regulators. The number of the binding sites of a gene at which the regulators bind varies substantially. This number shall be investigated in the following exercise which describes a procedure of model reduction that can be found in literature. Some complexes of low concentration that consist of a regulator and a binding site will be neglected.

Given is the following scheme of a regulation. Here, the regulator $R1$ binds only to the site $D1$ (represented by a small group of genes), while regulator $R2$ binds to $D1$ and $D2$ (represented by a large group of genes). Therefore, the count of $D2$ is much larger than $D1$. The binding of both regulators to $D1$ is independent from one another; the binding of the second regulator increases the affinity by a factor of α .

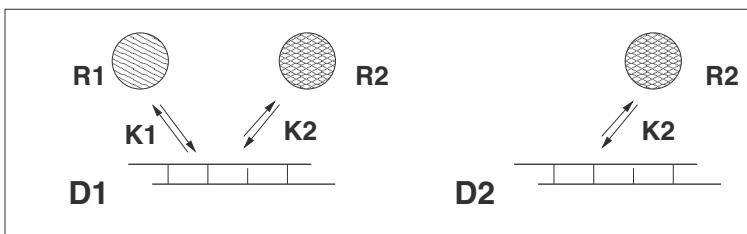


FIGURE 6.12: Interaction of two regulators.

a Set up a reaction scheme and then derive the algebraic equations. Which complexes must be accounted for? The algebraic state variables are $r1$ (free regulator $R1$), $r2$ (free regulator $R2$), $d1$ (free binding site $D1$) and $d2$ (free binding site $D2$). The first output is the one with the regulator $R2$ and with the occupied part ψ_1 at the binding sites $D1$ (occupied by $R1$ and $R2$). Plot the course of ψ_1 as a function of $R1$ ($R2$ is constant) and as a function of $R2$ ($R1$ is constant). The second output is the one with the $R2$ occupied part ψ_2 at the binding site $D2$. Plot the course of ψ_2 as a function of $R1$ and $R2$. What observation can be made for the model reduction?

b Consider the following simplified scheme: Only regulator $R2$ is considered that binds to $D1$ and to $D2$ as well. Set up the corresponding algebraic system.

The part ψ_{r1} with the $R2$ occupied binding site at $D1$ now is

$$\psi_{r1} = \frac{R2D1}{D10} . \quad (6.62)$$

Provide a relationship of the behavior

$$K = \frac{D1}{R2D1} \quad (6.63)$$

which only depends on ψ_{r1} . The simplified scheme for the regulator $R1$ looks like the following: The binding site $D1$ now exists as a free site and as a preoccupied binding site $D1^*$ by $R2$. The behavior is denoted by K and was calculated earlier. $R1$ can now according to the following reaction scheme, bind to $D1$ as well as to $D1^*$:



Calculate $\psi_1^{reduced}$ just like in the output model. To what extent was model reduction carried out? Determine the error arising from the model reduction.

Bibliography

- [1] A. Kremling. Comment on mathematical models which describe transcription and calculate the relationship between mRNA and protein expression ratio. *Biotech. & Bioeng.* 96, 2007, 815-819.
- [2] S. B. Lee and J. E. Bailey. Genetically structured models for *lac* promoter-operator function in the *Escherichia coli* chromosome and in multicopy plasmids: *lac* operator function. *Biotech. & Bioeng.* 26, 1984, 1372-1382.
- [3] A. Kremling and E. D. Gilles. The organization of metabolic reaction networks: II. Signal processing in hierarchical structured functional units. *Metab. Eng.* 3, 2001, 138-150.
- [4] S. Cooper. *Bacterial Growth and Division*. Academic Press, 1991 .

This page intentionally left blank

Chapter 7

Signal Transduction and Genetically Regulated Systems

Cellular systems quickly respond to changing environmental conditions. The response is achieved through activation and deactivation of metabolic pathways or chemotactical movements of the cells. In order to initiate or terminate the translation of DNA, a signal must be transmitted through the cellular membrane into the cell. Thus, signal transduction processes act as connectors between a stimulus and gene expression.

Signal transduction systems show an interesting dynamic behavior. Possible characteristics include switches, amplifications, adaptivity or oscillations. One or more of these attributes are needed to enable specific cellular functions. Signaling pathways are often linked with processes of gene regulation, where a regulated protein itself regulates genes. Control of transcription/translation will not be considered in this chapter (this was done in the previous chapter), yet the corresponding proteins will be shown in the figures. To illustrate genetic regulation, genes will be denoted by lower case italic letters, while upper case letters denote the protein/regulator. First, we will introduce and analyze a few basic signaling networks. Afterwards, genetic networks are considered.

In this chapter mathematical tools will be used to determine steady states of the system as well as stability properties. For an introduction into the analysis on non linear dynamical systems, see the Appendix.

7.1 Simple Schemes of Signal Transduction

Some basic networks from the literature [1] will be introduced. Tyson and coworkers present a number of very illustrative examples for small network schemes. They show basic mechanisms which are also found imbedded in more complex networks.

Activation Model

For the network in Figure 7.1, we can write down the following equations:

$$\begin{aligned}\dot{R}^P &= k_1 \cdot S \cdot R - k_2 \cdot R^P \\ \dot{R} &= -k_1 \cdot S \cdot R + k_2 \cdot R^P.\end{aligned}\quad (7.1)$$

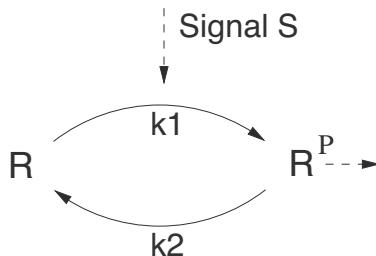


FIGURE 7.1: Simple signaling network.

Since the conservation law $R^P + R = R_0$ applies, the second equation can be eliminated. With

$$\dot{R}^P = k_1 \cdot S \cdot (R_0 - R^P) - k_2 \cdot R^P \quad (7.2)$$

one obtains the following stationary signal response curve (steady state).

$$R_{ss}^P = \frac{S \cdot R_0}{S + K_{12}} = \frac{S \cdot R_0}{S + \frac{k_1}{k_2}}. \quad (7.3)$$

The equation has the same form as the Michaelis-Menten kinetics (see Figure 7.2 on the left). Its curve can also be constructed graphically (Figure

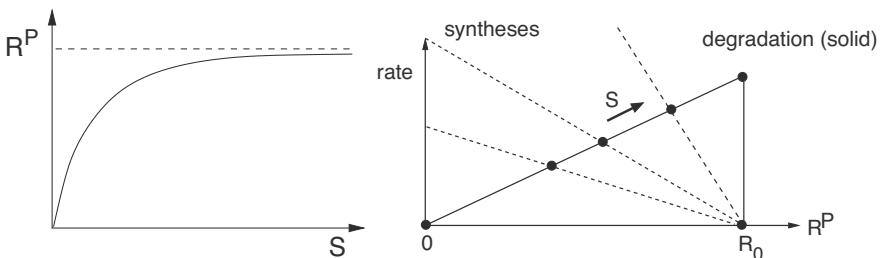


FIGURE 7.2: Left: Signal response curve. Right: Graphical construction of the signal response curve. The continuous line shows the degradation rate in dependence of R^P . It is independent of the signal. The dotted lines are rates of synthesis depending on signal S . With a stronger signal, the gain increases. However, larger values of S lead only to marginal alterations of R^P .

7.2 on the right) by plotting the synthesis rates against the degradation rates. By identifying the points of intersection, the stationary points are determined. Now, they can be transferred to a signal response diagram. A more detailed look on the model can be obtained if an enzyme kinetics approach is used for the reaction rates instead of first order reactions. With the conservation law for R_0 , one gets:

$$\dot{R}^P = \frac{k_1 \cdot S \cdot R}{K_1 + R} - \frac{k_2 \cdot R^P}{K_2 + R^P} = \frac{k_1 \cdot S \cdot (R_0 - R^P)}{K_1 + R_0 - R^P} - \frac{k_2 \cdot R^P}{K_2 + R^P}. \quad (7.4)$$

The graphical method described in Figure 7.3 shows that the signal response diagram has a sigmoid shape. The calculation of the steady states is more complex here. The above equation can be written as:

$$\frac{k_1 \cdot S \cdot (R_0 - R^P)}{K_1 + R_0 - R^P} = \frac{k_2 \cdot R^P}{K_2 + R^P}. \quad (7.5)$$

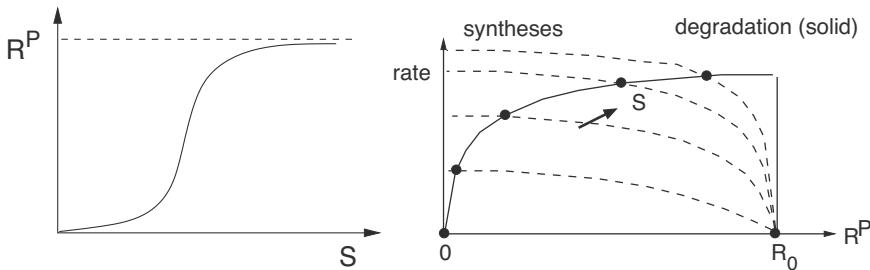


FIGURE 7.3: Switching behaviour — “buzzer.” The curve indicates the degradation rate in dependence of R_P . It is independent of the signal. The dotted line represents the signal-dependent rate of synthesis. All graphs have the same half saturation K_1 .

In a more general form, for example considering also a possible input in the back reaction, it results in:

$$\frac{v_1(S) \cdot (R_0 - R^P)}{K_1 + R_0 - R^P} = \frac{v_2(S) \cdot R^P}{K_2 + R^P} \quad (7.6)$$

where v_1 and v_2 now represent the maximum rates. Thus,

$$\begin{aligned} v_1 \cdot (R_0 - R^P) (K_2 + R^P) &= v_2 \cdot R^P (K_1 + R_0 - R^P) \\ \rightarrow v_1 \cdot (1 - r^P) (K_2^* + r^P) &= v_2 \cdot r^P (K_1^* + 1 - r^P). \end{aligned} \quad (7.7)$$

The second curve can be obtained through division with $R_0 \neq 0$. The quantities K^* represent the scaled Michaelis-Menten constants. Dividing by the right-hand side, one gets:

$$(v_2 - v_1) r_p^2 - (v_1(K_2^* - 1) + v_2(1 + K_1^*)) r_p + v_1 K_2^* = 0, \quad (7.8)$$

which is a quadratic equation with abbreviations:

$$p = \frac{v_1(K_2^* - 1) + v_2(1 + K_1^*)}{v_2 - v_1} \quad \text{and} \quad q = \frac{v_1 K_2^*}{v_2 - v_1}. \quad (7.9)$$

Solving the quadratic equation leads to (the plus solution of the quadratic equation leads to values for $r_p > 1$):

$$r_p = \frac{p}{2} - \frac{1}{2}\sqrt{p^2 - 4q}. \quad (7.10)$$

Resubstituting p and q, one gets:

$$r_p = \frac{P - \sqrt{P^2 - 4(v_2 - v_1)v_1 K_2^*}}{2(v_2 - v_1)} \quad (7.11)$$

$$\text{with } P = v_1(K_2^* - 1) + v_2(1 + K_1^*).$$

This form is disadvantageous, since a singularity occurs for $v_1 = v_2$. By extending it, the Goldbeter-Koshland function G can be obtained:

$$G := r_p = \frac{2v_1 K_2^*}{P + \sqrt{P^2 - 4(v_2 - v_1)v_1 K_2^*}}. \quad (7.12)$$

In many cases, G is needed to analyze the behavior of systems. Below, G is used in both a feedback system and an oscillating system.

Adaptive behaviour

Figure 7.4 shows a signal transduction system where the signal affects two sub-paths. Both paths are linked by the interaction of a component X and the degradation of R .

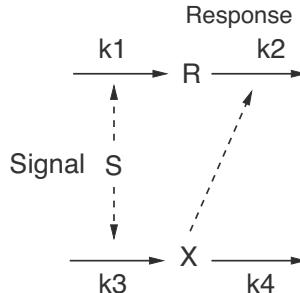


FIGURE 7.4: Loop that leads to adaptive behavior (Sniffer).

The corresponding equations are:

$$\begin{aligned}\dot{R} &= k_1 \cdot S - k_2 \cdot X \cdot R \\ \dot{X} &= k_3 \cdot S - k_4 \cdot X.\end{aligned}\quad (7.13)$$

The signal response curve is calculated and the steady states are determined:

$$\begin{aligned}X_{ss} &= \frac{k_3 \cdot S}{k_4} \\ R_{ss} &= \frac{k_1 \cdot S}{k_2 \cdot X_{ss}} = \frac{k_1 \cdot k_4}{k_2 \cdot k_3} = \text{const.} \neq f(S).\end{aligned}\quad (7.14)$$

The steady state of R is independent of the stimulation of the system. If the system is perturbed, it returns to its initial state after some time (Figure 7.5). This is called adaptive behavior.

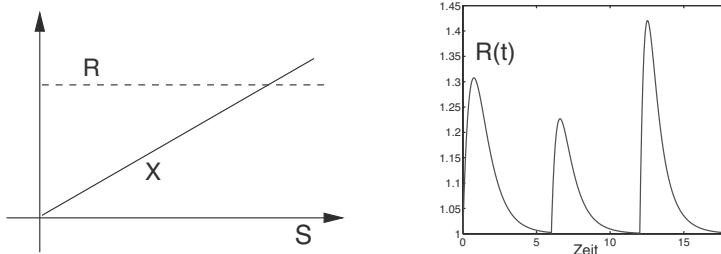


FIGURE 7.5: Adaptive behavior: quantity X depends of S while R is changed independently. A dynamic response of the system occurs if S is changed.

Feedback system

Now we consider a more complex situation as shown in Figure 7.6 that is an extension of the simple activation model. If the synthesis of E^P enhances the production of R which in turn stimulates the synthesis of E^P , then a positive feedback occurs. The output is R . The equations are:

$$\begin{aligned}\dot{R} &= k_0 E^P + k_1 S - k_2 R \\ \dot{E}^P &= \frac{k_3 R E}{K_1 + E} - \frac{k_4 E^P}{K_2 + E^P} = \frac{k_3 R (E_0 - E^P)}{K_1 + (E_0 - E^P)} - \frac{k_4 E^P}{K_2 + E^P}.\end{aligned}\quad (7.15)$$

Figure 7.7 shows the graphical solution of the steady states for different values of S . The Goldbeter-Koshland function was used to solve the equation. The

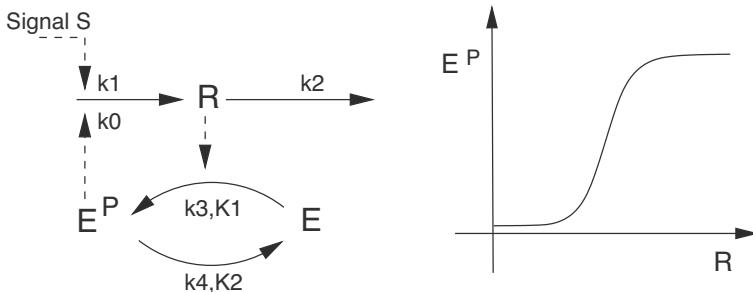


FIGURE 7.6: Left: Feedback mechanism. Right: Response curve when R is the input for the phosphorylation of E (adopted from Figure 7.3).

number of solutions for the whole system depends on the initial signal. We see that from the equations that S affects the synthesis additively. The solid curves move up and three, two or only one intersection point can be found.

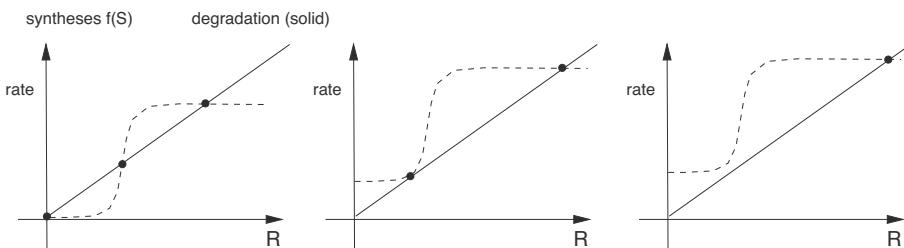


FIGURE 7.7: Steady state values in dependence on S . The curve moves up with the higher values of S .

In addition, multi-stationarity can occur. It can be shown that for three stationary points, the second one is unstable, in the case of two stationary points coinciding in the point of bifurcation. If the signal is increased above the bifurcation point and then lowered again to a value below the bifurcation, the signal output remains at a high value. This means that a switch off is not possible. The system irrevocably remains in a new stationary state. Such a simulation is shown in the right plot of Figure 7.8. Such processes can be found for example in cell division or apoptosis, the programmed cell death.

Influence of an effector

In this example, the first mechanism analyzed (Figure 7.1) is now extended to study the influence of an effector (here an inhibitor) [2]. The inhibitor can build a complex with signal S which does not support the transformation of R to R^P (Figure 7.9). The signal response curve can be determined with

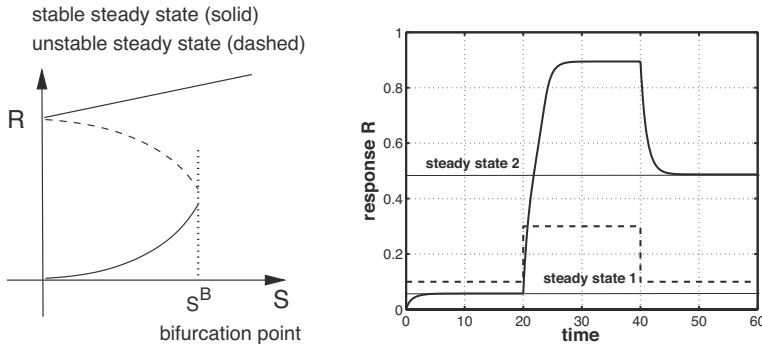


FIGURE 7.8: Left: All steady states as a function of S . For values larger than S^B , only one steady state exists. Right: Simulation of the feedback. First, an initial condition is chosen such that the lower stable branch is reached (steady state 1). Then, the incoming signal (scaled) is increased before returning to its original value. The system remains in steady state 2.

the assumption that the reversible reaction is in equilibrium. The following equations are obtained:

$$\begin{aligned}\dot{R}^P &= k_1 \cdot S \cdot (R_0 - R^P) - k_2 \cdot R^P \\ S_0 &= S + ES \\ E_0 &= E + ES.\end{aligned}\tag{7.16}$$

As in the previous example, one gets:

$$R^P = \frac{SR_0}{S + K_{12}}.\tag{7.17}$$

However, in this case, S depends on total inhibitor concentration E_0 which might be a non linear function. For stationary conditions, it can be shown that the conservation equations are as follows:

$$\begin{aligned}S_0 &= S + \frac{ES}{K_E} \\ E_0 &= E + \frac{ES}{K_E},\end{aligned}\tag{7.18}$$

where K_E is the binding constant of the reaction. It is assumed that S_0 and E_0 have a similar order of magnitude. S and E can be determined with the

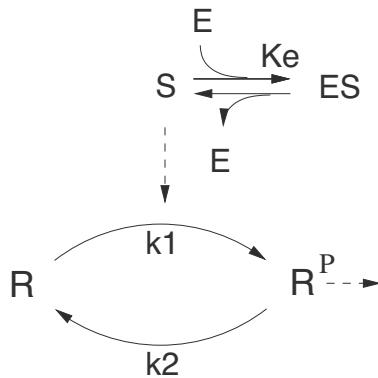


FIGURE 7.9: The effector E binds to the signal and deactivates the signaling pathway.

above two equations:

$$\begin{aligned} E &= \frac{E_0 K_E}{K_E + S} \\ \rightarrow S &= S_0 - \frac{S}{K_E} \frac{E_0 K_E}{K_E + S} = S_0 - \frac{E_0 S}{K_E + S}. \end{aligned}$$

The last equation leads to a quadratic equation for S .

$$S^2 + S(K_E + E_0 - S_0) - S_0 K_E = 0. \quad (7.19)$$

Solving this quadratic equation can be simplified for $E_0 \approx 0$ and small K_E :

$$S \approx S_0. \quad (7.20)$$

This is a bisection line. The Michaelis-Menten relation results after inserting S into Equation (7.17). For larger E_0 and smaller K_E one can estimate:

$$S \approx S_0 - E_0. \quad (7.21)$$

Thereby, one obtains a graph which is > 0 but has very small values if $S_0 < E_0$. For large values of S_0 , the equation becomes linear. Overall, the result is a function that can be transformed into a Hill equation. The signal response curve is shown in Figure 7.10. The Hill coefficient for different values of the effector is also shown.

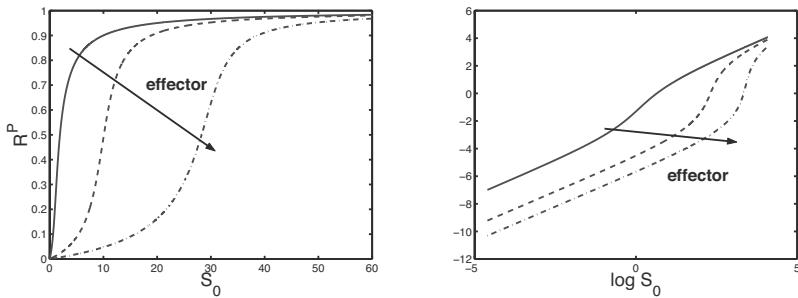


FIGURE 7.10: Influence of the effector on the result. With higher concentrations of the effector, the graph shifts from hyperbolic to sigmoid behavior. This can be illustrated with the Hill coefficient.

7.2 Oscillating Systems

Oscillations can be observed in many biological systems. They arise if feedback loops or activators (or both) are part of the network. Figure 7.11 shows a network which is an extension of the one in Figure 7.6. A component X is added to the system. R stimulates the synthesis of X while X stimulates the decomposition of R . Certain values of the signal S lead to oscillations of the system.

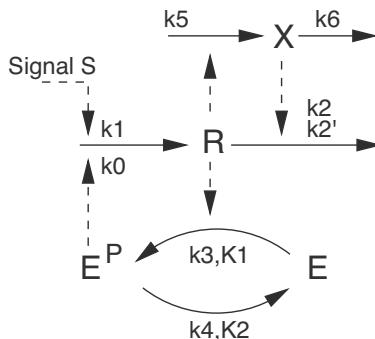


FIGURE 7.11: Signal transduction system which oscillates for particular values of S .

For the two components R and X one gets:

$$\begin{aligned}\dot{R} &= k_0 E^P + k_1 S - k_2 R - k'_2 R X \\ \dot{X} &= k_5 R - k_6 X.\end{aligned}\tag{7.22}$$

For E^P , it is assumed that the reaction is very fast and therefore stationary. The equation for E^P can be written like Equation (7.4). Simulations show that a limit cycle exists (Figure 7.12 on the left). On the right, the steady state is shown against the signal. If oscillations occur, maximum and minimum values are marked with “o.” If no oscillations occur, the steady states are marked with a continuous line.

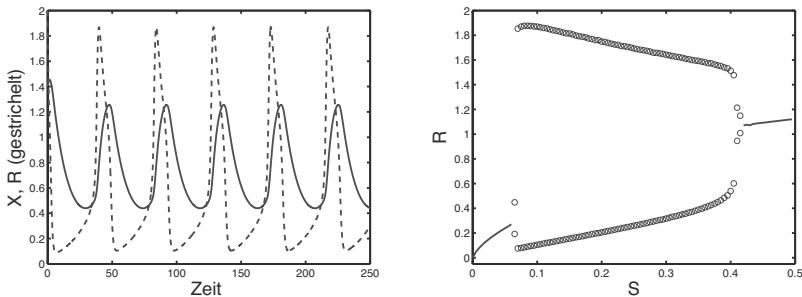


FIGURE 7.12: Left: Time response of R and X for a special value of S . Right: Signal response diagram. Plot of the stationary system for R in dependence of S . The maximum and minimum values of the oscillations are shown as symbols.

Now, the system is being examined more closely. The Goldbeter-Koshland function $G(R)$ can again be used for the calculation of E^P . Both nullclines of the system are denoted by $X(R)$:

$$\text{from Equation for } R: \quad X = \frac{k_0 E^P(R) + k_1 S - k_2 R}{k'_2} \quad (7.23)$$

$$\text{from Equation for } X: \quad X = \frac{k_5 R}{k_6}. \quad (7.24)$$

The first equation includes the sigmoid characteristic $E^P(R)$ and the signal S as an additional parameter; the second equation is linear.

Figure 7.13 shows the nullclines for various values of S . The linear nullcline intersects in a single point with each nullcline, which is the only fixed point. The vertical nullcline (the first equation) is characterized by two local extrema. One extrema lies above the straight line with the second one below in case of oscillations. For larger (or lower) values of S , both extrema are on the same side — above the straight line in case of low values of S . Figure 7.13 shows the behavior if parameter k_0 and k_2 are modified. As can be seen, there is only a small range where oscillations occur.

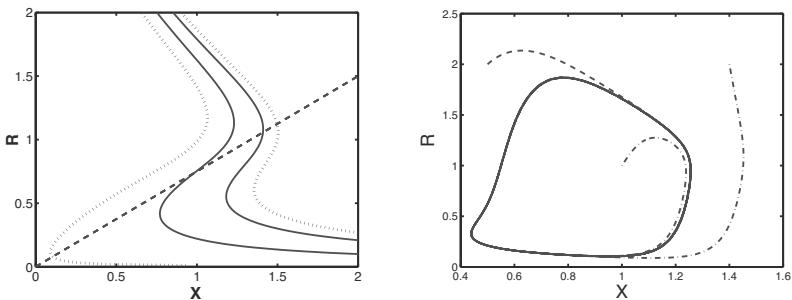


FIGURE 7.13: Left: Nullclines for different values of S . The values of S for which the system oscillates are solid. Right: Limit cycle; plotted for different initial values.

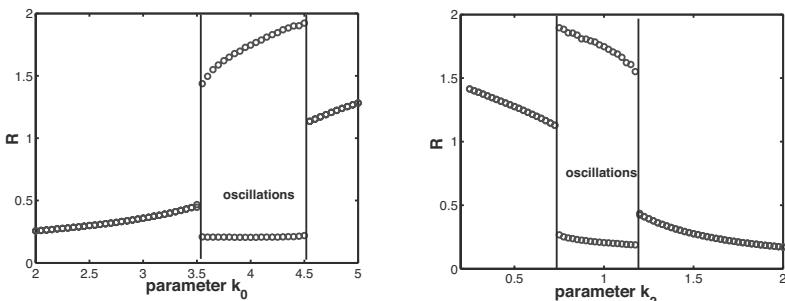


FIGURE 7.14: Left: Parameter study for parameter k_0 . Right: Parameter study for parameter k_2 . Oscillations occur only in a small range of the parameters. Shown are the maximal and the minimal values of R during the time course.

7.3 Genetically Regulated Networks

7.3.1 Inputs/outputs

In Chapter 6, equations for a detailed mathematical description of gene expression were shown whereby no regulation was considered in the sense of a feedback. This will be done now in a similar analysis. First, an inducer interacting with a regulator will be examined regarding inputs and outputs, as well as dynamics [3, 4]. Figure 7.15 shows different variations of connections (inductor i , regulator R , regulated gene g , genetic product G).

The first step considers only the logic of interaction of the two inputs, inducer i and regulator R . We ask about the output — gene is expressed or

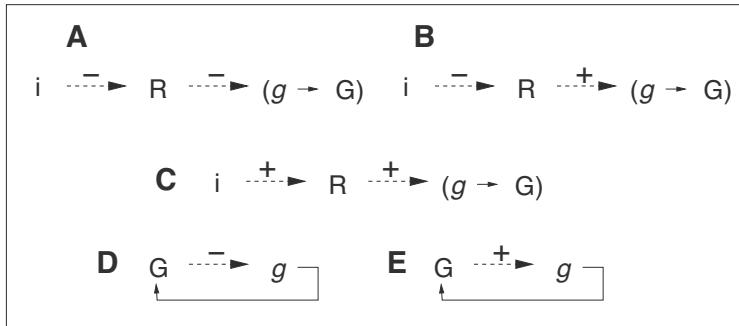


FIGURE 7.15: Model of the connections in gene regulation. **A** The inducer activates the repressor. **B** The inducer deactivates the activator. **C** The inducer activates the activator. **D** Negative self-control. **E** Positive self-control.

not — if one or both inputs are present or not. Table 7.1 shows the logical values for all cases from **A** to **C**. Here we see that only case **C** is a classical boolean function, an AND gate; only when both inputs are present is the output switched on. Case **A** is the negative of case **B** and can be written as: i OR NOT R . It is the classical scheme also valid for lactose induction. If allactose is available, repressor $LacI$ is deactivated and synthesis of the genes from the lac operon can start.

TABLE 7.1: Logic of the connections in plots **A** to **C**. The first two columns represent inputs inducer i and regulator R ; the other columns indicate if product is formed (+) or not (-).

i	R	case A	case B	case C
-	-	+	-	-
+	-	+	+	-
-	+	-	-	-
+	+	+	-	+

7.3.2 Negative and positive feedback reactions

The dynamic behavior of negative and positive feedback reaction will be analyzed in more detail. A basic model for both cases can be written as:

$$\dot{G} = \bar{f} = r_{syn} - r_d = f(G) - k_d G \quad (7.25)$$

where the function $f(G)$ models the feedback reaction:

$$\text{No regulation } f(G) = k_{syn} \quad (7.26)$$

$$\text{negative feedback reaction } f(G) = \frac{k_{syn1}}{1 + (G/K_A)^n} \quad (7.27)$$

$$\text{positive feedback reaction } f(G) = k_{syn2} \frac{(G/K_A)^n}{1 + (G/K_A)^n}. \quad (7.28)$$

Here, the Hill kinetics approach is used for describing the interaction with the DNA. A hyperbolic behavior can be observed if no regulation occurs. The parameters are chosen such that all steady states have the same value. Thus, strong promoters ($k_{syn,i} > k_{syn}$) for the positive and negative feedback are necessary. The negative feedback reaction shows a faster response than the other two systems, as its inhibition is slowed with increasing values of G . To reach the steady state, the system must be expressed quickly in the beginning. The time constant can be calculated with the following steps (see Appendix for analysis of non linear systems): Linearize the system, then compute the eigenvalues from the derivative of the system equation. If f is monotonically decreasing, its derivative with respect to G is negative and the eigenvalue can be calculated by the sum of two parts (df/dG and k_d) and therefore is absolutely larger than k_d . Larger eigenvalues correspond to faster time constants. The parameter k_d equals the eigenvalue of the unregulated system.

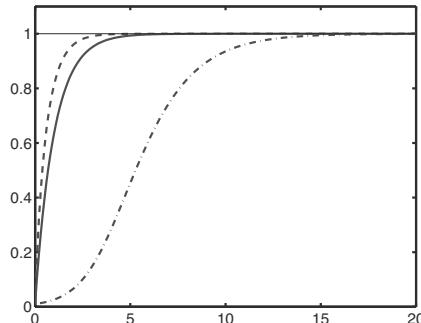


FIGURE 7.16: Negative autoregulation. Time response for the above systems: Continuous line: no regulation. Dashed line: negative feedback reaction. Semicolon: positive feedback reaction (parameters $K_A = 1$, $n = 1$, $k_{syn} = 1$, $k_{syn1} = 2$, $k_{syn2} = 2$).

If a positive feedback is considered it takes longer for systems to respond. Since only a few molecules of G are available at the very beginning, the system

needs to accelerate itself for those that need more time. Positively regulated systems show another interesting characteristic.

EXAMPLE 7.1 *Lactose uptake.*

Here again the lactose uptake system is considered in more detail (a first analysis was performed while comparing with the stochastic approach). For the example and to keep things simple, $n = 2$, $K_A = 1$ and $k_{ab} = 1$ are chosen.

$$\dot{LacY} = \bar{f} = k_{syn2} \frac{LacY^2}{1 + LacY^2} - LacY, \quad (7.29)$$

where k_{syn2} is the parameter that might have different values. The steady states arise from:

$$k_{syn2} \frac{LacY^2}{1 + LacY^2} = LacY, \quad (7.30)$$

where $LacY_1 = 0$ is an obvious stationary solution. The other two stationary solutions can be computed with

$$k_{syn2} LacY = 1 + LacY^2 \quad (7.31)$$

$$\rightarrow LacY_{2/3} = \frac{k_{syn2}}{2} \pm \frac{1}{2} \sqrt{k_{syn2}^2 - 4}. \quad (7.32)$$

One can see that $LacY_{2/3}$ only exists for $k_{syn2} > 2$. In the case of $k_{syn2} = 2$, the discriminant becomes zero and the two steady states coincide. One gets $LacY_{2/3} = 1$. For higher values of k_{syn2} , $LacY_{2/3}$ becomes > 1 , while for smaller k_{syn2} , $LacY_{2/3}$ is < 1 . To determine the stability of the steady states, a linearization is needed (see Appendix). Deriving $\bar{f} = f(LacY) - LacY$ leads to:

$$\frac{d\bar{f}}{dLacY} = \frac{2k_{syn2} LacY}{(1 + LacY^2)^2} - 1. \quad (7.33)$$

For the steady state $LacY_1 = 0$, one gets:

$$\left. \frac{d\bar{f}}{dLacY} \right|_{LacY=0} = -1. \quad (7.34)$$

Therefore, $LacY_1$ is stable. The equation is rewritten using Equation (7.31) to determine the other stationary points. It can be simplified as:

$$\left. \frac{d\bar{f}}{dLacY} \right|_{LacY_{2/3}} = \frac{1 - LacY^2}{1 + LacY^2}. \quad (7.35)$$

It can be seen that a stable steady state results for values $LacY > 1$ and an

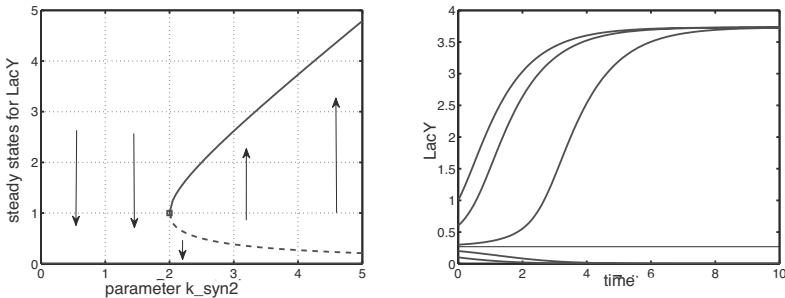


FIGURE 7.17: Left: Steady states in dependence of k_{syn2} . A supercritical bifurcation occurs for $k_{syn2} = 2$. Right: Simulation for $k_{syn2} = 4$. Depending on the initial value, one of the two solution branches is reached.

instable one for $LacY < 1$. The results are summed up in Figure 7.17. The trend of the steady states in dependence of k_{syn2} and the applicable directions of the velocity field are shown. To synthesize a protein efficiently the maximal rate of synthesis must not be too small, otherwise the system is always turned off.

A simulation for $k_{syn2} = 4$ is shown in the right part of the figure. It can be determined that for initial values of $LacY(0) > 0.27$, steady state 2 is reached while for values $LacY(0) < 0.27$, steady state $LacY = 0$ is reached. Now, one needs to analyze the bifurcation point at $k_{syn2} = 2$. The value is inserted into the first derivative: $\left. \frac{d\bar{f}}{dLacY} \right|_{LacY=1} = 0$. As a result no conclusion can be made so far since the value is neither positive nor negative. Thus, one derives a second time and gets:

$$\left. \frac{d^2 \bar{f}}{dLacY^2} \right|_{LacY=1} = \left. \frac{2 k_{syn2} - 6 k_{syn2} LacY^2}{(1 + LacY^2)^3} \right|_{LacY=1} = -\frac{1}{2}. \quad (7.36)$$

One gets for the linearized system:

$$\Delta \dot{LacY} = a_1 \Delta LacY + a_2 (\Delta LacY)^2, \quad (7.37)$$

where $a_1 = 0$ and $a_2 < 0$. The system is unstable since the quadratic term always produces positive results. If the system is close to the steady state and it is perturbed in a positive direction, it will return to the steady state. In case of negative perturbation, the distance to the steady state will increase. Positive autoregulation can be found in many other examples where an inducer is responsible for turning on or off metabolic pathways.

7.3.3 Toggle switch

Until now only systems with an input that switch it on or off were considered. For biotechnological problems, it is not only interesting to switch on

metabolic pathways but to switch another off at the same time. Here toggle switches come into play as it is illustrated in Figure 7.18 showing the connection of two regulatory proteins [5]. Here, the regulators inhibit each other mutually. With additional signals S_1 and S_2 one can realize a switching in both directions.

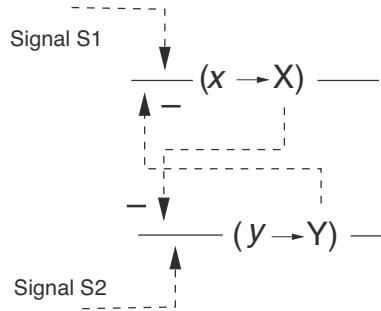


FIGURE 7.18: Toggle switch. Both regulators inhibit the synthesis of each other. A signal S_i leads to the fact that one direction is preferred (inducer, activator).

For the inhibition kinetics, a simple variation of the Michaelis-Menten kinetics is used. The following equations result after scaling the decay rates of both proteins:

$$\begin{aligned}\dot{X} &= \frac{S_1}{1 + Y} - X \\ \dot{Y} &= \frac{S_2}{1 + X} - Y.\end{aligned}\quad (7.38)$$

Before starting the calculation, the equations are simplified. Since the system is symmetric, the case $S_1 = S$ and $S_2 = 1$ is considered. For calculating the steady states, the equations are set equal to 0. One gets the following quadratic equation:

$$Y^2 + SY - 1 = 0. \quad (7.39)$$

If a signal input S was only possible for Y instead of X , this would lead to:

$$Y^2 + (2 - S)Y - S = 0. \quad (7.40)$$

The solution of the Equation (7.39) is:

$$Y^R = -\frac{S}{2} + \frac{1}{2}\sqrt{S^2 + 4}. \quad (7.41)$$

In this case, the steady states can well be determined geometrically. Both

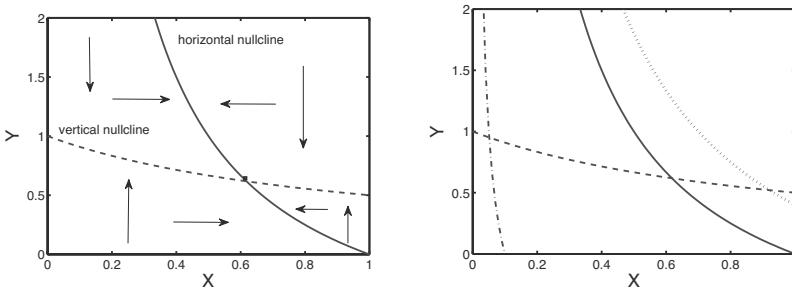


FIGURE 7.19: Left: Phase plane with nullclines; simulation for $S = 1$. Arrows give the direction of the movements of the state variable. Right: For values $S = 0.1, 1, 1.4$ the horizontal nullclines are shown. The steady state goes to the right side. The vertical nullcline doesn't depend on S .

nullclines (geometric locations for $\dot{X} = 0$ and $\dot{Y} = 0$) can be mapped to each other for $S = 1$ by inversion. Figure 7.19 shows the phase plane with both nullclines.

The intersection of the nullcline is the only steady state of the system. A stability analysis based on the nullclines can also be done. The nullclines are given as $Y = f(X)$ and are:

$$\text{Horizontal null cline: } Y = \frac{S - X}{X}; \text{ vertical null cline: } Y = \frac{1}{1 + X} \quad (7.42)$$

The whole phase plan can be subdivided into four parts and for every part the sign of the velocity for X and Y can be determined:

$$\begin{aligned} \text{Increasing X values: } X &< \frac{S}{1 + Y}, & \text{increasing Y values: } Y &< \frac{1}{1 + X} \\ \text{Decreasing X values: } X &> \frac{S}{1 + Y}, & \text{decreasing Y values: } Y &> \frac{1}{1 + X}. \end{aligned}$$

The conditions are shown by arrows in Figure 7.19. The system is stable since the actual velocity is a superposition of both directions. In all four parts, the vector points to the steady state. Only the X value is strongly influenced by a change of S , as can be seen. If a complete switch will be reached, the second quantity Y must also be changed. This is shown in Figure 7.20. Thereby, the signal was alternately switched on and off by S_1 and S_2 . The trend in the phase plane is shown on the right side.

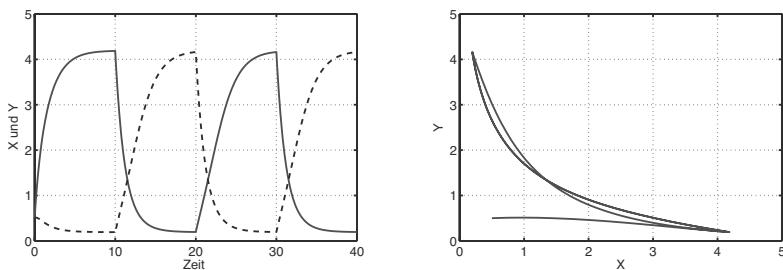


FIGURE 7.20: On the left: Time response of toggling between X and Y ($S_1 = 5$ or 1 and $S_2 = 1$ or 5). On the right: Trend in the phase plane. The starting value is $X = 0.5$, $Y = 0.5$.

7.4 Spatial Gradients by Signal Transduction

Components for signal transduction are often linked to a specific site inside the cell (membrane, compartments). To react with each other, the free components must be able to diffuse.

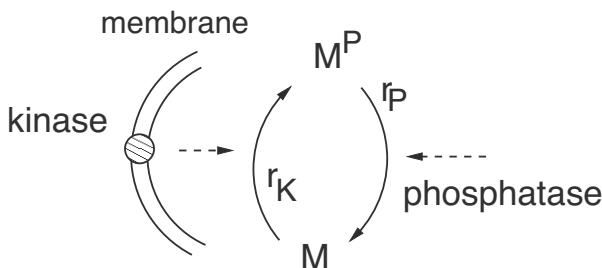


FIGURE 7.21: Schematic representation of the network to illustrate a spatial gradient. The kinase is bound to the membrane while the phosphatase is freely available in the cytosole.

The following example shall serve as an illustration: A kinase and a phosphatase are involved in protein phosphorylation and dephosphorylation. The kinase is considered as a membrane in relationship to where it interacts with certain other proteins. The phosphatase is freely available in the cytosole and leads to a dephosphorylation (see Figure 7.21). Here we follow the presentation of Kholodenko [6]. To deduce equations for the spatial and temporal changes, the cell is here considered tube-shaped. Then, a small volume element dV is considered in which diffusion and reaction take place at the same time.

For the volume element dV from Figure 7.22 the temporal change of the

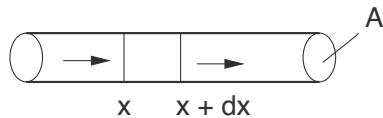


FIGURE 7.22: A rod-shaped bacteria can be abstracted by a tube with a constant cross section A .

substance concentration can be described in the following way:

$$\frac{dc}{dt} dV = j_x - j_{x+dx} - r dV \quad (7.43)$$

whereby diffusion j and a reaction rate r are considered. For the terms of diffusion and the diffusion coefficient D^* we obtain:

$$\begin{aligned} j_x &= -D^* \left. \frac{dc}{dx} \right|_x \\ j_{x+dx} &= -D^* \left. \frac{dc}{dx} \right|_{x+dx} \\ &= -D^* \left. \frac{dc}{dx} \right|_x - D^* \left. \frac{d}{dx} \left(\frac{dc}{dx} \right) \right|_x dx \end{aligned} \quad (7.44)$$

whereby a Taylor series approximation is used for the second term. By inserting, one gets:

$$\frac{dc}{dt} dV = D^* \frac{d^2c}{dx^2} dx - r dV \quad \rightarrow \quad \frac{dc}{dt} = \underbrace{\frac{D^*}{A}}_D \frac{d^2c}{dx^2} - r. \quad (7.45)$$

The diffusion model for the phosphorylated component with the rate of dephosphorylation r_P results then in the following equation:

$$\frac{dM^P}{dt} = D \frac{d^2M^P}{dx^2} - r_P \quad (7.46)$$

with side conditions:

$$-D \left. \frac{dM^P}{dx} \right|_{x=0} = r_K, \quad \left. \frac{dM^P}{dx} \right|_{x=l} = 0,$$

and for the nonphosphorylated component:

$$\frac{dM}{dt} = D \frac{d^2M}{dx^2} + r_P \quad (7.47)$$

with the side conditions:

$$D \left. \frac{dM}{dx} \right|_{x=0} = r_K, \quad \left. \frac{dM}{dx} \right|_{x=l} = 0$$

whereby l represents the distance to the middle of the cell (approx. $0.5 \mu\text{m}$ in *E. coli*). On the left boundary, phosphorylation happens with the rate r_K . The minus in the equation for M^P results from the direction of the gradient. For the right boundary (middle of the cell) a symmetric cell is assumed and the gradient disappears. Next we show how a system given as a partial differential equation can be simulated. Here, we discretize the spatial coordinate z . One gets a large system of differential equations which must be solved simultaneously:

$$\begin{aligned}\frac{dM_i^P}{dt} &= D \frac{M_{i-1}^P - 2M_i^P + M_{i+1}^P}{\Delta z^2} - k_2 M_i^P \\ \frac{dM_i}{dt} &= D \frac{M_{i-1} - 2M_i + M_{i+1}}{\Delta z^2} + k_2 M_i^P\end{aligned}\quad (7.48)$$

with two side conditions:

(i)

$$\begin{aligned}\left. \frac{dM^P}{dx} \right|_{x=0} &\approx \frac{M_1^P - M_0^P}{\Delta z} \\ \left. \frac{dM}{dx} \right|_{x=0} &\approx \frac{M_1 - M_0}{\Delta z}.\end{aligned}$$

(ii)

$$\begin{aligned}\left. \frac{dM^P}{dx} \right|_{x=l} &\approx \frac{M_l^P - M_{l-1}^P}{\Delta z} \\ \left. \frac{dM}{dx} \right|_{x=l} &\approx \frac{M_l - M_{l-1}}{\Delta z}.\end{aligned}$$

These conditions must be evaluated individually. One gets for M_0 :

$$D \frac{M_1 - M_0}{\Delta z} = r_K = k_1 K M_0 \quad (7.49)$$

whereby for the rate r_K a simple approach is done (K is a kinase activity, which corresponds to the stimulus). By transforming the equation, one gets:

$$M_0 = \frac{D M_1}{D + k_1 K \Delta z}. \quad (7.50)$$

Using an analogous approach as above, one gets for M_0^P :

$$M_0^P = \frac{D M_1^P + k_1 K \Delta z M_0}{D}. \quad (7.51)$$

The analysis of the side conditions for $x = l$ leads to:

$$M_l = M_{l-1} \quad \text{und} \quad M_l^P = M_{l-1}^P. \quad (7.52)$$

The following equation applies for the total amount of the component M_{tot} :

$$M_{tot} = M + M^P = M_0 + M_0^P + \sum M_i^P + \sum M_i. \quad (7.53)$$

Figure 7.23 shows the local profile of the phosphorylated and the nonphosphorylated component.

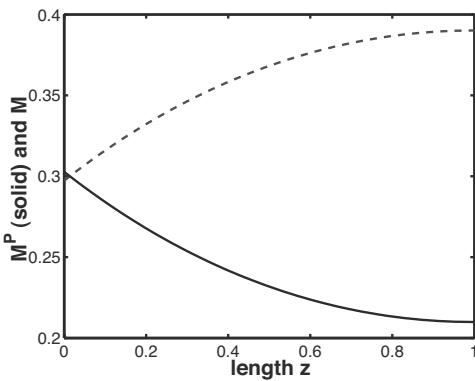


FIGURE 7.23: Stationary local profile with the length $l = 1$ of the phosphorylated and nonphosphorylated components.

7.5 Analysis of Signaling Pathways by Heinrich

Even though analyzing methods are the subject of a later chapter, this specific method for signaling pathways is now introduced. When analyzing signaling systems, there are a couple of interesting questions for characterizing the dynamics of the signal response. For example, the following questions are especially important [7]:

- (i) How long does the signal need to reach to the target point?
- (ii) How long does the signal have an effect?
- (iii) How strong is the signal?

The last question focuses on signal amplification characteristics of the signaling pathway while the first two questions look for the dynamics and

the temporal behavior of systems. The following restrictions are made for the analysis: The initial signal is abruptly changed from 0 to R_0 and decays with parameter λ . The corresponding equation for the time response is $R(t) = R_0 \cdot e^{-\lambda t}$. The method demands that the response of the system X is zero at time zero and approaches zero for extended intervals of time. Examples for possible input and output behavior are shown in Figure 7.24.

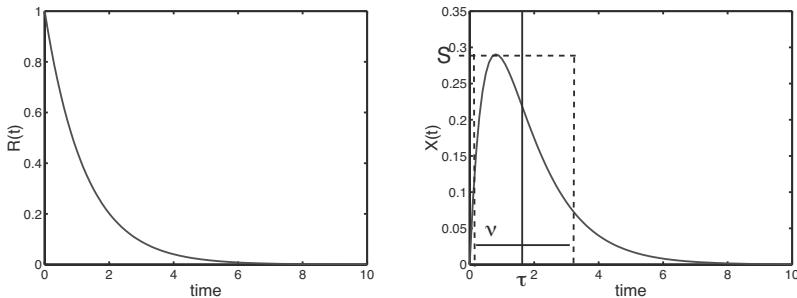


FIGURE 7.24: Time response of the input R and the output X .

To answer the questions asked above, first some characteristics must be determined. The definitions of the characteristics are formally borrowed from statistics (the course of x is considered as a distribution). One gets the following quantities:

$$\begin{aligned} I &= \int_0^\infty X(t) dt \equiv \text{area under the graph} \\ T &= \int_0^\infty X(t) \cdot t dt \equiv \text{with time weighted area.} \end{aligned} \quad (7.54)$$

From the two quantities one can calculate

$$\tau = \frac{T}{I} \quad (7.55)$$

as the average time (arithmetic mean) until the activation of the initial signal, which represents the activation time. The mean signal duration time (duration of activation) corresponds to the standard deviation and is calculated by means of

$$Q = \int_0^\infty X(t) \cdot t^2 dt. \quad (7.56)$$

It holds:

$$\nu = \sqrt{\frac{Q}{I} - \tau^2}. \quad (7.57)$$

The signal amplification is defined as the height of a rectangle with the width 2ν and the area $X(t)$ under a curve:

$$S = \frac{I}{2\nu}. \quad (7.58)$$

The introduced quantities are illustrated in Figure 7.24.

EXAMPLE 7.2 Simplified model of a phosphorelay chain.

As an example, a simplified model of a phosphorelay chain is used. Starting point is the activation model from above. The signal, here R , activates the phosphorylation of protein E . In subsequent steps the phosphoryl group is transferred to further protein E_j . For the first step the following equation holds, if we directly use the conservation equation $E_{10} = E_1 + E_1^P$:

$$\frac{dE_1^P}{dt} = k_1 \cdot E_{10} \cdot R \cdot \left(1 - \frac{E_1^P}{K_{10}}\right) - k_{-1} \cdot E_1^P.$$

Using new variables

$$X_i \equiv E_i^P, \text{ and if holds } E_{i0} = E_i + E_i^P \gg E_i^P,$$

which corresponds to a low activation, the system simplifies:

$$\begin{aligned} \frac{dx_1}{dt} &= \alpha_1 R - \beta_1 x_1 \\ \frac{dx_i}{dt} &= \alpha_i x_{i-1} - \beta_i x_i \end{aligned} \quad (7.59)$$

with the according parameters α, β . To calculate the parameters, firstly the correlation between I_i and I_{i-1} is determined. The integration of Equation 7.59 results in:

$$\int \frac{dx_i}{dt} dt = \alpha_i \int x_{i-1} dt - \beta_i \int x_i dt \stackrel{!}{=} 0 \quad (7.60)$$

since the initial and the last value are equal. Therefore all changes should be compensated for when looking at it for a longer time period. It applies:

$$0 = \alpha_i I_{i-1} - \beta_i I_i \rightarrow \frac{I_i}{I_{i-1}} = \frac{\alpha_i}{\beta_i}. \quad (7.61)$$

For the input signal one gets:

$$I_0 = \int_0^\infty R_0 e^{-\lambda t} = -\frac{R_0}{\lambda e^{-\lambda t}} \Big|_0^\infty = \frac{R_0}{\lambda} \quad (7.62)$$

and finally we get after iterative insertion:

$$I_i = \frac{R_0}{\lambda} \prod_j \frac{\alpha_j}{\beta_j}. \quad (7.63)$$

To calculate the response time τ of the last step, it is also approached iteratively. Firstly it is converted to:

$$\frac{dx_i}{dt} = \alpha_i \cdot x_{i-1} - \beta_i \cdot x_i \quad \longrightarrow \quad \int_0^\infty \frac{dx_i}{dt} \cdot t dt = \alpha \cdot T_{i-1} - \beta_i \cdot T_i. \quad (7.64)$$

The left side of the equation above results in:

$$\begin{aligned} \int_0^\infty \frac{dx_i}{dt} \cdot t dt &= \underbrace{x_i \cdot t|_0^\infty}_{=0} - \int_0^\infty x_i dt \\ &= -I_i. \end{aligned} \quad (7.65)$$

Comparing the two results leads to:

$$-I_i = \alpha_i \cdot T_{i-1} - \beta_i \cdot T_i \quad (7.66)$$

$$\begin{aligned} \longrightarrow 1 &= -\alpha_i \cdot \frac{T_{i-1}}{I_i} + \beta_i \cdot \frac{T_i}{I_i} = -\alpha_i \cdot \frac{T_{i-1}}{I_{i-1}} \cdot \frac{I_{i-1}}{I_i} + \beta_i \cdot \frac{T_i}{I_i} \\ &= -\alpha_i \cdot \tau_{i-1} \cdot \frac{\beta_i}{\alpha_i} + \beta_i \cdot \tau_i = -\tau_{i-1} \cdot \beta_i + \beta_i \cdot \tau_i. \end{aligned} \quad (7.67)$$

Since we are interested in τ we get:

$$\tau_i = \tau_{i-1} + \frac{1}{\beta_i}. \quad (7.68)$$

For the whole system we get:

$$\tau = \frac{1}{\lambda} + \sum_j \frac{1}{\beta_j}. \quad (7.69)$$

The other two quantities are as follows (see also Exercises):

$$\nu = \sqrt{\frac{1}{\lambda^2} + \sum_j \frac{1}{\beta_j^2}}, \quad S = \frac{\frac{R_0}{2} \prod_j \frac{\alpha_j}{\beta_j}}{\sqrt{1 + \lambda^2 \sum_j \frac{1}{\beta_j^2}}}. \quad (7.70)$$

The result can be summarized like the following:

- Quantities τ, ν are only depending on the phosphatase activities (β_i).
 - Regulation of the system only aims for signal amplification, not on fast dynamics ($\tau_i > \tau_{i-1}$, $\nu_i > \nu_{i-1}$).
 - A signal amplification or attenuation is possible from step to step; therefore different combinations can lead to the same final signal characteristics.
-

Exercises

EXERCISE 7.1 Simple signaling networks.

a Consider the signaling component X that is synthesized with rate r_1 in dependence on signal S and is degraded with rate r_2 :

$$\dot{X} = r_1 - r_2 = k_1(S)X - \frac{k_2 X}{K + X}. \quad (7.71)$$

Show graphically that there is a maximal one single steady state $X_{ss} > 0$ for X in dependence from S . Determine an analytical solution and give constraints for the case that a steady state exists. Confirm your results from the first part.

b Consider the signaling component X that is synthesized with rate r_1 in dependence on signal S and Y and is degraded with rate r_2 :

$$\dot{X} = r_1 - r_2 = \frac{Y}{K(S) + Y} - \frac{X}{1 + X}. \quad (7.72)$$

Components X and Y are two forms of a signaling complex and the sum $X_0 = X + Y$ is constant. Show graphically that there is one single steady state $X_{ss} > 0$ for X in dependence from S . Determine an analytical solution and confirm your results from the first part.

EXERCISE 7.2 Goldbeter-Koshland function.

Derive Equation (7.12) from Equation (7.11).

EXERCISE 7.3 Genetically regulated network.

An autocatalytic regulated synthesis of a protein G is described using the following model:

$$\dot{G} = \bar{f} = \frac{k G^2}{1 + G^2} - G, \quad k \geq 0. \quad (7.73)$$

The parameter k shall now be dependent on a signal S , i.e., $k(S)$. Next, determine all steady states of the system. Verify the calculation in the chapter:

Linearize and show that using algebraic manipulation that the following equation is the first derivative of the function \bar{f} for the steady states $G \neq 0$:

$$\frac{d\bar{f}}{dG} = \frac{1 - G^2}{1 + G^2}. \quad (7.74)$$

Carry out an analysis of the steady states of the system.

EXERCISE 7.4 *Analysis of signaling systems.*

In the text, the method of Heinrich was applied for a cascade, and time constants τ_i were determined explicitly. Verify also the calculations for the mean signal duration ν and signal amplification S .

Bibliography

- [1] J. J. Tyson et al. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Current Opinion in Cell Biology* 15: 221–231 2003.
- [2] J. E. Ferrell. Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. *TIBS* 21: 460-466, 1996.
- [3] M. Kaern et al. The engineering of gene regulatory networks. *Annu. Rev. Biomed. Eng.* 5: 179-206, 2003.
- [4] U. Alon. Network motifs: theory and experimental approaches. *Nature Rev. Gen.* 8: 450-460, 2007.
- [5] H. Kobayashi et al. Programmable cells: interfacing natural and engineered gene networks. *PNAS* 101: 8414-8419, 2004.
- [6] B. N. Kholodenko. Cell-signaling in time and space. *Molecular Cell Biology* 7: 165-176, 2006.
- [7] R. Heinrich et al. Mathematical models of protein kinase signal transduction. *Molecular Cell* 9: 957-970, 2002.

Part III

Analysis of Modules and Motifs

This page intentionally left blank

Chapter 8

General Methods of Model Analysis

The previous part of the book was dedicated to determining kinetic rates r_j that appear in differential equations for intracellular metabolites:

$$\dot{\underline{c}} = N \underline{r} - \mu \underline{c} = \underline{f}(\underline{c}, \underline{p}).$$

As we have seen, many possible combinations of simple first order reaction kinetics, Michaelis-Menten type kinetics or Hill equations with parameters p_j are applied that make the system quite complicated for an analysis. We also have already seen that the steady state of such a system can be calculated and we are interested in some properties of this steady state, especially if the system is perturbed to reach a different steady state. Here, we are interested in the time constants of the system, the influence of parameters p_j on the dynamics, structural properties that depend only weakly on the parameters and on stability properties of the steady states. These topics are covered in this chapter. Here, also two classical frameworks on modeling and model analysis, namely Metabolic Control Analysis and Biochemical Systems Theory, are covered.

8.1 Analysis of Time Hierarchies

Processes in cells proceed on different time scales. The different ranges of the time constants are denoted as hierarchies. The following time constants for cellular processes are known:

- Metabolic pathway with 8 reactions: $\sim 2\text{--}3\text{ s}$
- Phosphate transfer in a signal transduction pathway: $\sim \text{ms}$
- Synthesis of a protein molecule: $\sim 60\text{ s}$
- Cell division in bacteria: $\sim 20\text{--}100\text{ min}$
- Evolution (random mutation): $\sim \text{days/years}$

The notion of time constants shall be explained in a simple example. A component A is considered, which is formed through a constant reaction rate u and is consumed by a 1st order reaction:



The following equation is valid for the description of the dynamics of A :

$$\dot{A} = u - k_1 \cdot A. \quad (8.1)$$

If the system was initially at steady state A^* (for a specific $u > 0$) and then displaced with $u = 0$, the progression of A is then:

$$A(t) = A^* \cdot e^{-k_1 t}. \quad (8.2)$$

The time constant can be determined via the calculation of the tangents at A^* and the calculation of the cutting point of this tangent with the t axis. Formally:

$$\text{Slope: } \dot{A}(0) = -k_1 \cdot A(0) = -k_1 \cdot A^*$$

$$\text{Axis intersection: } y = -k_1 \cdot A^* \cdot \tau + A^* \stackrel{!}{=} 0$$

$$\rightarrow \tau = \frac{1}{k_1}. \quad (8.3)$$

Figure 8.1 illustrates the procedure. On the right the partitioning of the ranges of the time constants is shown. In general, one is only interested in a specific time range.

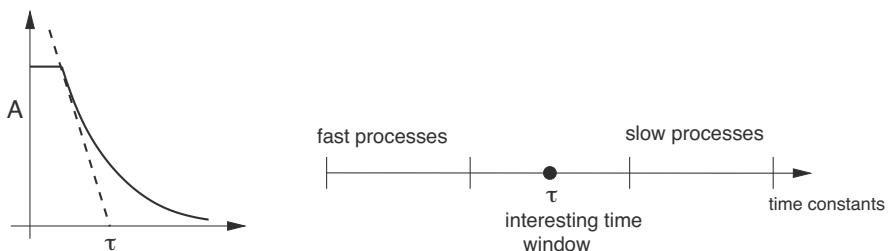


FIGURE 8.1: Illustration of time constants.

8.1.1 Linear systems

In the classical approach, one considers the processes near an equilibrium or a steady state. One can therefore start with the behavior in the proximity of an operating point which can be described by a linearized model. To derive

the linearized model, the steady states are calculated by setting the time derivatives of the non linear model to zero. This leads to a system of the following algebraic form:

$$\dot{\underline{x}} = \underline{0} = f(\underline{x}), \quad (8.4)$$

where the number of steady states depends on the structure of the equations and can be between 0 and ∞ . Linearization then proceeds by calculation of the Jacobian matrix J of the system (see Appendix for details); it is a matrix containing the partial derivatives of the individual functions f_i with respect to the state variables x_j . The Jacobian matrix is quadratic in its nature and has n^2 entries, where n is the system's order. A linearization at steady states \underline{x} , if a small deviation \underline{x}' is considered, results in system Equation (8.5) of the following form:

$$\boxed{\dot{\underline{x}'} = J \underline{x}'} \quad (8.5)$$

with the entries j_{ij} of the Jacobian matrix

$$j_{ij} = \left. \frac{\partial f_i}{\partial x_j} \right|_{\underline{x}=\underline{x}}, \quad (8.6)$$

where each steady state is considered separately. A system of linear differential equations can be solved using the approximation

$$\underline{x}' = \underline{v} e^{\lambda t} \quad (8.7)$$

where the vector \underline{v} and the λ are to be determined. Note that the λ are reciprocal to time and therefore represent time constants:

$$\boxed{\tau = \frac{1}{|\text{Re}(\lambda)|}}; \quad (8.8)$$

the larger the value, the faster the system is. On the other hand, small values for λ indicate a slow system with a large time constant. The calculation leads to an eigenvalue problem of the form:

$$|\lambda I - J| \stackrel{!}{=} 0. \quad (8.9)$$

The associated eigenvectors are the vectors \underline{v} . As the calculation yields n eigenvalues (not necessarily n different), this allows writing the progression of state variables as a superposition of various exponential functions. Alternatively, solving a linear system can proceed in analogy to solving a one dimensional system. One uses here the fundamental matrix:

$$\underline{x} = \underline{x}_{(0)} \cdot e^{Jt}, \quad e^{Jt} \equiv \text{matrix function, fundamental matrix} \quad (8.10)$$

where for the fundamental matrix the following relation is valid:

$$e^{Jt} = \sum_{m=0}^{\infty} \frac{J^m \cdot t^m}{m!}. \quad (8.11)$$

For a second order system with Jacobian matrix

$$J = \begin{pmatrix} j_{11} & j_{12} \\ j_{21} & j_{22} \end{pmatrix} \quad (8.12)$$

one obtains an equation for the eigenvalues:

$$\begin{aligned} \lambda^2 - \text{trace}(J) \lambda + \det(J) &= \\ \lambda^2 - (j_{11} + j_{22}) \lambda + (j_{11} j_{22} - j_{12} j_{21}) &= 0. \end{aligned} \quad (8.13)$$

The position of the eigenvalue in the complex plane determines the stability behavior. The different cases for a 2nd order system are summarized in the Appendix. An interesting analysis option is to transform the system, so that for every new state variables, an individual time constant can be assigned. With this, the new state variables are composed of a superposition of the old state variables. For the transformation in the new \underline{z} coordinates with matrix T :

$$\begin{aligned} \underline{x}' &= T \cdot \underline{z} \\ \rightarrow \quad T \cdot \dot{\underline{z}} &= J \cdot T \cdot \underline{z} \\ \rightarrow \quad \dot{\underline{z}} &= T^{-1} \cdot J \cdot T \cdot \underline{z} = \Lambda \underline{z}. \end{aligned} \quad (8.14)$$

T must be chosen such that the eigenvalues are only in the diagonals of $\Lambda = T^{-1} J T$. Having the system transformed in such a way the solution for the individual z_i can be determined easily:

$$z_i(t) = z_0 e^{\lambda_i t}; \quad (8.15)$$

in this way the new variables are characterized by their individual eigenvalues λ_i . From the transformation matrix conclusions can be made, for example, in which way the components have to be combined to run on different time scales. Although the method is straightforward from a theoretical point, it is not clear if the new variables z_i have a clear meaning or biological interpretation with respect to the old variables.

EXAMPLE 8.1 Simple reaction system.

Consider the following reaction network:



The equations here read (assume the volume is constant and no growth occurs):

$$\begin{aligned} \dot{A} &= -r_1 = -k_1 \cdot A + k_{-1} \cdot B \\ \dot{B} &= r_1 - r_2 = k_1 \cdot A - k_{-1} \cdot B - k_2 \cdot B \\ \dot{C} &= r_2 = k_2 \cdot B. \end{aligned} \quad (8.17)$$

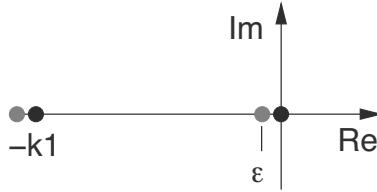


FIGURE 8.2: Positions of the eigenvalues for both cases $k \approx 0$ (gray) and $k = 0$ (black).

For the variables of interest to us, A and B , the system (already linear) can be written in matrix form:

$$\begin{pmatrix} \dot{A} \\ B \end{pmatrix} = \begin{pmatrix} -k_1 & k_{-1} \\ k_1 & -(k_{-1} + k_2) \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix}. \quad (8.18)$$

The characteristic equation (for determining the eigenvalues) reads:

$$|\lambda I - J| = \det(\lambda I - J) \stackrel{!}{=} 0. \quad (8.19)$$

One calculates with the equations from above:

$$\begin{aligned} \lambda^2 + (k_1 + k_{-1} + k_2) \cdot \lambda + k_1 \cdot (k_{-1} + k_2) - k_1 \cdot k_{-1} &= 0 \\ \rightarrow \lambda_{1,2} &= -\frac{k_1 + k_{-1} + k_2}{2} \pm \sqrt{\frac{(k_1 + k_{-1} + k_2)^2}{4} - k_1 \cdot k_2}. \end{aligned} \quad (8.20)$$

In what follows, the special case $k = k_{-1} = k_2 \ll k_1$ will be considered. One then obtains for the eigenvalues:

$$\lambda_{1,2} = -\frac{k_1 + 2k}{2} \pm \sqrt{\frac{(k_1 + 2k)^2}{4} - k_1 k} \quad (8.21)$$

$$= -\frac{k_1 + 2k}{2} \pm \frac{1}{2}\sqrt{k_1^2 + 4k^2}. \quad (8.22)$$

The equation is still complex. Therefore, in the first step, a further simplification is done: $k \approx 0$. This leads to $\lambda_1 = 0$ and $\lambda_2 = -k_1$. Consider now a change for k in a small positive value. Then we get: $\lambda_1 = -\epsilon$ and $\lambda_2 = -k_1 - \epsilon$ with a undefined but small value for ϵ . This is illustrated in Figure 8.2. If the real parts $|Re(\lambda)|$ are very small, this leads to large time constants; for very large $|Re(\lambda)|$, the system is very fast and has small time constants. Using eigenvectors, a statement regarding the system can be made, for example, how the original variable must be superimposed. In the example, the equation to

determine the eigenvector reads:

$$\begin{aligned} J \cdot \underline{x} &= \lambda \cdot \underline{x} \\ \begin{bmatrix} -k_1 & k \\ k_1 & -2k \end{bmatrix} \cdot \begin{pmatrix} A \\ B \end{pmatrix} &= \lambda \cdot \begin{pmatrix} A \\ B \end{pmatrix}. \end{aligned} \quad (8.23)$$

This leads to an algebraic system where a single variable can be freely chosen. For $B = 1$ one can determine both eigenvectors. The second row from earlier gives:

$$\begin{aligned} k_1 \cdot A - 2k &= \lambda \\ \rightarrow A &= \frac{\lambda + 2k}{k_1} = \frac{\lambda}{k_1} + \frac{2k}{k_1}. \end{aligned} \quad (8.24)$$

For the eigenvalue λ_1 (slow) this yields:

$$A^1 = \frac{\lambda_1 + 2k}{k_1} \approx \frac{2k}{k_1} \approx \epsilon' \ll 1 \quad (8.25)$$

and for the eigenvector λ_2 (fast):

$$A^2 = \frac{\lambda_2 + 2k}{k_1} \approx \frac{-k_1 + 2k}{k_1} \approx -1 + \epsilon'. \quad (8.26)$$

The eigenvectors forming the transformation matrix T are then

$$T = \begin{bmatrix} \epsilon' & -1 + \epsilon' \\ 1 & 1 \end{bmatrix} \rightarrow T^{-1} = \begin{bmatrix} 1 & 1 - \epsilon' \\ -1 & \epsilon' \end{bmatrix}. \quad (8.27)$$

For the new variables z , this yields the following mapping

$$\underline{z} = T^{-1} \cdot \underline{c} = \begin{bmatrix} 1 & -1 - \epsilon' \\ -1 & \epsilon' \end{bmatrix} \cdot \begin{pmatrix} A \\ B \end{pmatrix}, \quad (8.28)$$

and therefore:

$$z_1 = (A + (1 - \epsilon') B) \approx A + B \text{ with } \tau_1 = \frac{1}{|Re(\lambda_1)|} \gg 1 \text{ (slow)}$$

$$z_2 = (-A + \epsilon' \cdot B) \approx -A \text{ with } \tau_2 = \frac{1}{|Re(\lambda_2)|} \approx \frac{1}{k_1} \text{ (fast)}. \quad (8.29)$$

Here, the new states can be assigned directly to old variables. Variable z_1 represents a pool of both A and B while z_2 presents only variable A . For the fast variables z_2 , quasi-stationarity is assumed. This means that the fast state is considered in steady state ($\dot{z}_2 = 0$). If the ratio $\tau_{slow}/\tau_{fast} \approx 100$ simulation, studies show that the assumption is fair. One obtains for the second state:

$$\dot{z}_2 = 0 \Rightarrow A = \frac{k}{k_1} B. \quad (8.30)$$

The transformation shows that $A = \frac{k}{k_1} B$ represents the equilibrium relation of the reaction. Figure 8.3 shows simulation results for different values of $\epsilon = k/k_1$. As can be seen for small values, the original system and the reduced one are nearly identical.

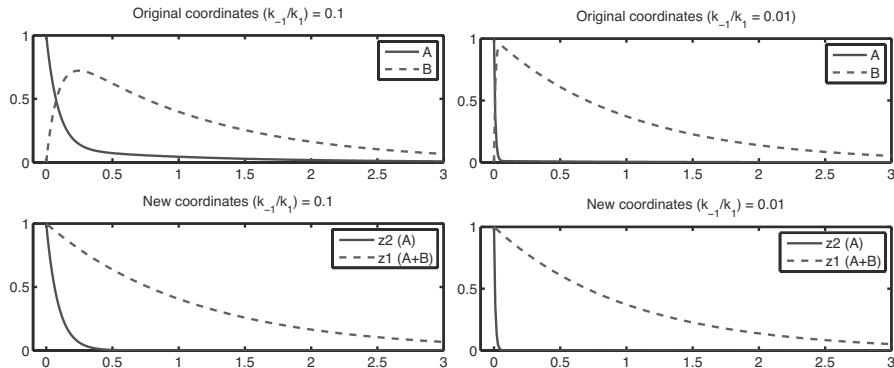


FIGURE 8.3: Time course of the original system (first row) and the transformed system for different values of k/k_1 . Left: Original and transformed system are different due to a high value of k/k_1 . Right: Original and transformed system show almost the same behavior since k/k_1 is very small.

8.1.2 Classification of fast and slow reactions

Once the eigenvalues for a large system are determined, a question arises which reactions contribute to the fast modes. This case emerge when one determines parameters via parameter estimation and carries out the analysis directly. Conversely if it is known that fast reactions are present in the system (if the single parameters are taken from literature or data base systems), a question that arises is how the system can be simplified without prior linearization and calculation of the eigenvalues. The starting point for the following considerations is that the system has the form:

$$\dot{\underline{c}} = N \cdot \underline{r}.$$

Now two cases are considered to approach the two problems mentioned above.

Case i: Eigenvalues and eigenvectors are calculated directly, for example, from a parameter estimation analysis. Corresponding to the eigenvalues analysis, matrix T^{-1} can be split (by sorting the columns of N) into a fast component (f) and a slow component (s):

$$T^{-1} = \begin{pmatrix} T_s^{-1} \\ T_f^{-1} \end{pmatrix}. \quad (8.31)$$

Introducing the transformation:

$$T^{-1} \underline{\dot{c}} = T^{-1} N \underline{r} \quad (8.32)$$

yields for the z system

$$\underline{\dot{z}} = \begin{pmatrix} T_s^{-1} \\ T_f^{-1} \end{pmatrix} (N_s \ N_f) \begin{pmatrix} \underline{r}_s \\ \underline{r}_f \end{pmatrix} \quad (8.33)$$

and for the slow variables \underline{z}_s one obtains:

$$\underline{\dot{z}}_s = T_s^{-1} N_s \underline{r}_s + T_s^{-1} N_f \underline{r}_f. \quad (8.34)$$

Should no fast reactions have an influence on $\underline{\dot{z}}_s$ (the z_{s_i} should be slow variables) then what must be valid: $T_s^{-1} N_f = \underline{0}$. This means that the influence of the fast reactions is then excluded. As now it is not known which reactions are fast or slow, fast reactions can be determined from the relation

$$T_s^{-1} N \stackrel{!}{=} \underline{0}. \quad (8.35)$$

Inspection of the columns of $T_s^{-1} N$ and looking for entries ≈ 0 then reveal the fast reactions.

Case ii: Eigenvalues and eigenvectors are not known; slow and fast reactions are distinguishable, for example, from data from a data base. For the splitting of the reactions into fast and slow ones, we get:

$$\begin{aligned} \underline{\dot{c}} &= (N_s \ N_f) \begin{pmatrix} \underline{r}_s \\ \underline{r}_f \end{pmatrix} \\ &= N_s \underline{r}_s + N_f \underline{r}_f. \end{aligned} \quad (8.36)$$

Introduce now a transformation with a matrix T^* so that:

$$T^* \underline{\dot{c}} = T^* N_l \underline{r}_l + T^* N_s \underline{r}_s. \quad (8.37)$$

This enables determining the slow modes as the fast reactions will be quasi-hidden:

$$T^* N_f = \underline{0} \rightarrow N_f^T T^{*T} = \underline{0} \quad (8.38)$$

where T^{*T} represents the null space of N_f^T . With T^* , the transformation of the whole system can be carried out. Then it is guaranteed that in the new coordinates no fast reactions are present.

EXAMPLE 8.2 *Simple reaction system (continued).*

The system description using the stoichiometric equation N instead of the Jacobian matrix J reads:

$$\dot{\underline{c}} = J \underline{c} = N \underline{r} = \begin{pmatrix} -1 & 0 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \quad (8.39)$$

where r_1 represents a fast reaction and r_2 represents a slow reaction.

Case i: Matrix T is determined from the parameters. Then the following hold:

$$T_s^{-1} N = (1 \ 1 - \epsilon') \begin{pmatrix} -1 & 0 \\ 1 & -1 \end{pmatrix} = (-\epsilon' \ -1 + \epsilon'). \quad (8.40)$$

For small ϵ' this yields:

$$T_s^{-1} N \approx [0 \ -1]. \quad (8.41)$$

The entry in the first columns vanishes; hence r_1 is the fast reaction.

Case ii: It is known that r_1 is fast and r_2 is slow. Following the procedure above we get:

$$N_f^T T^{*T} = 0 \longrightarrow (-1 \ 1) T^{*T} = 0. \quad (8.42)$$

In this case, the null space can be easily determined:

$$T^{*T} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \longrightarrow T^* = (1 \ 1). \quad (8.43)$$

Applying the transformation to \underline{c} yields:

$$\begin{aligned} T^* \dot{\underline{c}} &= (1 \ 1) N_s r_s = (1 \ 1) \begin{pmatrix} 0 \\ -1 \end{pmatrix} r_2 \\ \rightarrow (A + B) \dot{r}_2 &= -r_2. \end{aligned} \quad (8.44)$$

Also, here one obtains the results that the components A and B form a pool of values and that the dynamics is then only due to the slow reactions. Every combination of the null space vectors of N_f^T eliminates the fast modes. One distinguishable linear combination represents the matrix T^{-1} , which transforms the system such that time constants of the new system can be directly assigned to the eigenvalues. This is an interesting result since it allows different possibilities for a larger system to get a reduced model description; that is, the number of differential equations is reduced and replaced by algebraic equations that describe the interrelation between state variables.

8.1.3 Complete procedure for model reduction

The results from earlier can now be used for a general model reduction, where the systems can also be non linear [1] in the state variables. One can divide the system of equations as following (s slow, f fast):

$$\dot{\underline{c}} = N_s \underline{r}_s + N_f \underline{r}_f \quad (8.45)$$

with p fast reactions and $q - p$ slow reactions. A common approach to simplify the system is scaling. In this case, scaling with respect to the smallest of the fast reaction rate constants $r_{f_i} = r_{max,i}$ $f(\underline{c})$ can be carried out:

$$\underline{r}_f = r_{max}^* \underbrace{\begin{pmatrix} \frac{r_{max,1}}{r_{max}^*} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \ddots \end{pmatrix}}_{=: D} \bar{\underline{r}}_f \quad (8.46)$$

where r_{max}^* is the smallest value of all $r_{max,i}$. This results in a system with $\epsilon = 1/r_{max}^*$ and the above diagonal matrix D :

$$\frac{dc}{dt} = N_s \underline{r}_s + \epsilon^{-1} N_f D \bar{\underline{r}}_f. \quad (8.47)$$

The following considerations should be valid for the case: N_f and $\frac{\partial \bar{\underline{r}}_f}{\partial \underline{c}}$ have rank p . Now if ϵ approaches zero, this implies that $\bar{\underline{r}}_f$ goes to zero as well. Introducing the following function \underline{z}_f :

$$\underline{z}_f = \lim_{\epsilon \rightarrow 0} \frac{D \bar{\underline{r}}_f}{\epsilon}; \quad (8.48)$$

this function now shall represent this limiting value and one obtains:

$$\dot{\underline{c}} = N_s \underline{r}_s + N_f \underline{z}_f, \quad (8.49)$$

where \underline{z}_f contains implicitly the state variables. \underline{z}_f can be specified as follows: The limiting case $\bar{\underline{r}}_f = 0$ applies and hence also $\frac{\partial \bar{\underline{r}}_f}{\partial t} = 0$ applies and one obtains:

$$\frac{\partial \bar{\underline{r}}_f}{\partial t} = \frac{\partial \bar{\underline{r}}_f}{\partial \underline{c}} \frac{\partial \underline{c}}{\partial t} = \frac{\partial \bar{\underline{r}}_f}{\partial \underline{c}} N_s \cdot \underline{r}_s + \frac{\partial \bar{\underline{r}}_f}{\partial \underline{c}} N_f \underline{z}_f = 0. \quad (8.50)$$

As the derivative now has full rank, the inverse can be derived and one obtains for \underline{z}_s :

$\underline{z}_f = - \left[\frac{\partial \bar{\underline{r}}_f}{\partial \underline{c}} N_f \right]^{-1} \frac{\partial \bar{\underline{r}}_f}{\partial \underline{c}} N_s \underline{r}_s.$

(8.51)

Inserting the previous equation results in:

$$\frac{dc}{dt} = N_s \underline{r}_s - N_f \left[\frac{\partial \bar{\underline{r}}_f}{\partial \underline{c}} N_f \right]^{-1} \frac{\partial \bar{\underline{r}}_f}{\partial \underline{c}} \cdot N_s \underline{r}_s. \quad (8.52)$$

The fast modes are now eliminated in the system though the number of the state variables has not been reduced. A system of equations of only the $n -$

p slow modes can be obtained by the transformation used above. The new variables \underline{w} are obtained by:

$$\underline{w} = K^T \underline{c} \rightarrow \dot{\underline{w}} = K^T \dot{\underline{c}} \quad (8.53)$$

where K is the null space of N_f^T . Besides the equations of the original system also the equilibrium conditions for the fast reactions should be used for the transformation as shown in the following example.

EXAMPLE 8.3 Simple reaction system (continued).

To apply the described method the system is reformulated:

$$\begin{pmatrix} \dot{A} \\ \dot{B} \end{pmatrix} = \underbrace{\begin{pmatrix} 0 \\ -1 \end{pmatrix}}_{N_s} \underbrace{r_2}_{\underline{r}_s} + \underbrace{\begin{pmatrix} -1 \\ 1 \end{pmatrix}}_{N_f} \cdot \underbrace{r_1}_{\underline{r}_f}. \quad (8.54)$$

Fast reaction r_1 is given as:

$$r_1 = k_1 A - k_{-1} B = k_1 (A - K_S B), \quad (8.55)$$

with parameter k_1 as scaling factor $\epsilon = 1/k_1$. Then we obtain:

$$\frac{d\bar{r}_1}{dc} = (1 \quad -K_S). \quad (8.56)$$

Inserting in the equation for z_f (in this case scalar):

$$z_f = \frac{K_S}{1 + K_S} r_2. \quad (8.57)$$

For the state variables this results in the following two differential equations:

$$\dot{A} = -\frac{K_S}{1 + K_S} r_2 \approx 0 \quad (8.58)$$

$$\dot{B} = -\frac{1}{1 + K_S} r_2 \approx -r_2. \quad (8.59)$$

Since $k_1 \gg k_{-1}$, K_S approaches zero. This means that A is the fast variable reaching the equilibrium that is a small value. As described above, a transformation with the null space of N_f^T is further possible.

8.1.4 Singular perturbation theory

Singular perturbation theory is an approximation for non linear models running on different time scales. With it, one tries to cleverly scale the system's equation so that statements can be made concerning the system's characteristics via the newly defined parameters. The procedure will be clarified by the next example.

EXAMPLE 8.4 Enzyme catalyzed reaction (for details of the mechanism, see the enzyme kinetics chapter).

The equation of the reaction reads:



The differential equations for the substrate S and the complex of the enzyme and the substrate ES with the conservation relation $E_0 = E + ES$ read:

$$\begin{aligned}\dot{S} &= -k_1 E S + k_{-1} ES = -k_1 \cdot (E_0 - ES) S + k_{-1} ES \\ \dot{ES} &= k_1 (E_0 - ES) S - (k_{-1} + k_2) ES,\end{aligned}\quad (8.60)$$

with the initial conditions $S(t = 0) = S_0$ and $ES(t = 0) = 0$ (at beginning the enzyme is in the free form). A normalization or scaling reduces the number of parameters and simplifies the transferability to other system examples. The following scaling is suggested here:

$$s = \frac{S}{S_0}, \quad es = \frac{ES}{E_0}, \quad \tau = k_1 E_0 t. \quad (8.61)$$

$[\tau]$ is dimensionless: $\frac{1}{mol \cdot h}$. To obtain the following equations for the scaled states s and es insert:

$$\begin{aligned}\frac{ds}{d\tau} &= -(1 - es) s + \frac{k_{-1}}{k_1} \frac{1}{S_0} es \\ \frac{des}{d\tau} &= (1 - es) s \cdot \frac{S_0}{E_0} - \frac{(k_{-1} + k_2)}{k_1 S_0} \frac{S_0}{E_0} es.\end{aligned}\quad (8.62)$$

Introducing dimensionless parameters

$$K = \frac{k_{-1}}{k_1 S_0} = \frac{K_S}{S_0}, \quad \lambda = \frac{k_2}{k_1 S_0}, \quad \epsilon = \frac{E_0}{S_0} \quad (8.63)$$

the following equations are obtained:

$$\begin{aligned}\frac{ds}{d\tau} &= -(1 - es) s + K es = (K + s) es - s \\ \frac{des}{d\tau} &= (1 - es) s \cdot \frac{1}{\epsilon} - (K + \lambda) \frac{1}{\epsilon} es \\ &= -(K + \lambda + s) es + s\end{aligned}\quad (8.64)$$

The last equation can be rewritten as:

$$\epsilon \cdot \frac{des}{d\tau} = -(K + \lambda + s) es + s. \quad (8.65)$$

In this form, a singular perturbation problem is presented, which for $\epsilon \rightarrow 0$ is equivalent to the assumption of a steady state for es . The algebraic equation reads:

$$0 = -(K + \lambda + s) es + s \quad \longrightarrow \quad es = \frac{s}{K + \lambda + s}. \quad (8.66)$$

Calculating $\dot{P} \sim es$ as above we obtain once again the Michaelis-Menten kinetics. Figure 8.4 shows the course of the trajectory in the s - es phase plane.

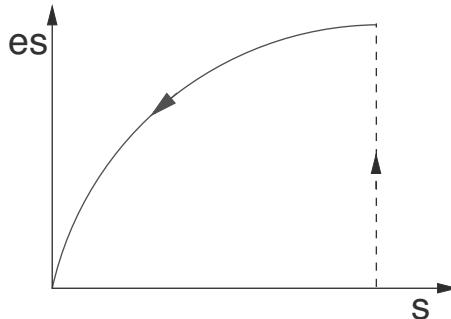


FIGURE 8.4: Idealized course of the trajectory in the phase plane. Solid line: dynamics in the τ time. Dashed line: dynamics in the σ time.

For systems of differential equations, one is free to select the initial conditions. Here, a differential equation is replaced by an algebraic equation and leads to the problem that the initial condition for es can no longer be chosen freely as for the initial system; here, the value is fixed by the algebraic relationship:

$$es(0) = \frac{s(0)}{K + \lambda + s(0)} \quad (8.67)$$

and is not equal to zero as for the original system. In other words, in the current form there is an inconsistency and we ask in which way we can resolve the problem. So far, we have neglected the fast response of the system. So, can the behavior be determined for a second (fast) time scale? In this new time scale es shall be set from $es(t = 0) = 0$ to the value obtained above. Another scaling will be carried out therefore. The analysis of the system for small times proceeds with $\sigma = \frac{\tau}{\epsilon}$. Very small times τ are transformed and turned into larger times ($\tau \approx \epsilon \rightarrow \sigma \approx 1$). The new equations read:

$$\begin{aligned} \frac{ds}{d\tau} &= \epsilon \cdot \left((K + s) es - s \right) \\ \frac{des}{d\tau} &= -(K + \lambda + s) es + s. \end{aligned} \quad (8.68)$$

Now s can be considered quasi-stationary with the initial condition $s = s(t=0) = s_0$. For es this yields the following dynamics:

$$\frac{des}{d\sigma} = s_0 - (K + \lambda + s_0) es. \quad (8.69)$$

This is an ordinary differential equation with the solution:

$$es(\sigma) = -\frac{s(0)}{K + \lambda + s(0)} \left(1 + e^{-(K+\lambda+s(0))\sigma} \right) \quad (8.70)$$

and it describes the dynamics of es . One determines for the initial value and the large times:

$$es(0) = 0, \quad \lim_{\sigma \rightarrow \infty} es = \frac{s_0}{K + \lambda + s_0}, \quad (8.71)$$

which corresponds to the above determined value. The idealized course of the curve is included in Figure 8.4: The dashed line gives the fast dynamics (σ time). S does not change. The slow trajectory is then reached (solid line) and the system then has the course of S coupled again with ES (τ time). If the original system is simulated, one would observe the passage from the fast to the slow trajectory.

8.2 Sensitivity Analysis

From the term sensitivity analysis, one understands that it is related to the effect of variation of the kinetic parameters or the model structure on the progression of state variables. A system that exhibits small sensitivity then is considered robust. Sensitivities can be defined and calculated for a number of properties:

- Sensitivity of a function of a stoichiometric network, for example, yields coefficients for products from metabolism with respect to variations of stoichiometry (this can be realized by performing mutations).
- Sensitivity of state variables in case of parameter variations or structural variations. This concerns, for example, the dynamic characteristics like response times, response duration, oscillation amplitudes, phase, gain factors, etc.
- Sensitivity of a signal response curve (steady state characteristics) for variation for the total amount of a protein in the network.

Here, the influence of kinetic parameters on stationary characteristic curves is investigated and finally the method is extended to dynamic systems. Sensitivity analysis is considered today a standard procedure and a fundamental

instrument for the analysis of cellular systems. Of special importance are applications during model calibration and for experimental design: if a parameter possesses low sensitivity, it can hardly be estimated or not at all. Moreover, whenever a system is designed, that is, modified or engineered with the aim to improve it, a sensitivity analysis points out crucial parts that need the most attention.

8.2.1 Definition of the parametric sensitivity w_{ij}

Formally, the parametric sensitivity w_{ij} denotes the influence of the parameter p_j on the quantity (or state variables) x_i :

$$w_{ij} = \frac{\Delta x_i}{\Delta p_j} \approx \frac{dx_i}{dp_j} \quad \text{for the limiting case } \Delta p_j \rightarrow 0. \quad (8.72)$$

To compare sensitivities with one another, it is often useful to use the normalized sensitivities which are defined as follows:

$$\bar{w}_{ij} = \frac{\Delta x_i / x_i}{\Delta p_j / p_j} = \frac{\Delta x_i}{\Delta p_j} \frac{p_j}{x_i} \approx \frac{dx_i}{dp_j} \frac{p_j}{x_i}. \quad (8.73)$$

EXAMPLE 8.5 Michaelis-Menten kinetics.

The influence of both kinetic parameters r_{max} and K on the quantity r is studied. The kinetics read:

$$r = r_{max} \frac{S}{K + S}.$$

For the first sensitivity, this results in:

$$w_{r,K} = \frac{dr}{dK} = \frac{-r_{max} S}{(K + S)^2} \quad (8.74)$$

$$\bar{w}_{r,K} = \frac{-r_{max} S}{(K + S)^2} \cdot \frac{K \cdot (K + S)}{r_{max} S} = -\frac{K}{K + S}. \quad (8.75)$$

Figure 8.5 on left illustrates the progression in dependence on S .

For the second sensitivity, this results in:

$$w_{r,r_{max}} = \frac{dr}{dr_{max}} = \frac{S}{K + S} \quad (8.76)$$

$$\bar{w}_{r,r_{max}} = \frac{S}{K + S} \cdot \frac{r_{max} \cdot (K + S)}{r_{max} S} = 1. \quad (8.77)$$

The results can be interpreted as follows: For parameter K a positive deviation of the value in the proximity of 0 leads to reduction of r , therefore the value is negative; for large values of S the influence becomes weaker. For parameter r_{max} a deviation of the value leads to a uniform deviation of r over the whole range of S .

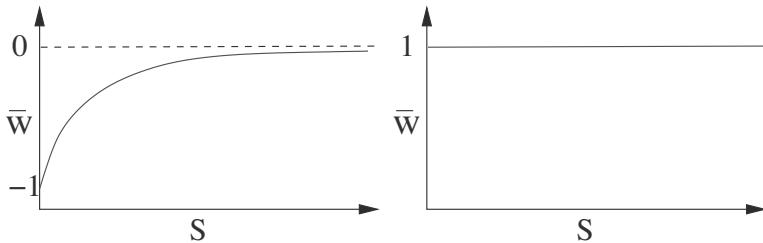


FIGURE 8.5: Behavior of both normalized sensitivities over S .

EXAMPLE 8.6 Reversible kinetics.

For the kinetics:

$$r = k_{max} \cdot \left(S_1 - \frac{S_2}{K} \right) \quad (8.78)$$

one obtains the following sensitivity in relation to the equilibrium constants K :

$$w_{r,K} = \frac{dr}{dK} = k_{max} \cdot \frac{S_2}{K^2} \quad (8.79)$$

$$\bar{w}_{r,K} = k_{max} \cdot \frac{S_2}{K^2} \cdot \frac{K}{k_{max} \cdot \left(S_1 - \frac{S_2}{K} \right)} = \frac{S_2}{K S_1 - S_2}. \quad (8.80)$$

One can see from the equation that the normalized sensitivity has a singularity when S_1 and S_2 are in equilibrium. The denominator goes to zero and the normalized sensitivity becomes large.

In subsequent sections often the derivative of the reaction rates in vector \underline{r} are differentiated with respect to the concentration of the involved metabolites \underline{c} . Since every rate is differentiated with respect to every concentration, the result is a matrix \mathbf{E} which is defined as:

$$\mathbf{E} = \frac{dr}{d\underline{c}} = \begin{pmatrix} \frac{dr_1}{dc_1} & \frac{dr_1}{dc_2} & \frac{dr_1}{dc_3} & \dots \\ \frac{dr_2}{dc_1} & \frac{dr_2}{dc_2} & \frac{dr_2}{dc_3} & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}. \quad (8.81)$$

For q rates and n metabolites, it follows that \mathbf{E} is a $q \times n$ matrix. We will use throughout the text \mathbf{E} for this type of derivative if nothing different is explained.

8.2.2 Sensitivity analysis using Hearne's method

To obtain a ranking of the parameters, a procedure is proposed, the procedure by Hearne [2]. To start with a general non linear equation consider:

$$\underline{0} = g(\underline{z}, \underline{p}) \quad (8.82)$$

where \underline{z} is the vector representing the state variables and \underline{p} represents the parameters' vector. The deviation of the parameters is not arbitrary but rather limited. Now a new parameters' combination is sought with condition

$$\left(\frac{\Delta p_1}{p_1} \right)^2 + \left(\frac{\Delta p_2}{p_2} \right)^2 + \dots = 1, \quad (8.83)$$

so that the curve $\underline{0} = g(\underline{z}, \underline{p})$ achieves a maximal deflection; in the case of two parameters, this signifies that the allowed variations Δp_1 and Δp_2 lie on the unit circle. Figure 8.6 illustrates the method.

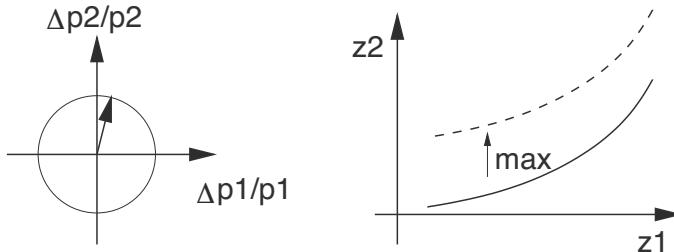


FIGURE 8.6: Hearne method: A new combination of parameters is sought that changes the courses maximally.

The following values are obtained for Equation (8.83) in the case of two parameters:

$\Delta p_1/p_1$	$\Delta p_2/p_2$
0.10	0.99
0.20	0.98
0.30	0.95
0.40	0.92
0.50	0.87
0.71	0.71

How the system behaves under small perturbations will now be investigated. The non linearity is approximated using a Taylor series. For function g we get for the small deviations $\Delta \underline{p}$:

$$g(\underline{p} + \Delta \underline{p}, \underline{z}) = g(\underline{p}, \underline{z}) + \frac{\partial g}{\partial \underline{p}} \Big|_{\underline{p}} \Delta p_1 + \frac{\partial g}{\partial \underline{p}_2} \Big|_{\underline{p}} \Delta p_2 + \dots \quad (8.84)$$

so that one obtains for the variation Δg of g :

$$g(\underline{p} + \Delta \underline{p}, \underline{z}) - g(\underline{p}, \underline{z}) = \Delta g = \frac{\partial g}{\partial p_1} \Delta p_1 + \frac{\partial g}{\partial p_2} \Delta p_2 + \dots \quad (8.85)$$

Relating to g :

$$\begin{aligned} \frac{\Delta g}{g} &= \frac{\partial g}{\partial p_1} \frac{\Delta p_1}{g} + \frac{\partial g}{\partial p_2} \frac{\Delta p_2}{g} + \dots \\ &= \frac{\partial g}{\partial p_1} \frac{\Delta p_1}{p_1} \frac{p_1}{g} + \frac{\partial g}{\partial p_2} \frac{\Delta p_2}{p_2} \frac{p_2}{g} + \dots \end{aligned} \quad (8.86)$$

The normalized sensitivities are now used, $\bar{w}_i = \frac{\partial g}{\partial p_i} \frac{p_i}{g}$, and one obtains:

$$\frac{\Delta g}{g} = \bar{w}_1 \frac{\Delta p_1}{p_1} + \bar{w}_2 \frac{\Delta p_2}{p_2} + \dots = \underline{w}^T \underline{s} \quad (8.87)$$

with the vectors:

$$\underline{s} = \begin{pmatrix} \Delta p_1/p_1 \\ \Delta p_2/p_2 \\ \vdots \end{pmatrix}, \quad \underline{w} = \begin{pmatrix} \bar{w}_1 \\ \bar{w}_2 \\ \vdots \end{pmatrix}. \quad (8.88)$$

The relative variation must be summed over z and finally maximized. The functional is given by:

$$\begin{aligned} \Omega &= \int_{z_0}^z \left(\frac{\Delta g}{g} \right)^2 dz = \int_{z_0}^z (\underline{w}^T \underline{s})^2 dz = \int_{z_0}^z \underline{s}^T \underline{w} \underline{w}^T \underline{s} dz \\ &= \underline{s}^T \int_{z_0}^z \underline{w} \underline{w}^T dz \quad \underline{s} = \underline{s}^T W \underline{s} \end{aligned} \quad (8.89)$$

where W is the matrix of the normalized sensitivities. The objective function and constraints read:

max _{\underline{s}}	Ω	.
s.t.	$\underline{s}^T \underline{s} = 1$	

(8.90)

To solve the problem, constraints can be incorporated in the objective function. To determine the maximum, the derivative with respect to the parameters must be zero (here the rules for derivatives with respect to vectors are applied, which are summarized in the Appendix):

$$\begin{aligned} \frac{d}{d\underline{s}} \left(\underline{s}^T W \underline{s} - \lambda (\underline{s}^T \underline{s} - 1) \right) &= 0 \\ \longrightarrow 2 W \underline{s} - 2 \lambda \underline{s} &= 0. \end{aligned} \quad (8.91)$$

The solution of the problem leads to the calculation of eigenvalues of matrix W .

$$W \underline{s} = \lambda \underline{s}. \quad (8.92)$$

The eigenvector of the maximal eigenvalues provides the parameter deviation that deflects the curve g maximally (for more theoretical background we refer to the original publication from Hearne).

EXAMPLE 8.7 Michaelis-Menten kinetics (*continued*).

For Michaelis-Menten kinetics, one obtains for \bar{w} :

$$\bar{w} = \left[1 - \frac{K}{K + c_S} \right]. \quad (8.93)$$

A numerical calculation with defined parameters values is shown in Figure 8.7.

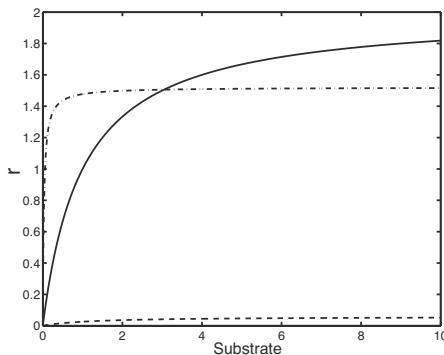


FIGURE 8.7: Simulation study with varied parameters values. The solid line shows r , dashed one shows the maximal deviation and dot-dashed lines show the minimal deviation. Note that in any case the parameter deviation has to fulfill the constraint Equation (8.83).

8.2.3 Sensitivity matrix

An alternative possibility to obtain the ranking of the parameters for the steady state situation is to place all the sensitivities in a single matrix W_R . One obtains for the state variables \underline{x} and the parameters vector \underline{p} :

$$W_R = \begin{pmatrix} \frac{dx_1}{dp_1} & \frac{dx_1}{dp_2} & \dots \\ \frac{dx_2}{dp_1} & \frac{dx_2}{dp_2} & \dots \\ \dots & \dots & \dots \end{pmatrix} = \begin{pmatrix} w_{11} & w_{12} & \dots \\ w_{21} & w_{22} & \dots \\ \dots & \dots & \dots \end{pmatrix}. \quad (8.94)$$

The ranking can be determined from an analysis of rows and columns. Considering, for example, rows in the matrix the length of this vector can be taken as a measure of the strength of all parameters on state variables. For the first state variable x_1 one gets with normalized sensitivities:

$$\omega_{x1} = \sqrt{\bar{\omega}_{11}^2 + \bar{\omega}_{12}^2 + \bar{\omega}_{13}^2 + \dots} \quad (8.95)$$

To determine the most important parameter the columns are considered; the strength of the influence of the first parameter p_1 on all states can be determined by:

$$\omega_{p1} = \sqrt{\bar{\omega}_{11}^2 + \bar{\omega}_{21}^2 + \bar{\omega}_{31}^2 + \dots} \quad (8.96)$$

EXAMPLE 8.8 Comparison of two kinetics.

The following two reaction kinetics will be compared:

$$r_1 = k \frac{S^2}{S^2 + K^2} \quad \text{and} \quad r_2 = k \left(1 - e^{-S/K}\right). \quad (8.97)$$

Plotting both kinetics reveals a very similar behavior (not shown). The scaled derivatives with respect to k are equal in both cases and have the numerical value 1 (see above). The derivatives with respect to K are:

$$\begin{aligned} \bar{\omega}_{r1,K} &= \frac{-2K^2}{S^2 + K^2} \\ \bar{\omega}_{r2,K} &= \frac{-S}{K} \frac{e^{-S/K}}{1 - e^{-S/K}}. \end{aligned} \quad (8.98)$$

If we are interested in evaluation of $W_R(S = K)$ we get the matrix:

$$W_R = \begin{pmatrix} 1 & -1 \\ 1 & -0.58 \end{pmatrix}. \quad (8.99)$$

We conclude that the influence of k is stronger than that of K (comparison of the two columns) and that the influence of the parameters on r_1 is stronger than on r_2 (comparison of the two rows).

8.2.4 Sensitivities for dynamic systems

For a dynamic system of the form $\dot{\underline{x}} = \underline{f}(\underline{x}, \underline{p})$ where sensitivities $w_{ij} = \frac{dx_i}{dp_j}$ are time dependent, the differential equation is used as a starting point. To calculate the w_{ij} the initial equation must be differentiated with respect to the parameters. One obtains:

$$\frac{d}{dp} \dot{\underline{x}} = \frac{d}{dp} \underline{f}(\underline{x}(p), \underline{p}) \quad \rightarrow \quad \frac{d}{dt} \frac{d\underline{x}}{dp} = \frac{d\underline{f}}{d\underline{x}} \frac{d\underline{x}}{dp} + \frac{d\underline{f}}{dp}. \quad (8.100)$$

To summarize this equation reads:

$$\boxed{\frac{d}{dt} W = J W + F_{\underline{p}}, \quad W(t=0) = 0,} \quad (8.101)$$

where the matrix W is given by the derivatives of vector \underline{x} with respect to vector \underline{p} ; J represents the Jacobian and the matrix $F_{\underline{p}}$ represents differentiation of functions \underline{f} with respect to the parameter vector. The initial values as a rule are freely selectable and do not depend on parameters; hence the derivative is zero. Note, that this leads to additional equations that must be solved together with the original system for the state variables. Considering two state variables with two parameters, we get the following equations system:

$$\frac{d}{dt} \begin{pmatrix} \frac{dx_1}{dp_1} & \frac{dx_1}{dp_2} \\ \frac{dx_2}{dp_1} & \frac{dx_2}{dp_2} \end{pmatrix} = \begin{pmatrix} \frac{df_1}{dx_1} & \frac{df_1}{dx_2} \\ \frac{df_2}{dx_1} & \frac{df_2}{dx_2} \end{pmatrix} \begin{pmatrix} \frac{dx_1}{dp_1} & \frac{dx_1}{dp_2} \\ \frac{dx_2}{dp_1} & \frac{dx_2}{dp_2} \end{pmatrix} + \begin{pmatrix} \frac{df_1}{dp_1} & \frac{df_1}{dp_2} \\ \frac{df_2}{dp_1} & \frac{df_2}{dp_2} \end{pmatrix}; \quad (8.102)$$

hence four equations have to be solved together with the original equation.

EXAMPLE 8.9 *Example with a single state variable and a single parameter.*

For the system

$$\dot{x} = b - x \quad (8.103)$$

one determines the differential equation for the sensitivity as follows:

$$\frac{d}{dt} \left(\frac{dx}{db} \right) = -1 \left(\frac{dx}{db} \right) + 1 \rightarrow \dot{w} = -w + 1 \quad (8.104)$$

with the solution

$$w(t) = (1 - e^{-t}). \quad (8.105)$$

8.3 Robustness in Stoichiometric Networks

In the previous section, the sensitivity of state variables is analyzed with respect to changes of kinetic parameters or an external disturbance. In the following analysis, systems are considered that are not sensitive to changes in some of the input variables. In contrast, the variables only show a transient dynamics and come back exactly to the steady state value that they had before the system was disturbed. This observation was already studied in the signaling section with an example and the behavior was called adaptation.

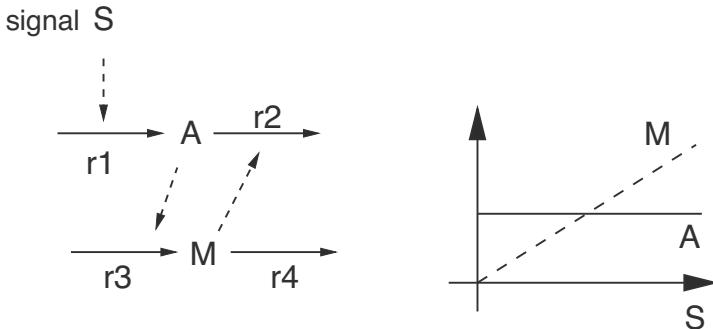


FIGURE 8.8: Example for ACR. Both metabolites are interacting. Compound A shows ACR while M does not.

There are theoretical concepts available that allow to check if one (or more) state variable possesses this characteristic, called ACR (absolute concentration robustness). The system shown in Figure 8.8 where two metabolites are interacting is considered.

The differential equations are as follows:

$$\begin{aligned}\dot{A} &= r_1(S) - r_2(A, M) = k_1 S - k_2 A M \\ \dot{M} &= r_3(A) - r_4 = k_3 A - k_4.\end{aligned}\quad (8.106)$$

Rate r_4 is assumed first; do not depend on M . The consequences are discussed later on. The steady state solution can be calculated easily and one gets:

$$A = \frac{k_3}{k_4}; \quad M = \frac{k_1 k_3}{k_2 k_4} S. \quad (8.107)$$

It could be seen that A is a constant and doesn't depend on the input S into the system while M is a linear function of S (see Figure 8.8). As long as r_4 is not a function of M we also can choose more complex rate equations for r_1 , r_2 and r_3 and the statement that $A \neq f(S)$ is valid, too. In the following a systematic approach is introduced that allows to check ACR for selected state variables [3]. It is based on the basic equation that characterizes steady states:

$$\underline{\dot{c}} = \underline{0} = N \underline{r}(\underline{c}, \underline{p}) \quad (8.108)$$

with the stoichiometric matrix N , concentrations in steady state \underline{c} and a parameter vector \underline{p} . Now let's assume that some state variables in \underline{c} possess ACR and the vector is divided accordingly into two parts: state variables \underline{c}_M that don't show ACR and \underline{c}_A that show ACR. Now let's disturb the system by changing one parameter in rate $r_j(p_i)$ and a new steady state is reached. As an example, one equation with two rate equations r_1 and r_2 is considered:

$$0 = [1 \ -1] \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} = r_1(p, c_1, c_2) - r_2(p, c_1, c_2). \quad (8.109)$$

From the total derivative of this equation, it could be seen that a small change in parameter p will lead to new values in c_1 and c_2 in the new steady state. The total derivative is given by:

$$\begin{aligned} 0 &= \frac{\partial r_1}{\partial p} dp + \frac{\partial r_1}{\partial c_1} dc_1 + \frac{\partial r_1}{\partial c_2} dc_2 - \\ &\quad \frac{\partial r_2}{\partial p} dp - \frac{\partial r_2}{\partial c_1} dc_1 - \frac{\partial r_2}{\partial c_2} dc_2. \end{aligned} \quad (8.110)$$

The first three summands can be expressed also as:

$$0 = \frac{\partial r_1}{\partial p} dp \frac{r_1}{p} \frac{p}{r_1} + \frac{\partial r_1}{\partial c_1} dc_1 \frac{r_1}{c_1} \frac{c_1}{r_1} + \frac{\partial r_1}{\partial c_2} dc_2 \frac{r_1}{c_2} \frac{c_2}{r_1} - \dots \quad (8.111)$$

Rearranging of this equation leads to:

$$0 = \left(\bar{\omega}_{r_1,p} \frac{dp}{p} + \bar{\omega}_{r_1,c_1} \frac{dc_1}{c_2} + \bar{\omega}_{r_1,c_2} \frac{dc_2}{c_2} \right) r_1 - \dots \quad (8.112)$$

with scaled sensitivities $\bar{\omega}$. The sensitivities could be arranged in matrices \mathbf{P} , \mathbf{M} , \mathbf{A} depending on parameters and state variables, respectively, and for the general case the following condition hold true:

$$0 = N \text{ diag}(\bar{\omega}) (\mathbf{P} \Delta \underline{p} + \mathbf{M} \Delta \underline{c}_M + \mathbf{A} \Delta \underline{c}_A) \quad (8.113)$$

with

$$\begin{aligned} P_{ij} &= \bar{\omega}_{r_i,p_j} = \frac{\partial r_i}{\partial p_j} \frac{p_j}{r_i}, \quad M_{ij} = \bar{\omega}_{r_i,c_{Mj}} = \frac{\partial r_i}{\partial c_{Mj}} \frac{c_{Mj}}{r_i}, \\ A_{ij} &= \bar{\omega}_{r_i,c_{Aj}} = \frac{\partial r_i}{\partial c_{Aj}} \frac{c_{Aj}}{r_i} \end{aligned} \quad (8.114)$$

and

$$\Delta p_j = \frac{dp_j}{p_j} \quad \Delta c_i = \frac{dc_i}{c_i}. \quad (8.115)$$

Now let's go back to the request that some of the state variables don't change: $\Delta c_{Ai} = 0$. This leads to a condition for the other state variables:

$$N \text{ diag}(\bar{\omega}) \mathbf{P} \Delta \underline{p} = -N \text{ diag}(\bar{\omega}) \mathbf{M} \Delta \underline{c}_M. \quad (8.116)$$

Here, n equations (matrix N has n rows) are given and it might be the case that a solution cannot be found. In special situations a solution is possible: if the columns of \mathbf{P} can be expressed as linear combinations of \mathbf{M} . Expressed in equations this leads to the following condition that must be checked. Given the nullspace K of $N \text{ diag}(\bar{\omega})$ the rank of two matrices must be equal:

$$\text{Rank}(\mathbf{P}|\mathbf{M}|K) = \text{Rank}(\mathbf{M}|K), \quad (8.117)$$

or, in other words the columns of \mathbf{P} could be expressed by the columns of $(\mathbf{M}|K)$. For many examples, however, kinetic parameters are unknown and the required derivatives cannot be determined quantitatively. Therefore, one tries to find a representation of $(\mathbf{M}|K)$ that is free of parameters; this is called the invariant subspace \mathbf{I} . This can be achieved by standard algebraic manipulations.

For the example from above and taking S as our parameter that is disturbed, the following equations are derived. The stoichiometry and the steady state fluxes — given $r_1 = \bar{r}_1$ and $r_3 = \bar{r}_3$ as input fluxes — are:

$$N = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}; \quad \text{steady state fluxes: } r_2 = \bar{r}_1; \quad r_4 = \bar{r}_3 \quad (8.118)$$

and matrix P reads:

$$\mathbf{P} = \begin{pmatrix} \frac{dr_1}{dS} \frac{S}{r_1} \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \alpha \\ 0 \\ 0 \\ 0 \end{pmatrix}. \quad (8.119)$$

It follows:

$$N \operatorname{diag}(\bar{r}) = \begin{pmatrix} \bar{r}_1 & -\bar{r}_1 & 0 & 0 \\ 0 & 0 & \bar{r}_3 & -\bar{r}_3 \end{pmatrix}. \quad (8.120)$$

Having two independent rows and four columns, the rank is 2 and the dimension of the null space is also 2. K is therefore:

$$K = \begin{pmatrix} \frac{1}{\bar{r}_1} & 0 \\ \frac{1}{\bar{r}_1} & 0 \\ 0 & \frac{1}{\bar{r}_3} \\ 0 & \frac{1}{\bar{r}_3} \end{pmatrix} = \begin{pmatrix} \rho_1 & 0 \\ \rho_1 & 0 \\ 0 & \rho_2 \\ 0 & \rho_2 \end{pmatrix} \quad (8.121)$$

and matrix M is:

$$\mathbf{M} = \begin{pmatrix} 0 \\ \frac{dr_2}{dM} \frac{M}{r_2} \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \beta \\ 0 \\ 0 \end{pmatrix}, \quad (8.122)$$

since in the example only reaction rate r_2 depends on M . Combining the two equations we get for $(\mathbf{M}|K)$:

$$(\mathbf{M}|K) = \begin{pmatrix} 0 & \rho_1 & 0 \\ \beta & \rho_1 & 0 \\ 0 & 0 & \rho_2 \\ 0 & 0 & \rho_2 \end{pmatrix}. \quad (8.123)$$

This matrix is now modified to make it independent from parameters. Since in this example, all columns share a common factor, the columns are divided

by this factor:

$$(\mathbf{M}|K) = \mathbf{I} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}. \quad (8.124)$$

Now the rank of $(\mathbf{P}|\mathbf{I})$ and \mathbf{I} is compared and it can be seen that the rank is three in both cases:

$$\text{Rank}(\mathbf{P}|\mathbf{I}) = \text{Rank} \begin{pmatrix} \alpha & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \text{Rank}(\mathbf{I}). \quad (8.125)$$

Now a different kinetic scheme for the same example is considered: It is assumed now that also r_4 depends on the M . In this case \mathbf{M} is given by:

$$\mathbf{M} = \begin{pmatrix} 0 \\ \frac{dr_2}{dM} \frac{M}{r_2} \\ 0 \\ \frac{dr_4}{dM} \frac{M}{r_4} \end{pmatrix} = \begin{pmatrix} 0 \\ \beta_1 \\ 0 \\ \beta_2 \end{pmatrix}, \quad (8.126)$$

and $(\mathbf{M}|K)$ is:

$$(\mathbf{M}|K) = \begin{pmatrix} 0 & \rho_1 & 0 \\ \beta_1 & \rho_1 & 0 \\ 0 & 0 & \rho_2 \\ \beta_2 & 0 & \rho_2 \end{pmatrix} \rightarrow (\mathbf{M}|K) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ \beta_2/\beta_1 & 0 & 1 \end{pmatrix}. \quad (8.127)$$

So, the parameter independent subspace is given by:

$$\mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, \quad (8.128)$$

and it could be seen immediately that $\text{Rank } (\mathbf{P}|\mathbf{I}_2)$ and $\text{Rank } \mathbf{I}_2$ are different. Even in the case $\beta_2 = \beta_1$ the situation is the same. Here

$$\mathbf{I}_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \quad (8.129)$$

with $\text{Rank } \mathbf{I}_3 = 3$ but $\text{Rank } (\mathbf{P}|\mathbf{I}_3) = 4$.

The method from Steuer et al. is based mainly on algebraic modifications and rank calculations. This can be done today also in a symbolic way in different software tools. Therefore, the method can also be applied to larger networks and is well suited to detect variables that show ACR in these networks.

8.4 Metabolic Control Analysis

Even if the term Systems Biology was coined at the beginning of the 21st century, the first fundamental approaches addressing its key questions can be tracked far back into the 20th century. Biochemical Systems Theory (BST) [4] in 1969, and Metabolic Control Analysis (MCA) [5] four years later, proposed frameworks to analyze biological problems from a systems perspective. While the former approach established a general modeling framework, the latter focused exclusively on sensitivity analysis. This section will present sensitivity analysis as considered in MCA and the next will introduce additional concepts within BST. Starting from the question of how the variations of enzyme amounts affect the stationary flow distribution, a series of the so-called control coefficients are defined and the relationship between those coefficients is determined from the theory.

The starting equation for the analysis is therefore Equation (3.85). The theory presents a closed system of equations. Not only the state variables are to be taken into consideration, but also the stationery flux distribution. In general, one speaks of system variables. In analogy to the earlier method, the control coefficients are also considered in the non normalized and in normalized form.

8.4.1 Control coefficients

The first coefficient that is considered in the original publications is the flux control coefficient C_{Ei}^J . The index refers to the influence of enzyme levels Ei on the stationary flux J through a pathway.

$$C_{Ei}^J = \frac{\Delta J}{\Delta Ei}. \quad (8.130)$$

Normally one lets the delta be small and following a normalization this gives:

$$C_{Ei}^J = \frac{dJ}{dEi} \frac{Ei}{J}. \quad (8.131)$$

As the enzyme levels are not the only parameters in a reaction kinetic expression that exerts influence on the stationary fluxes, one can generalize the approach using the rate r_i that is dependent on various factors p_k like the turn over number k , K_M values or values of metabolites as effectors that can be considered as constants. This leads to:

$$C_{ri}^J = \frac{dJ/dp_k}{dr_i/dp_k} \frac{r_i}{J} = \frac{dJ}{dr_i} \frac{r_i}{J}, \quad (8.132)$$

where one assumes that the parameter p_k has a direct influence on r_i . The second important coefficient is the concentration control coefficient which is

defined as following in the analogy to the previous definition:

$$C_{ri}^{Sj} = \frac{dS_j/dp_k}{dr_i/dp_k} \frac{r_i}{S_j} = \frac{dS_j}{dr_i} \frac{r_i}{S_j}. \quad (8.133)$$

It describes the influence of the chosen parameters on the concentration of the metabolites Sj (S is simplified like earlier and is equivalent to c_S).

Starting from Equation (3.85) equations for the control coefficients could be derived. As a pre-condition, no conservation relations are valid; that is, the rows of matrix N are all independent. We restrict our analysis to this case; however, for the more general case the equations below can easily be adapted.

$$N \underline{r}(\underline{c}, \underline{p}) = \underline{0} \quad (8.134)$$

that is dependent now on the vector of the metabolites \underline{c} and the vector of the parameters \underline{p} . Differentiation with respect to the vector of the parameters leads to [6]:

$$N \frac{dr}{dc} \frac{dc}{dp} + N \frac{dr}{dp} = \underline{0}. \quad (8.135)$$

The equation can now be rearranged and one obtains for the dependence of the stationary metabolite concentration on the parameters with matrix \mathbf{E} from above:

$$\frac{dc}{dp} = - (N \mathbf{E})^{-1} N \frac{dr}{dp}. \quad (8.136)$$

The following correlation is now applicable for stationary fluxes dependent on the stationary metabolite values and parameters:

$$\underline{J} = \underline{r}(\underline{c}(p), \underline{p}). \quad (8.137)$$

Differentiating with respect to the parameters leads to:

$$\frac{d\underline{J}}{dp} = \frac{dr}{dp} + \frac{dr}{dc} \frac{dc}{dp}, \quad (8.138)$$

which when combined with the equation above gives:

$$\frac{d\underline{J}}{dp} = \left(I - \mathbf{E} (N \mathbf{E})^{-1} N \right) \frac{dr}{dp}. \quad (8.139)$$

The matrices can now be determined from both equations that represent the correlation between the perturbation of the parameters and the system response:

Matrix \mathbf{C}^S ($n \times q$) of concentration control coefficients:

$$\boxed{\mathbf{C}^S = - (N \mathbf{E})^{-1} N} \quad (8.140)$$

and matrix \mathbf{C}^J ($q \times q$) of flux control coefficients:

$$\boxed{\mathbf{C}^J = I - \mathbf{E} (\mathbf{N} \mathbf{E})^{-1} \mathbf{N}} \quad (8.141)$$

where the last equation can be expressed as:

$$\mathbf{C}^J = I + \mathbf{E} \mathbf{C}^S. \quad (8.142)$$

In the previous section, examples were introduced regarding differentiation with respect to the parameters. One obtains the following relationships for the differentiation of the rate with respect to the concentration, also called elasticity, for the Michaelis-Menten kinetics:

EXAMPLE 8.10 Elasticity of the Michaelis-Menten kinetics.

Starting from:

$$r = r_{max} \frac{S}{K + S}$$

differentiation leads to:

$$\epsilon_{rS} = r_{max} \frac{K}{(K + S)^2} \quad (8.143)$$

and after normalization:

$$\epsilon_{rS} \frac{S}{r} = \frac{K}{(K + S)}. \quad (8.144)$$

The value of the normalized elasticity always lies between zero and one. For large substrate concentration values in the range of saturation, any variation of the values has little effect on the rate and ϵ approaches zero.

8.4.2 Summation and connectivity theorems for linear pathways

For the individual elements from the previously introduced matrices, one can now derive theorems. These theorems are related to the rows/columns of the matrices and are known as summation and connectivity theorems. Let q represent the number of reactions that are connected together in series (q reactions lead to $m = q - 1$ metabolites in the pathway). The summation theorems read:

$$\boxed{\text{i)} \quad \sum_{k=1}^q C_{rk}^S = 0 \quad \text{ii)} \quad \sum_{k=1}^q C_{rk}^J = 1} \quad (8.145)$$

and the connectivity theorems read:

$$\boxed{\text{iii)} \quad \sum_{k=1}^q C_{r_k}^S \epsilon_{r_k,S} = -1 \quad \text{iv)} \quad \sum_{k=1}^q C_{r_k}^J \epsilon_{r_k,S} = 0.} \quad (8.146)$$

For iii) the sum of the concentration control coefficients is multiplied with the apparent elasticities; otherwise the sum is zero. The summation theorem ii) can be illustrated as following: Consider a linear reaction path with a specified number of enzymes that inter-convert the metabolites. Now, one increases the concentration of the enzymes simultaneously by a small value λ , which in turn increases the stationary flux through the reaction path by a similar factor λ . The metabolites remain the same if steady states are considered. One can now calculate the variation of the stationary fluxes as an alternative to the summation of the variations of all enzyme amounts. Variations will be summed up where all the other enzyme amounts will be considered as constants. This gives:

$$\frac{dJ}{J} = \sum \frac{1}{J} \left(\frac{dJ}{dEi} \right)_{E_j=const} dEi. \quad (8.147)$$

Expanding to obtain:

$$\frac{dJ}{J} = \sum \frac{Ei}{J} \left(\frac{dJ}{dEi} \right)_{E_j=const} \frac{dEi}{Ei} = \sum C_{Ei}^J \frac{dEi}{Ei}. \quad (8.148)$$

As the left side is equal to λ and the same is true for the second factor in the sum, all the flux control coefficients sum up to 1.

The theorems can be summarized now using a single matrix equation. With q being the number of reactions and $m = q - 1$ being the number of metabolites in the linear reaction paths:

$$\boxed{\begin{pmatrix} I_{1 \times q} \\ \mathbf{E}^T_{m \times q} \end{pmatrix} (\mathbf{C}^J_{q \times 1} - \mathbf{C}^S_{q \times m}) = I_{q \times q}.} \quad (8.149)$$

The equation denotes that with the knowledge of only the local characteristics of the reactions, that is, the elasticities, the global characteristics in form of control coefficients can be determined. This is accomplished by an equation in which the matrix of elasticities is inverted:

$$(\mathbf{C}^J_{q \times 1} - \mathbf{C}^S_{q \times m}) = \begin{pmatrix} I_{1 \times q} \\ \mathbf{E}^T_{m \times q} \end{pmatrix}^{-1} I_{q \times q}. \quad (8.150)$$

EXAMPLE 8.11 Reaction path with two reactions and one metabolite.

In the represented metabolic pathway, the metabolite S is formed by the reversible reaction r_1 and is consumed by the reversible reaction r_2 . The quantities X_1 and X_2 shall be treated as constants:



Using equation (8.149), all theorems can be immediately written down:

$$\begin{pmatrix} 1 & 1 \\ \epsilon_{r_1 S} & \epsilon_{r_2 S} \end{pmatrix} \begin{pmatrix} C_{r_1}^J & -C_{r_1}^S \\ C_{r_2}^J & -C_{r_2}^S \end{pmatrix} = I_{2 \times 2}. \quad (8.152)$$

The two flux control coefficients are calculated to be:

$$C_{r_1}^J = \frac{-\epsilon_{r_2 S}}{\epsilon_{r_1 S} - \epsilon_{r_2 S}} \quad \text{and} \quad C_{r_2}^J = \frac{\epsilon_{r_1 S}}{\epsilon_{r_1 S} - \epsilon_{r_2 S}}. \quad (8.153)$$

If, for example, now one lets the reaction r_1 be irreversible and independent of S , the corresponding elasticity is then zero and the control is totally on the first enzyme, which represents a bottleneck. The two concentration coefficients are determined to be:

$$C_{r_1}^S = \frac{-1}{\epsilon_{r_1 S} - \epsilon_{r_2 S}} \quad \text{and} \quad C_{r_2}^S = \frac{1}{\epsilon_{r_1 S} - \epsilon_{r_2 S}}. \quad (8.154)$$

For the special case that the first enzyme is irreversible we get $C_{r_1}^S = 1/\epsilon_{r_2 S}$ and $C_{r_2}^S = -1/\epsilon_{r_2 S}$. The concentration of S will increase if the first rate is increased and it is decreased if the parameter in the second rate is increased.

8.4.3 General summation and connectivity theorems

As for general networks, branching occurs and the stationary fluxes vary in the individual branches; this must be taken into account in the calculation. For networks that have no more conservation relations, this leads to the following central equation:

$$\boxed{\begin{pmatrix} \mathbf{C}^J \\ \mathbf{C}^S \end{pmatrix} (K \quad \mathbf{E}) = \begin{pmatrix} K & \mathbf{0} \\ \mathbf{0} & -I \end{pmatrix}} \quad (8.155)$$

where K is the null space of N .

EXAMPLE 8.12 Branched reaction path with three reactions and one metabolite.

The considered network looks like the following:



The stoichiometric matrix of the network is $N = [1 \ -1 \ -1]$. One then obtains the null space K :

$$K = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (8.157)$$

Matrix \mathbf{E} in this case is a row vector with three entries:

$$\mathbf{E} = \begin{pmatrix} \frac{dr_1}{dS} \\ \frac{dr_2}{dS} \\ \frac{dr_3}{dS} \end{pmatrix} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix}, \quad (8.158)$$

and the complete structure from Equation 8.155 reads:

$$\begin{pmatrix} C_{11}^J & C_{12}^J & C_{13}^J \\ C_{21}^J & C_{22}^J & C_{23}^J \\ C_{31}^J & C_{32}^J & C_{33}^J \\ C_1^S & C_2^S & C_3^S \end{pmatrix} \begin{pmatrix} 1 & 1 & \epsilon_1 \\ 1 & 0 & \epsilon_2 \\ 0 & 1 & \epsilon_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \quad (8.159)$$

There are 9 flux control coefficients in total and one gets the following result:

$$\mathbf{C}^J = \frac{1}{\epsilon_1 - \epsilon_2 - \epsilon_3} \begin{pmatrix} -(\epsilon_2 + \epsilon_3) & \epsilon_1 & \epsilon_1 \\ -\epsilon_2 & \epsilon_1 - \epsilon_3 & \epsilon_2 \\ -\epsilon_3 & \epsilon_3 & \epsilon_1 - \epsilon_2 \end{pmatrix}, \quad (8.160)$$

and for the vector of concentration coefficients:

$$\mathbf{C}^S = \frac{1}{\epsilon_2 + \epsilon_3} (1 \ -1 \ -1). \quad (8.161)$$

For the special case $\epsilon_1 = 0$ and $\epsilon_2 = \alpha \epsilon_3$, that is, S has an influence on r_2 and in a different way on r_3 , one gets:

$$\mathbf{C}^J = \frac{1}{\alpha + 1} \begin{pmatrix} \alpha + 1 & 0 & 0 \\ \alpha & 1 & -\alpha \\ 1 & -1 & \alpha \end{pmatrix}; \mathbf{C}^S = \frac{1}{(\alpha + 1)\epsilon_3} (1 \ -1 \ -1). \quad (8.162)$$

Let's put numerical values in. For $r_1 = 1$ and two Michaelis-Menten kinetics:

$$r_2 = k_2 \frac{S}{K_2 + S}; \quad r_3 = k_3 \frac{S}{K_3 + S}. \quad (8.163)$$

The steady state solution for $k_2 = k_3 = 1$ is $S = \sqrt{K_2 K_3}$. Plugging this value in the respective derivative:

$$\epsilon_{21} = \left. \frac{dr_2}{dS} \right|_{S=\sqrt{K_2 K_3}} = \epsilon_{31}. \quad (8.164)$$

Here we found that $\alpha = 1$ and the matrix of control coefficients is:

$$\mathbf{C}^J = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 1/2 & -1/2 \\ 1/2 & -1/2 & 1/2 \end{pmatrix}. \quad (8.165)$$

Irrespective of the actual values of K_2 and K_3 the system responds to changes of all three rates in the same manner: Increasing the input flux r_1 is equally distributed to J_2 and J_3 . The influence of both rates r_2 and r_3 on J_2 is in such a way that a positive change in r_2 increases J_2 in the same ratio that J_3 is decreased and vice versa. For the same input flux and values $k_2 = 2$, $k_3 = 0.1$, $K_2 = 0.1$ and $K_3 = 0.1$ we found $\alpha = 20$ and matrix \mathbf{C}^J therefore read:

$$\mathbf{C}^J = \begin{pmatrix} 1 & 0 & 0 \\ 20/21 & 1/21 & -20/21 \\ 1/21 & -1/21 & 20/21 \end{pmatrix}. \quad (8.166)$$

The second and third row describe the influence of the reaction rates on the steady flux J_2 and J_3 , respectively. For this large value of α the influence of r_1 is stronger on J_2 than on J_3 . In contrast, the influence of r_2 on the steady fluxes J_2 and J_3 is smaller than the influence of r_3 . To increase the flux J_2 one therefore has to modify the input flux. A direct modification of r_2 is not promising.

The representation of the relationship between the elasticities and the control coefficients is not a minimal one since the fluxes are not independent from each other (this is also true in general for metabolites; however we restrict our analysis for systems where the metabolites are independent). For a further analysis, the system can be split into dependent and independent parts; see [7] for details. In this way a very elegant form of the general theorems is obtained.

The application of the metabolic control theory is in the analysis of networks in biotechnology and medicine. In biotechnology, one is interested in how a metabolic pathway can be optimized and which inputs are required in order to increase yield and productivity. The practical application is somewhat difficult, as the elasticities are not known and are difficult to be determined experimentally and, thus, a more qualitative approach is possible in order to vary the flux through selected reaction pathways. In medicine, one is interested in the action of drugs and would like to determine the enzymes that have a large impact on the metabolic pathway, as it represents a potential target for the drug.

8.5 Biochemical Systems Theory

The availability of tools like control coefficients and the other forms of sensitivity analysis discussed above has provided a rigorous approach to study

how biological systems remain robust to perturbations and nevertheless responsive to the right type of signals. The old concept of rate limiting steps has grudgingly conceded terrain to a more quantitative view where control is distributed among different components. But one may wonder: is this enough? Sometimes it may seem that textbooks have just replaced sentences like: “Phosphofructokinase is the rate limiting enzyme of glycolysis” with politically correct versions such as “Phosphofructokinase is the enzyme with the highest Flux Control Coefficient.” But the fact remains that overexpression of one or two enzymes seldom increases the flux through a pathway even if they have high flux control coefficients [11].

There is obviously much more to the regulation of cellular processes than sensitivity analysis. Modeling cellular processes as non linear dynamical systems immediately draws attention to issues like the existence of multiple steady states, stability, oscillations, etc. It is difficult to approach those problems using a systematic framework like that presented in the previous section. Non linear models are incredibly diverse and most of what is known about them comes from *ad hoc* analyses performed in small models with two or three variables, which cannot be generalized to bigger models.

8.5.1 Building a theoretical framework

But difficult as the above mentioned problems may be, they are not unsurmountable. Most biological subsystems exhibit relatively simple behaviors and are known to be very robust against perturbations in the environment and mutations. These are logical consequences of evolution, since biological systems must normally achieve certain goals like biosynthesis or energy production regardless how conditions might change. Moreover, certain patterns can be frequently observed like end-product inhibition of biosynthetic pathways in bacteria. Evolution often seems to result in the repetition of certain patterns or design principles, which is consistent with natural selection being an opportunistic process in which many components are reused in order to achieve certain goals.

Upon dealing with biological complexity, hard theoretical results that provide a firm ground on which to work are extremely valuable. If it were at all possible to develop a theoretical framework for biological systems, it would certainly look very different from the laws of Newton or the Maxwell equations. The huge diversity of biology is linked to the fact that living beings are the result of evolution, solutions found at a certain time to solve a particular problem. For that reason, modern biology is steadily becoming more similar to engineering and systems science than to physics and chemistry. Just as general methods in systems theory can be applied to many different kinds of control systems, be they electronic circuits, servomechanisms or communication systems, Biochemical Systems Theory enables the development models that fit specific cases but that can be analyzed in a standardized fashion. This com-

promise between specificity and generality can be partly achieved by choosing an appropriate modelling formalism.

Reducing the number of parameters or variables in a model is not the only way to cope with complexity. Systems with a regular structure enable the development of tailor-made techniques and facilitate the study of their properties. Of course the best example of systems with a regular structure are linear systems, but there are also nonlinear formalisms that are sometimes referred to as canonical models. An example of well studied formalism is the Lotka-Volterra (LV) equations, which have been used in ecology for roughly a century. In a LV model, each equation has the form:

$$\dot{X}_i = X_i \left(\lambda_i + \sum_j a_{ij} X_j \right), \quad (8.167)$$

and the information about the model can be compiled in a numerical matrix ($\mathbf{A}|\boldsymbol{\lambda}$). Analyzing it, many properties of the model, like existence, number and stability of steady states, can be established regardless of the size of the system.

In spite of their structural rigidity, LV models exhibit a rich dynamical behavior. Actually, practically any nonlinear system can be rewritten as a Lotka-Volterra by performing some simple variable changes and adding some auxiliary variables. In the following sections we will cover the power-law formalism, which is closely related to LV systems but has a more flexible structure, one more adequate to describe processes such as biochemical reactions. Biological processes like metabolism, gene expression, etc. can be described as linear combinations of fluxes that either produce or consume each component.

$$\dot{\underline{x}} = N \underline{r}(\underline{x}). \quad (8.168)$$

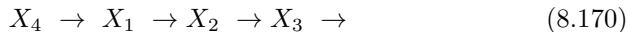
There is no standard mathematical expression to describe reaction rates, as already shown in previous sections. There are many different types of “rate laws,” each with its pros and cons. In the power-law formalism, each rate has the form:

$$r_i = k_i \prod_j X_j^{f_{ij}}, \quad (8.169)$$

where the parameters borrow their names from chemical kinetics: k_i are rate constants and f_{ij} are kinetic orders. All these parameters are real numbers. There are two kinds of magnitudes, X_j , that influence the rate in Equation (8.169): dependent variables that will change in time, and independent variables that will be considered constant during simulations. In order to clearly distinguish between them, the dependent variables will be numbered first $j = 1 \dots n$ and independent variables will receive the subsequent numbers $j = n + 1 \dots m$. All the variables can be collected in a vector \underline{x} , and this vector can be partitioned in a vector containing only the dependent variables \underline{x}_d and another containing only the independent variables \underline{x}_i .

EXAMPLE 8.13 *An unbranched pathway.*

The pathway:



has three intermediates or dependent variables: X_1, X_2 and X_3 that are converted sequentially into one another. The only independent variable of this system would be the initial substrate, X_4 , that is assumed constant.

8.5.2 Setting up a power-law model

There are three main ways to obtain a power-law model: Fundamental representation, recasting and approximation. Fundamental representation: Sometimes, cellular processes are modeled using detailed descriptions of their mechanisms, which are described using mass action kinetics. For instance, the process



would have a rate law:

$$r = k A^2 B. \quad (8.172)$$

All mass action models are particular cases of power-laws where the kinetic orders are positive integers matching the molecularity of each compound in the process. That is the reason why power-law models are called Generalized Mass Action (GMA). When reactions occur on dimensionally constrained environments — for example membranes or the crowded cytoplasmic compartment — even their elemental processes can have fractional kinetic orders.

Recasting: Describing processes using elementary mechanisms requires including many variables that are often impractical to measure, like enzyme-substrate complexes. For that reason the use of approximate rate laws such as the Hill equation or Michaelis-Menten is normally preferred. Such models can be rewritten or recast within the power-law formalism by simple changes of variables which normally involve the creation of additional auxiliary variables.

EXAMPLE 8.14 *A simple case of recasting.*

For instance the equation:

$$\dot{S} = \frac{r_{max} S}{S + K} - k_d S, \quad (8.173)$$

can be recast into power law form by defining the auxilliary variable $Z = S + K$, so $\dot{Z} = \dot{S}$. Now the equations are:

$$\dot{S} = r_{max} S Z^{-1} - k_d S \quad (8.174)$$

$$\dot{Z} = r_{max} S Z^{-1} - k_d S, \quad (8.175)$$

which is a power-law model and will behave exactly as the original one as long as the initial values satisfy the condition $Z(0) = S(0) + K$.

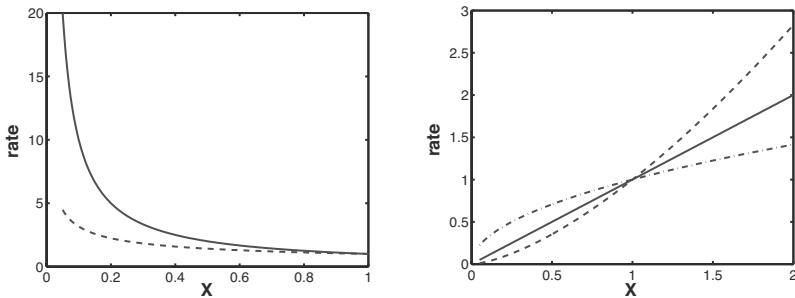


FIGURE 8.9: Power laws $r = k X^f$ with different exponents. Left: Negative exponent. Right: Positive exponent. For values smaller than one, a hyperolic curve can be seen while for values larger than one, a more sigmoid curve is detected.

Approximation: The last way of obtaining a power-law model and the most frequently used is to approximate a non power-law system. Unlike the previously discussed techniques, the obtained system is not identical to the original but it is often much simpler since no additional variables are required. Just as non linear equations can be approximated by linearizing around a reference steady state, each of the rates in a model can be approximated by a power-law. Power-laws are more versatile than linear functions as can be seen in Figure 8.9, and thus tend to provide better approximations than standard linearizations. Since power-laws become straight lines in logarithmic coordinates, approximating any function with a power-law can be done by expanding its Taylor series in log-log coordinates around a reference steady state. For a single variable that would be:

$$\log r = \log r|_0 + \left. \frac{\partial \log r}{\partial \log X} \right|_0 (\log X - \log X|_0), \quad (8.176)$$

where the notation $\cdot|_0$ means that whatever magnitude lies to the left of the bar is computed at the reference steady state and is therefore a number. This equation can be rearranged to obtain:

$$\log r = f \log X + \log k, \quad (8.177)$$

where

$$f_{ij} = \left. \frac{\partial \log V}{\partial \log X} \right|_0 = \left. \frac{\partial V_i}{\partial X_j} \frac{X_j}{V_i} \right|_0 \quad \text{and} \quad k_i = \left. \frac{r_i}{\prod X_j^{f_{ij}}} \right|_0. \quad (8.178)$$

Removing the logarithms from Equation 8.177 results in a power-law.

It can be easily seen that Equation 8.178 is equivalent to the definition of elasticity given in the previous section on MCA. The fact that kinetic

orders — or elasticities — calculated at a certain reference state give a good approximation to the kinetics of the system is partly because biochemical systems tend to maintain metabolite concentrations within narrow margins, thus staying within the region of validity of the Taylor series.

EXAMPLE 8.15 Kinetic order of the Michaelis-Menten kinetics

For instance, in a classical Michaelis-Menten reaction $r = \frac{r_{max}S}{K+S}$ the kinetic order would be

$$f_{ij} = \left. \frac{K}{K+S} \right|_0 , \quad (8.179)$$

just as the final result of Example 8.10. Kinetic orders for Michaelis-Menten kinetics can therefore vary from zero, when the enzyme is saturated to one when the enzyme is very far from saturation. A typical value would be $f_{ij} = 0.5$, which corresponds to substrate concentrations approximately equal to the saturation constant K . In general, as a rule of thumb, kinetic orders for inhibitors range between -1 and 0 while activators range from 0 to 2 .

8.5.3 S-systems

An S-system is a particular case of GMA model that only has two terms per equation: A positive term accounting for production and a negative one accounting for depletion.

$$\frac{dX_i}{dt} = r_i^+ - r_i^- = \alpha_i \prod_j X_j^{g_{ij}} - \beta_i \prod_j X_j^{h_{ij}} . \quad (8.180)$$

S-systems can arise as the natural representation for some systems but can also be obtained by lumping all synthesis or degradation fluxes together in one power-law each. Such an approximation is equivalent to the Taylor series explained above. S-systems have many desirable properties that facilitate their analysis. In subsequent sections we will consistently use the example of the unbranched metabolic network to explain the most common forms of analysis:

EXAMPLE 8.16 S-system equations for an unbranched pathway.

The unbranched pathway described in Example 8.13 could be described by the following S-system:

$$\begin{aligned} \dot{X}_1 &= \alpha_1 X_4^{g_{14}} - \beta_1 X_1^{h_{11}} \\ \dot{X}_2 &= \alpha_2 X_1^{g_{21}} - \beta_2 X_2^{h_{22}} \\ \dot{X}_3 &= \alpha_3 X_2^{g_{32}} - \beta_3 X_3^{h_{33}} , \end{aligned} \quad (8.181)$$

where all the parameters are positive. The exact kinetics of the system can be of many different types but as has been shown, the power-law rate laws will be

a good approximation around the steady state. By keeping the kinetic orders in symbolic form, a generic steady state can be analyzed without specific assumptions on kinetics or parameter values.

The number of unknown parameters can be further reduced because due to mass conservation, the degradation term of each variable is equal to the production term of the next:

$$\begin{aligned}\dot{X}_1 &= \alpha_1 X_4^{g_{14}} - \beta_1 X_1^{h_{11}} \\ \dot{X}_2 &= \beta_1 X_1^{h_{11}} - \beta_2 X_2^{h_{22}} \\ \dot{X}_3 &= \beta_2 X_2^{h_{22}} - \beta_3 X_3^{h_{33}}.\end{aligned}\tag{8.182}$$

Steady states

An important consequence of the structure of S-systems is that in spite of being fully nonlinear systems, their steady states are the solutions of a linear equation. In the steady state, the two terms of each equation must be equal, so taking logarithms and rearranging:

$$\sum_j (g_{ij} - h_{ij}) \log X_j = \log \frac{\beta_i}{\alpha_i}.\tag{8.183}$$

For convenience we can define $y_i = \log X_i$, and $b_i = \log \frac{\beta_i}{\alpha_i}$. We can also collect the kinetic orders in a matrix \mathbf{A} so that $a_{ij} = g_{ij} - h_{ij}$. The coefficients of dependent and independent variables can be partitioned in two matrices, $\mathbf{A_d}$ and $\mathbf{A_i}$, respectively. Then

$$\mathbf{A_d} \underline{y}_d + \mathbf{A_i} \underline{y}_i + \underline{b} = 0,\tag{8.184}$$

and simple linear algebra can be used to solve \underline{y}_d as a function of the independent variables and the parameters of the system. Also, the number of possible steady states is as in linear systems. There will be a unique steady state if $\det(\mathbf{A_d}) \neq 0$ or there can be infinite steady states otherwise but not two or three.

EXAMPLE 8.17 Steady states of an unbranched pathway.

In our particular example:

$$\begin{aligned}g_{10} y_4 - h_{11} y_1 &= b_1 \\ h_{11} y_1 - h_{22} y_2 &= b_2 \\ h_{22} y_2 - h_{33} y_3 &= b_3,\end{aligned}\tag{8.185}$$

solving for the dependent variables:

$$\begin{aligned} y_1 &= \frac{g_{14} y_4 - b_1}{h_{11}} \\ y_2 &= \frac{g_{14} y_4 - b_1 - b_2}{h_{11}} \\ y_3 &= \frac{g_{14} y_4 - b_1 - b_2 - b_3}{h_{22}}. \end{aligned} \quad (8.186)$$

Sensitivity analysis

Sensitivity analysis of S-systems is straightforward. Equation (8.184) can be rewritten as:

$$\underline{y}_d = \mathbf{L}(\underline{x}_d, \underline{x}_i) \underline{y}_i + \mathbf{S}(\underline{x}_d, \underline{\alpha}) \underline{b}, \quad (8.187)$$

where

$$\mathbf{L}(\underline{x}_d, \underline{x}_i) = -\mathbf{A}_d^{-1} \mathbf{A}_i \quad \text{and} \quad \mathbf{S}(\underline{x}_d, \underline{\alpha}) = -\mathbf{A}_d^{-1}. \quad (8.188)$$

It can be seen by mere inspection that $\mathbf{L} = \frac{\partial \log \underline{x}_d}{\partial \log \underline{x}_i}$ and $\mathbf{S} = \frac{\partial \log \underline{x}_d}{\partial \log \underline{\alpha}}$. These matrices are called logarithmic gains — or simply log-gains — and parameter sensitivities, respectively. Just as elasticities can be translated as kinetic orders, control coefficients can be translated as log-gains. All the summation and connectivity theorems of MCA can be derived from the steady state equations and there is extensive literature on exploring and discussing such equivalence. These theorems, however, play a negligible role in BST, where the focus lies on dynamic systems and not in relations between sensitivities. As a rule of thumb, we will interpret sensitivities as the change experimented by a variable or flux in steady state when a parameter or independent variable increases by 1%. Thus, positive (negative) values indicate positive (negative) correlations. Absolute values greater than one indicate amplification of the initial perturbation and absolute values less than one indicate a dampening.

EXAMPLE 8.18 Substrate accumulation in unbranched pathways.

The logarithmic gains for the unbranched pathway are:

$$\frac{\partial y_1}{\partial y_4} = \frac{g_{14}}{h_{11}} > 0, \quad \frac{\partial y_2}{\partial y_4} = \frac{g_{14}}{h_{22}} > 0, \quad \frac{\partial y_3}{\partial y_4} = \frac{g_{14}}{h_{33}} > 0. \quad (8.189)$$

Since they are all positive, any increase in the initial substrate will not only result in an increased flux through the pathway but also in an accumulation of all the intermediates.

Normalization

It is often convenient to normalize/scale S-systems using their steady state values, since this results in a reduction of parameters and brings the equations

to a particularly convenient form. Defining the normalized variables

$$u_i = \frac{X_i}{X_i|_0}, \quad (8.190)$$

the equations become:

$$\begin{aligned} \dot{u}_i &= \frac{1}{X_i|_0} \left(\alpha_i \prod_j (X_j|_0 u_j)^{g_{ij}} - \beta_i \prod_j (X_j|_0 u_j)^{h_{ij}} \right) \\ &= \frac{1}{X_i|_0} \left(\alpha_i \prod_j X_j|_0^{g_{ij}} \prod_j u_j^{g_{ij}} - \beta_i \prod_j X_j|_0^{h_{ij}} \prod_j u_j^{h_{ij}} \right). \end{aligned} \quad (8.191)$$

But in steady state both fluxes are equal:

$$r_i|_0 = \alpha_i \prod_j X_j|_0^{g_{ij}} = \beta_i \prod_j X_j|_0^{h_{ij}}. \quad (8.192)$$

So, the normalized system can be simplified to:

$$\dot{u}_i = \frac{V_i}{X_i|_0} \left(\prod_j u_j^{g_{ij}} - \prod_j u_j^{h_{ij}} \right), \quad (8.193)$$

which always has its steady state at $u_i = 1 \forall i$. The constant factor multiplying the equation, $F_i = \frac{V_i}{X_i|_0}$, has inverse time units and provides an estimate of the time scale of the variable. Big multipliers result in fast variables and small multipliers correspond to somewhat “sluggish” dynamics. The inverse of F_i , mass over flux, corresponds to the definition of residence times in a compartment and also to that of transition times in biochemistry. It is also noteworthy that all the rate constants and independent variables are factored into F .

EXAMPLE 8.19 Scaled unbranched pathway.

In the unbranched pathway example the normalized system would look like this:

$$\dot{u}_1 = F_1 \left(1 - u_1^{h_{11}} \right), \quad \dot{u}_2 = F_2 \left(u_1^{h_{11}} - u_2^{h_{22}} \right), \quad \dot{u}_3 = F_3 \left(u_2^{h_{22}} - u_3^{h_{33}} \right) \quad (8.194)$$

where $F_i = \beta_i x_i^{h_{ii}-1}$.

Stability

The stability of an S-system can be analyzed in the normalized system. Linearizing gives:

$$\dot{\underline{u}} = J \underline{u} \quad (8.195)$$

where

$$J = \text{diag}(F_1 \dots F_n)(G - H). \quad (8.196)$$

Expanding the equation $\det(J) = 0$ yields the characteristic polynomial of the system, which has the form:

$$\lambda^n + p_1 \lambda^{n-1} + p_2 \lambda^{n-2} \dots p_n \stackrel{!}{=} 0. \quad (8.197)$$

The steady state will be stable if the characteristic polynomial is Hurwitz — in other words, if the real part of all its roots are negative. There are methods to find out necessary conditions to guarantee that a polynomial is Hurwitz, even without solving it. These methods result in a series of relations between the coefficients that, if fulfilled, guarantee the stability of the system.

EXAMPLE 8.20 Stability of an unbranched pathway.

The linearized system would be

$$\begin{aligned} \dot{u}_1 &= -F_1 h_{11} u_1 \\ \dot{u}_2 &= F_2 (h_{11} u_1 - h_{22} u_2) \\ \dot{u}_3 &= F_3 (h_{22} u_2 - h_{33} u_3). \end{aligned} \quad (8.198)$$

leading to the Jacobian:

$$\begin{vmatrix} -(F_1 h_{11} + \lambda) & 0 & 0 \\ F_2 h_{11} & -(F_2 h_{22} + \lambda) & 0 \\ 0 & F_3 h_{22} & -(F_3 h_{33} + \lambda) \end{vmatrix} = 0. \quad (8.199)$$

The third degree characteristic polynomial has coefficients:

$$\begin{aligned} p_1 &= F_1 h_{11} + F_2 h_{22} + F_3 h_{33} \\ p_2 &= F_1 F_2 h_{11} h_{22} + F_1 F_3 h_{11} h_{33} + F_2 F_3 h_{22} h_{33} \\ p_3 &= F_1 F_2 F_3 h_{11} h_{22} h_{33}. \end{aligned} \quad (8.200)$$

Stability will be guaranteed when $p_1 > 0$, $p_3 > 0$ and $p_1 p_2 > p_3$, leading to the equation:

$$\left(2 + \frac{F_1 h_{11}}{F_2 h_{22}} + \frac{F_2 h_{22}}{F_1 h_{11}} + \frac{F_1 h_{11}}{F_3 h_{33}} + \frac{F_3 h_{33}}{F_1 h_{11}} + \frac{F_2 h_{22}}{F_3 h_{33}} + \frac{F_3 h_{33}}{F_2 h_{22}}\right) h_{33} > 0. \quad (8.201)$$

Since all the parameters are positive, the pathway will always be stable, disregarding the values of the kinetic parameters.

8.5.4 Design principles and end-product inhibition

Now that the basic tools have been defined lets use them in order to analyze a simple case. The unbranched pathway analyzed in the examples above shows a relatively simple behavior. Increasing the available substrate, the flux of the pathway will increase but so will all the intermediates. Since the cytoplasm of a cell is extremely crowded with all kinds of compounds and macromolecules, unnecessary accumulation of intermediates is not a desirable feature. How can this accumulation be avoided? It has been observed that in many biosynthetic pathways the end-product of the pathway inhibits the first reaction. Modifying the model to include such an inhibition amounts to change the first equation to:

$$\dot{X}_1 = \alpha_1 X_4^{g_{14}} X_3^{g_{13}} - \beta_1 X_1^{h_{11}}. \quad (8.202)$$

The new kinetic order g_{13} must be negative, since it accounts for an inhibition. Now we can compare the old model and the new version to observe the effects of end-product inhibition. For instance, the steady state can be calculated from the modified equations

$$b_1 = g_{10} y_0 - h_{11} y_1 + g_{13} y_3, \quad b_2 = h_{11} y_1 - h_{22} y_2, \quad b_3 = h_{22} y_2 - h_{33} y_3, \quad (8.203)$$

and the logarithmic gains become:

$$\begin{aligned} \frac{\partial y_1}{\partial y_0} &= \frac{h_{33}}{h_{33} - g_{13}} \frac{g_{10}}{h_{11}} > 0, & \frac{\partial y_2}{\partial y_0} &= \frac{h_{33}}{h_{33} - g_{13}} \frac{g_{10}}{h_{22}} > 0, \\ \frac{\partial y_3}{\partial y_0} &= \frac{h_{33}}{h_{33} - g_{13}} \frac{g_{10}}{h_{33}} > 0. \end{aligned} \quad (8.204)$$

The new logarithmic gains are like those of the original system, but they are all multiplied by the same factor $\frac{h_{33}}{h_{33} - g_{13}}$ which is always less than one, making the new log-gains smaller than those of the original system. Therefore, the inclusion of a feedback decreases the accumulation of intermediates proportionally to the strength of the feedback inhibition — the absolute value of g_{13} . Is there a limit to the strength of the feedback that can be imposed on the pathway? In order to answer this question, the stability conditions can be updated to include the feedback loop. The new linearized system is:

$$\begin{aligned} \dot{u}_1 &= F_1 (g_{13} u_3 - h_{11} u_1) \\ \dot{u}_2 &= F_2 (h_{11} u_1 - h_{22} u_2) \\ \dot{u}_3 &= F_3 (h_{22} u_2 - h_{33} u_3), \end{aligned} \quad (8.205)$$

and Equation (8.201) becomes:

$$\left(2 + \frac{F_1 h_{11}}{F_2 h_{22}} + \frac{F_2 h_{22}}{F_1 h_{11}} + \frac{F_1 h_{11}}{F_3 h_{33}} + \frac{F_3 h_{33}}{F_1 h_{11}} + \frac{F_2 h_{22}}{F_3 h_{33}} + \frac{F_3 h_{33}}{F_2 h_{22}} \right) h_{33} > -g_{13}, \quad (8.206)$$

which establishes an upper limit to the strength of the feedback loop. This result is a very important complement to the sensitivity analysis. Even if the system becomes more robust as the strength of the feedback increases, there is a point where the stability is lost, which would make robustness meaningless. Inspecting the left hand side of the stability condition we can see that the terms include all possible fractions that can be built using the terms $F_1 h_{11}$, $F_2 h_{22}$ and $F_3 h_3$ dividing one another. Such a sum has a minimum value when $F_1 h_{11} = F_2 h_{22} = F_3 h_3$, so the stability region will be bigger as the difference between the kinetic parameters of the different enzymes grows wider. A similar analysis to the one presented here has shown that the length of the pathway has a negative effect on stability. In those cases, the dispersion of kinetic parameters is a critical factor to obtain viable pathways.

This simple analysis has provided some insight on the observed architecture of metabolic pathways. The lessons extracted from it are specially valuable as engineering of pathways *de novo* has become a reality.

8.5.5 Design space analysis

Many biological systems are expected to have different modes of operation, and each may respond differently to the same stimulus. The response of the TCA cycle to an increase in pyruvate, for instance, is not the same under aerobic and anaerobic conditions. Furthermore, changes in the sensitivities of the model may result in different opportunities for model reduction. Design Space Analysis (DSA) is a relatively recent technique that addresses these issues and will be presented here through an example.

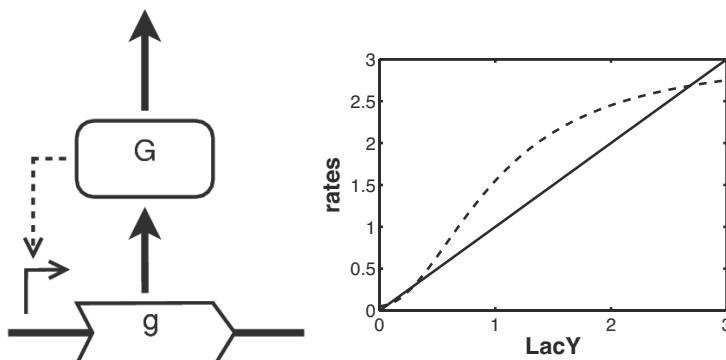


FIGURE 8.10: Left: A gene induces its own expression. Right: Rate of synthesis and degradation. Depending on the values of the parameters, the curves may intersect in one point or in three.

Let's consider a gene that induces its own expression like in the lactose

uptake system (see also previous chapters).

$$\dot{LacY} = \beta + \frac{r_{max} LacY^2}{LacY^2 + K_s^2} - k_d LacY. \quad (8.207)$$

The first two terms are constitutive and inducible expressions while the third term is degradation/dilution. Depending on the values of the parameters, the gene can have a unique induction level in the steady state or three different levels, two of them stable and one of them unstable that acts as a threshold between them. The existence of more than a stable state is extremely important because it can result in the differentiation of a genetically homogeneous population into two completely different expression profiles.

The first step to study this model will be reducing the number of parameters as much as possible. Defining nondimensional variables:

$$x = \frac{LacY}{K_s}, \quad \tau = k_d t \quad \rightarrow \quad \frac{dLacY}{dt} = K_s k_d \frac{dx}{d\tau} \quad (8.208)$$

results in the system:

$$\dot{x} = b + a \frac{x^2}{x^2 + 1} - x \quad \text{with} \quad a = \frac{r_{max}}{k_d K_s}, \quad b = \frac{\beta}{k_d K_s}. \quad (8.209)$$

Since the system has only two nondimensional parameters and they are genetically determined, each possible phenotype can be represented as a point in the plane with the parameters being the coordinates. This plane will be called Design Space because the position of a phenotype in it will determine its performance. The basic idea is to partition the parameter space into discrete regions that are qualitatively similar. The next step of the analysis will be to bring the system into power-law form and that will be done through recasting. Defining an auxiliary variable z we obtain a system of algebraic differential equations in power-law form:

$$\begin{aligned} \dot{x} &= b + a \frac{x^2}{z} - x \\ 0 &= z - x^2 - 1. \end{aligned} \quad (8.210)$$

Now we can greatly simplify the model for such cases in which some terms are much bigger than the rest, so they dominate their respective equations. Particularly, when there is a single positive and a single negative term that dominate in each equation, the model simplifies to an S-system. In order to calculate how many of these particular cases can be found for a given system, we can build a signature containing two numbers for each equation: number of positive terms and number of negative terms. In our particular example, the system signature would be [2 1 1 2], since there are two positive terms in the first equation and two negative terms in the last. The total number of combinations, or cases, is simply the product of all the entries in the system

TABLE 8.1: A summary of all the possible cases in the example.

case	signature	conditions	steady state	S-system
1	[1 1 1 1]	$a \ll b, b^2 \gg 1$	$x _0 = b, z _0 = b^2$	$\dot{x} = b - x$
2	[2 1 1 1]	$a \gg b, a^2 \gg 1$	$x _0 = a, z _0 = a^2$	$\dot{x} = a - x$
3	[1 1 1 2]	$a b \ll 1, b^2 \ll 1$	$x _0 = b, z _0 = 1$	$\dot{x} = b - x$
4	[2 1 1 2]	$a b \ll 1, a^2 \gg 1$	$x _0 = \frac{1}{a}, z _0 = 1$	$\dot{x} = a x^2 - x$

signature, in this case four. Now we can also assign a signature to each case, so we can identify it. For instance, the case where all the first candidates are dominant would be [1 1 1 1]. The conditions for dominance of the fluxes are:

$$b \gg a \frac{x^2}{z} \quad \text{and} \quad x^2 \gg 1. \quad (8.211)$$

If these conditions are satisfied, the system will be very similar to the S-system that only takes the dominant fluxes into account and the system becomes

$$\dot{x} = b - x \quad \text{and} \quad z = x^2 \quad (8.212)$$

and which has a steady state at:

$$x|_0 = b \quad \text{and} \quad z|_0 = b^2. \quad (8.213)$$

Now we can substitute the steady state variables into the dominant conditions to remove the dependent variables and leave only the parameters.

$$a \ll b \quad \text{and} \quad b^2 \gg 1. \quad (8.214)$$

So we have found well defined a region in the Design Space (a polygon within the plane) where the system behaves like an S-system. We can also remove the auxiliary variable now to obtain the reduced model:

$$\dot{x} = b - x, \quad (8.215)$$

which is a full dynamic system but much simpler. It can also be analyzed with all the tools described previously for S-systems. If parameter values fall well within the region defined by Equation (8.214), the dominant flux will drive the dynamics and the corresponding S-system will be an excellent approximation to the full system. In the neighborhood of a boundary with another case, the system will exhibit properties of both cases. It is even possible that two or more cases overlap and this has interesting consequences that will be seen later.

Table 8.1, which summarizes all cases, shows several interesting features. First of all, case signatures can be read in biological terms. In case 1, for

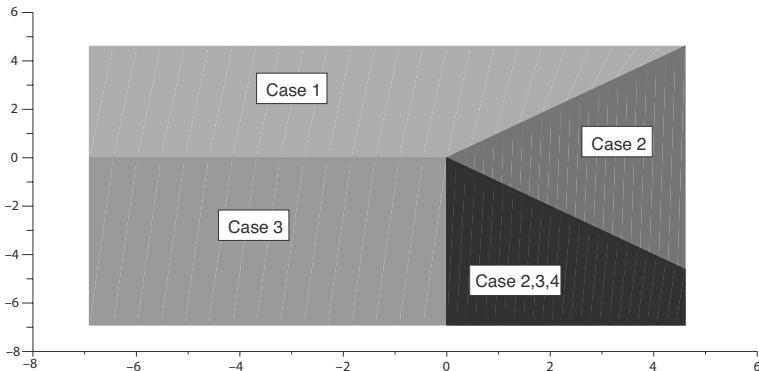


FIGURE 8.11: Design space for parameters a and b . Shown are the different cases. In the black area different cases overlap.

instance, the gene is not induced since the first digit in the case signature indicates that the constitutive expression term dominates over the inducible one. Furthermore, the last digit indicates that induction is saturated, since the dominant term for the second equation results in $z = x^2$, which substituted in the inducible term would cancel its dependency on x . This apparent contradiction of the gene not being induced and being nevertheless saturated could only happen if the constitutive expression level of the gene was greater than the maximum value of inducible expression $\beta > r_{max}$. This unlikely configuration of parameters clearly indicates that it is extremely unlikely to find a gene whose phenotype lies within case 1. It is also easy to see that the dynamic system of case 4 is unstable since it will grow when the variable is above the steady state and decrease when it is below. Like case 1, no phenotype of the system can be expected to behave like case 4, but it still plays a key role in the dynamic of the system, since it overlaps with cases 2 and 3. Actually, case 4 only exists in the overlapping between the other two cases, as shown in Figure 8.11. Such areas where several cases overlap may result in multi-stability, since more than one steady state can become relevant.

In the overlapping area shown in Figure 8.11, all the conditions for cases 2, 3 and 4 are satisfied. By collecting all these conditions and eliminating redundancies we obtain a set of conditions for bistability:

$$ab \ll 1, \quad a \gg 1, \quad b \ll 1. \quad (8.216)$$

Figure 8.12, left plot, shows steady state levels for different values of a . The solid lines are steady states calculated by simulation. In a first round, the steady state was calculated for increasing values of parameter a . A second round of simulations was performed for decreasing values of a which resulted in a classic hysteresis cycle characteristic of a bistable system. The three dotted lines show the steady states corresponding to each stable case and that of the unstable case (in red) that provides the threshold. The left side of the graph

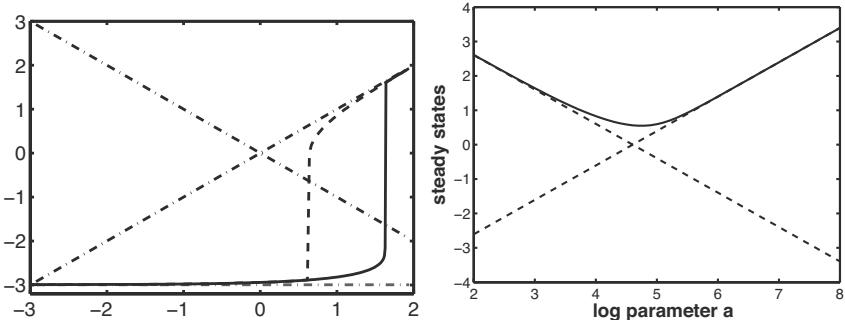


FIGURE 8.12: Left: Hysteresis cycle. Solid and dashed lines are real steady states. Dash-dotted lines are steady states of each case. Right: Real steady states as dots. Steady state of cases in dashed lines.

shows a system with a unique steady state very close to that predicted by case 2. Within the overlapping area, all three cases are valid and the system exhibits bistability, and as the system exits the overlapping area, only a steady state remains that is increasingly close to case 3.

As has already been stated, case 1 is not realistic by itself, but it does play a role by affecting the behavior of the system at the boundaries of other, biological realistic cases. Figure 8.12, right plot, shows the steady state of the system calculated for different values of the parameters. As the parameter values move from case 1 to case 2, the steady states make a smooth transition from one case to the other.

Design Space Analysis also shows promise as a framework for automatic model reduction. The simple example above shows well designated areas of the design space where the model can be reduced to a much simpler form that also happens to be an S-system. In the two stand-alone modes, cases 2 and 3, the whole system reduces to: $\dot{x} = a - x$ and $\dot{x} = b - x$, respectively. But this simplification does not have to be limited to a single case. In the boundary between cases 1 and 2, for example, a partial reduction can be made and the system $\dot{x} = a + b - x$ also provides a very good estimate associated to less strict conditions, namely those common to both cases.

Biochemical systems theory has been growing steadily for the last forty years with the aim to provide a solid theoretical framework for biology. During this time it has established mathematically sound foundations and thoroughly explored some issues like sensitivity analysis, the properties of S-systems, mathematically controlled comparisons between alternative designs of pathways, etc. But the topic and the scope are ambitious enough that new and exciting concepts keep appearing, as is the case of Design Space Analysis. By combining the concepts introduced in this chapter, complex biological systems can be decomposed in alternative modes of operation that are valid in different conditions. These modes of operation have a fixed mathematical structure

(S-system) and can then be very thoroughly analyzed in order to explore its local properties. Moreover, the relative positioning of the different modes to one another enable the study of possible transitions between them, putting together the pieces of the local analysis in a puzzle of possible phenotypes.

Building a theoretical framework for systems biology is such an ambitious goal that not even the most optimistic can claim that we are anywhere near it. Even though it has come a long way, modern Biology is only starting the long journey from describing to explaining. It will be an exciting journey that many scientists will claim to be never-ending. But remember the old motto: we must be realistic and demand the impossible.

8.6 Structured Kinetic Modeling

The analysis of dynamic characteristics of the biochemical network is difficult due to the fact that many kinetic parameters are either known with low precision or not known at all. Therefore, many interesting characteristics like the number of steady states, time constants or stability characteristics can be hardly analyzed quantitatively. Hence, research in recent times has focused on the structured characteristics of systems. An interesting combination of pure structural and parametric analysis was introduced by Steuer [9], which enables studying the stability characteristics of linear systems as independently as possible from the choice of parameters. In this method, the entries of the Jacobian matrix are limited by scaling the model equations as shown earlier, that is, the entries possess well defined intervals which can then be investigated. For instance, considering the dilution during growth as an element of the reaction vector, the basic equation is given by:

$$\dot{\underline{c}} = N \underline{r}(\underline{c}), \quad (8.217)$$

which results from balance equations for intracellular components in cellular networks. The single requirement for the method is the existence of a steady state \underline{c}_0 , which need not be unique. For the steady state, the following applies:

$$0 = N \underline{r}(\underline{c}_0). \quad (8.218)$$

Now, the scaling is performed using the new variables: \underline{x} for the concentration, Λ for the stoichiometry and ρ for the rate vector:

$$x_i := \frac{c_i}{c_{i0}}, \quad \Lambda_{ij} := N_{ij} \frac{r_j(\underline{c}_0)}{c_{i0}}, \quad \rho_j := \frac{r_j(\underline{c})}{r_j(\underline{c}_0)}. \quad (8.219)$$

Inserting the terms in Equation (8.217) results in:

$$\dot{\underline{x}} = \Lambda \rho(\underline{x}). \quad (8.220)$$

The matrix Λ now contains scaled stoichiometric coefficients. Both the values of the steady states of the concentration and the stationary fluxes are inserted. Those values must be known, and can be determined like in the case of flux distributions (see chapter on Metabolic Flux Analysis) and from metabolome measurements. The aim is now a stability analysis, and the entries of the Jacobian matrix of the system at steady state ($x_0 = 1$) can be explored:

$$\boxed{J = \Lambda \left. \frac{d\rho(\underline{x})}{d\underline{x}} \right|_{\underline{x}_0=1} = \Lambda \Theta|_{x_0=1}.} \quad (8.221)$$

In matrix Θ , the derivatives of the scaled kinetic terms with respect to the state variables can be calculated for general reaction kinetics. Next, the Michaelis-Menten kinetics will be used as an example. Using the basic form:

$$r = r_{max} \frac{c}{K + c} \quad (8.222)$$

the scaled terms result in:

$$\rho = \frac{r}{r_0} = \frac{x c_0}{K + x c_0} \frac{K + c_0}{c_0} = \frac{x (K + c_0)}{K + x c_0}. \quad (8.223)$$

Now forming the derivative of ρ with respect to x and inserting $x = 1$:

$$\left. \frac{d\rho}{dx} \right|_{x_0=1} = \Theta|_{x_0=1} = \frac{K}{K + c_0}. \quad (8.224)$$

One observes that for the derivative, independent of its concrete numerical values, it lies always valued between 0 and 1 (this scaling was formally shown previously). For a general kinetics of the form:

$$r = r_{max} \frac{c^n}{f_m(\underline{c}, \underline{p})} \quad (8.225)$$

with a polynomial f_m of m th order in the denominator and parameters \underline{p} , the following applies:

$$\Theta|_{x_0=1} = n - \alpha m \quad (8.226)$$

where α is guaranteed to be between 0 and 1. α depends on the value of the steady state c_0 and the following limits apply:

$$\lim_{c_0 \rightarrow 0} \alpha = 1, \quad \lim_{c_0 \rightarrow \infty} \alpha = 0. \quad (8.227)$$

Using the general form, kinetics that describe inhibition or activation can also be treated since polynomials are also often used in these cases.

EXAMPLE 8.21 Network with two components and three reactions.

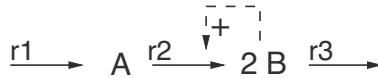


FIGURE 8.13: Network with two components and three reactions.

Figure 8.13 shows the network under consideration. Here, the second component B activates its own synthesis (rate r_2). The following kinetic rate expressions are chosen for $r_2(A, B)$ and $r_3(B)$:

$$\begin{aligned} r_2 &= k_2 \frac{A}{K_2 + A} \frac{B^n}{K_B + B^n} \\ r_3 &= k_3 \frac{B}{K_3 + B}. \end{aligned} \quad (8.228)$$

The differential equations for the original system read:

$$\begin{pmatrix} \dot{A} \\ \dot{B} \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 2 & -1 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2(A, B) \\ r_3(B) \end{pmatrix}. \quad (8.229)$$

For the system, one rate must be predetermined, so that the other two are then uniquely determined. Let the rate r_1 be the given one; the other two are:

$$r_2 = r_1, \quad r_3 = 2r_1. \quad (8.230)$$

For the input flux, the value $r_1 = u$ is selected. Using the previously introduced scaling, the matrix Λ is then:

$$\Lambda = \begin{pmatrix} \frac{u}{A_0} & -\frac{u}{A_0} & 0 \\ 0 & 2\frac{u}{B_0} & -2\frac{u}{B_0} \end{pmatrix}. \quad (8.231)$$

Denoting the scaled rates with ρ_1 , ρ_2 and ρ_3 , and the concentration with a for A and b for B , this yields the matrix with the scaled derivatives:

$$\Theta = \begin{pmatrix} 0 & 0 \\ \Theta_a^{\rho_2} & \Theta_b^{\rho_2} \\ 0 & \Theta_b^{\rho_3} \end{pmatrix} \quad (8.232)$$

with the corresponding derivatives:

$$\Theta_a^{\rho_2} = \frac{d\rho_2}{da}, \quad \Theta_b^{\rho_2} = \frac{d\rho_2}{db}, \quad \Theta_b^{\rho_3} = \frac{d\rho_3}{db}. \quad (8.233)$$

So, there are three terms to consider. For such (rather small) systems the complete parameter space can be analyzed. For large systems, the number of calculations quickly exceeds the numerical capacities. To overcome this, one takes for every entry in the matrix a normal distribution with values between 0 and those given in Equation (8.226) and then values are drawn from those distributions for all parameters' values. If a large number of runs is made, one can assume that the parameter space is well formed without the need to calculate all combinations.

Consider a two dimensional problem, and let the eigenvalues be split into stable (both eigenvalues are small or equal to zero) and unstable (at least one eigenvalue larger than zero). The interesting case is the one where oscillations occur. Here, the imaginary part of the eigenvalues is not equal to zero. Permanent oscillations are obtained when both real parts of the eigenvalues are zero and only imaginary eigenvalues are present (see the classification for two dimensional systems in the Appendix). Figure 8.14 shows an exemplary result of the calculations for the parameter space $\Theta_a^{\rho_2}$, $\Theta_b^{\rho_2}$, when the parameter $\Theta_b^{\rho_3} = 1$.

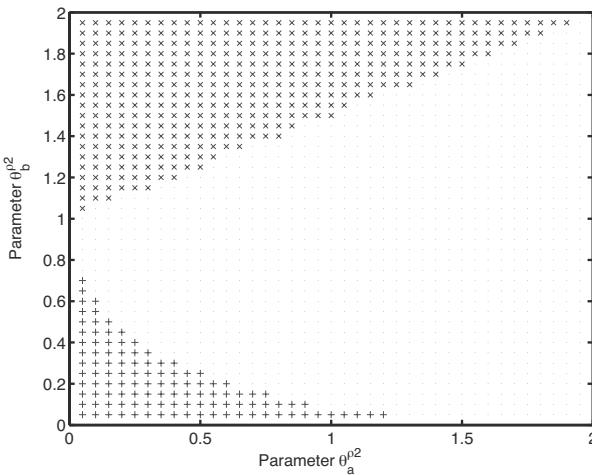


FIGURE 8.14: For the previous network, the parameters are varied between 0 and 2. Crosses indicate the unstable region while the plus signs show the region where stable nodes can be found. Points mark the parameters' range of a stable focus.

If an interesting range is found, one would be also interested in obtaining parameters of the original system via the simulation study. The following relationships can be utilized to (re-)calculate parameters: Start from a general approach that the reaction rate r is of the form:

$$r = k \frac{c_1^n}{K + c_1^n} f(c_2) \quad (8.234)$$

where the term f refers to the activation/inhibition of a second metabolite c_2 , and has the form:

$$\text{Activation} \quad f_A = \frac{c_2^n}{K_A + c_2^n} \quad (8.235)$$

$$\text{Inhibition} \quad f_I = \frac{K_I}{K_I + c_2^n}. \quad (8.236)$$

The corresponding Θ entries depend only on the stationary values and the half-saturation parameters. The equations can then be simply rearranged with respect to K_A and K_I yielding:

$$\text{Activation} \quad \Theta = \frac{n}{1 + c_{20}^n / K_A} \rightarrow K_A = \frac{\Theta c_{20}^n}{n - \Theta} \quad (8.237)$$

$$\text{Inhibition} \quad \Theta = \frac{-n}{1 + K_I / c_{20}^n} \rightarrow K_I = \frac{(n + \Theta) c_{20}^n}{|\Theta|}. \quad (8.238)$$

The maximal rates k are:

$$k = \frac{r_{c0}}{\frac{c_{10}^n}{K + c_{10}^n} f(c_{20})}. \quad (8.239)$$

In general, one can insert the reaction speed r for the K_i factors; this yields:

$$\text{Activation} \quad f_A = \frac{c^n}{K_A + c^n} = 1 - \frac{\Theta}{n} \quad (8.240)$$

$$\text{Inhibition} \quad f_I = \frac{K_I}{K_I + c^n} = 1 + \frac{\Theta}{n}. \quad (8.241)$$

8.7 Model Reduction for Signal Proteins

Similarly to enzymes, proteins that participate in signal transduction have one or more domains to which effectors can bind and modify their activity. In higher cells, the number of effectors is fairly large, and on the basis of combinational complexity a fairly large number of combination conformations result. A method [10] of model reduction is introduced here which allows using a smaller number of equations, when one is interested in defined physiological quantities, for example, the degree of binding of a protein with an effector.

The idea that would be followed is a transformation of the system, so that the new system has a simplified structure and one does not have to calculate

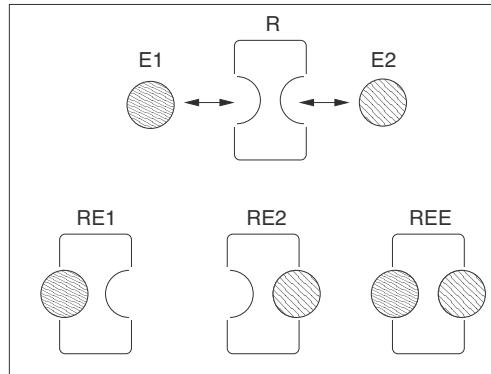
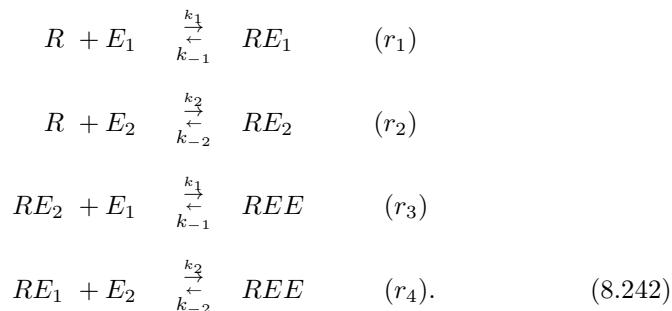


FIGURE 8.15: The receptor R can integrate with $E1$ as well as with $E2$. Interesting quantities are the degree of allocation of the protein with $E1$ and $E2$.

all the properties of the new system but those required for determining the degree of binding. The procedure shall be shown for the example for a protein R and both ligands $E1$ and $E2$ as shown in Figure 8.15. The following reaction scheme results from the figure:



From the figure, one determines that there are 4 possible conformations. This number can also be determined using the same considerations used for the enzyme kinetics. It is assumed that each single binding site can bind only a single type of effector. Therefore, the following possibilities arise for a single effector to bind to the protein:

$$l'_i = \sum_{k=0}^1 \binom{m_i}{k} = m_i + 1. \tag{8.243}$$

Multiplying now for all the effectors to yield the various conformations:

$$l = \prod_{i=1}^n (m_i + 1) \tag{8.244}$$

The transformation will now be carried out, which possesses a hierarchical structure. The new state z_0 is the sum of all conformations. The rest of the new states, z_1 and z_2 , take into account all the conformations with E_1 and E_2 each that bind to R . State z_3 takes into account the conformations where $E1$ as well as $E2$ bind to the protein. The original system and the reaction rates following the previous scheme are:

$$\begin{aligned}\dot{R} &= -r_1 - r_2, & r_1 &= k_1 \text{ } E_1 \text{ } R - k_{-1} \text{ } RE_1 \\ \dot{RE}_1 &= r_1 - r_4, & r_2 &= k_2 \text{ } E_2 \text{ } R - k_{-2} \text{ } RE_2 \\ \dot{RE}_2 &= r_2 - r_3, & r_3 &= k_1 \text{ } RE_2 \text{ } E_1 - k_{-1} \text{ } REE \\ \dot{REE} &= r_3 + r_4, & r_4 &= k_2 \text{ } RE_1 \text{ } E_2 - k_{-2} \text{ } REE.\end{aligned}\quad (8.245)$$

The new states are defined as follows:

$$\begin{aligned}z_0 &= R + RE_1 + RE_2 + REE \\ z_1 &= RE_1 + REE \\ z_2 &= RE_2 + REE \\ z_3 &= REE\end{aligned}\quad (8.246)$$

and one obtains the transformation matrix T :

$$\underline{z} = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_T \begin{bmatrix} R \\ RE_1 \\ RE_2 \\ REE \end{bmatrix}. \quad (8.247)$$

The inverse transformation matrix T^{-1} reads then:

$$T^{-1} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (8.248)$$

The dynamics of the new states are:

$$\begin{aligned}\dot{z}_0 &= 0, & \dot{z}_2 &= \dot{RE}_2 + \dot{REE} \\ \dot{z}_1 &= \dot{RE}_1 + \dot{REE}, & \dot{z}_3 &= \dot{REE}.\end{aligned}\quad (8.249)$$

The old states can be reconstructed, where one uses the transformation matrix

T^{-1} . One obtains:

$$\begin{aligned} R &= z_0 - z_1 - z_2 + z_3 = R_0 \\ RE_1 &= z_1 - z_3 \\ RE_2 &= z_2 - z_3 \\ REE &= z_3. \end{aligned} \tag{8.250}$$

Of interest now are the states that each denote a degree of binding: z_1 and z_2 . Those dynamics are independent of the z_3 dynamics. Therefore, it is sufficient to consider only those two new states. One obtains:

$$\begin{aligned} \dot{z}_1 &= r_1 + r_3 = k_1 E_1 R - k_{-1} RE_1 + k_1 RE_2 E_1 - k_{-1} REE \\ &= k_1 E_1 (z_0 - z_1 - z_2 + z_3) - k_{-1} (z_1 - z_3) + k_1 E_1 (z_2 - z_3) - k_{-1} z_3 \\ &= k_1 E (z_0 - z_1) - k_{-1} z_1. \end{aligned} \tag{8.251}$$

$$\begin{aligned} \dot{z}_2 &= r_2 + r_4 = k_2 E_2 R - k_{-2} RE_2 + RE_1 E_2 - k_{-2} REE \\ &= k_2 E_2 (z_0 - z_1 - z_2 + z_3) - k_{-2} (z_2 - z_3) + k_2 E_2 (z_1 - z_3) - k_{-2} z_3 \\ &= k_2 E_2 (z_0 - z_2) - k_{-2} z_2 \end{aligned} \tag{8.252}$$

Hence, the original system is reduced to two differential equations: both (new) equations do not depend on z_3 . The reduction method can be very helpful in large systems where only a few state quantities are of interest.

Exercises

EXERCISE 8.1 Enzyme kinetics (*Simulation exercise*).

Use the standard scheme for Michaelis-Menten to show that the system

$$\begin{aligned} \dot{S} &= -k_1(E_0 - ES)S + k_{-1}ES + F \\ \dot{ES} &= k_1(E_0 - ES)S - (k_{-1} + k_2)ES \end{aligned}$$

with input flux F run on two time scales. Do the following steps: Linearize the system of two state variables S and ES . Calculate the eigenvalues and eigenvectors with steady state data for S and ES . Based on these results, perform a model reduction by assuming that the fast state is in steady state. Parameters are: $k_1 = 70$, $k_{-1} = 10$, $k_2 = 1.5$, $E_0 = 1.0$, $F = 1.4$.

EXERCISE 8.2 Time scale separation of a small network.

Consider the system



with the reaction rates: $r_1 = k_1 A$, $r_2 = k_2 B - k_{-2} C$ and $r_3 = k_3 C$.

Set up the differential equations for the system. Assume that r_2 is the fast reaction. Apply the method described above for time scale separation. How can the system be reduced to only two differential equations and one algebraic equation?

SOLUTION: The differential equations are then:

$$\begin{aligned}\dot{A} &= -r_1 \\ \dot{B} &= r_1 - r_2 \\ \dot{C} &= r_2 - r_3;\end{aligned}\quad (8.254)$$

with the fast reaction r_2 , the system is reformulated:

$$\begin{pmatrix} \dot{A} \\ \dot{B} \\ \dot{C} \end{pmatrix} = \underbrace{\begin{pmatrix} -1 & 0 \\ 1 & 0 \\ 0 & -1 \end{pmatrix}}_{N_s} \underbrace{\begin{pmatrix} r_1 \\ r_3 \end{pmatrix}}_{\underline{r}_s} + \underbrace{\begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}}_{N_f} \cdot \underbrace{r_2}_{z_f}. \quad (8.255)$$

For the fast reaction r_2 (k_2 is taken as scaling factor) with constant $K_2 = \frac{k_{-2}}{k_2}$ the following applies:

$$r_f = k_2 (B - K_2 C) = \frac{1}{\epsilon} \bar{r}_f \quad \text{and thus} \quad z = \lim_{\epsilon \rightarrow 0} \frac{\bar{r}_f}{\epsilon}. \quad (8.256)$$

Applicable for z_f :

$$z_f = - \left[\frac{\partial \bar{r}}{\partial c} N_s \right]^{-1} \frac{\partial \bar{r}}{\partial c} \cdot (N_s \underline{r}_s). \quad (8.257)$$

Differentiating to obtain:

$$\begin{aligned}\frac{\partial \bar{r}_f}{\partial c} &= \begin{pmatrix} 0 & 1 & -K_2 \end{pmatrix} \\ \rightarrow \quad \frac{\partial \bar{r}_f}{\partial c} N_f &= \begin{pmatrix} 0 & 1 & -K_2 \end{pmatrix} \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} \\ &= -1 - K_2 = -(1 + K_2);\end{aligned}\quad (8.258)$$

$$(8.259)$$

and finally one gets:

$$\begin{aligned} z_f &= \frac{1}{1+K_2} \begin{pmatrix} 0 & 1 & -K_2 \end{pmatrix} \begin{pmatrix} -r_1 \\ r_1 \\ r_3 \end{pmatrix} \\ &= \frac{1}{1+K_2} (r_1 + K_2 r_3). \end{aligned} \quad (8.260)$$

A model reduction can now be carried out where only the slow modes are considered. The previous differential equations now are:

$$\begin{aligned} \begin{pmatrix} \dot{A} \\ \dot{B} \\ \dot{C} \end{pmatrix} &= N_s \cdot \begin{pmatrix} r_1 \\ r_3 \end{pmatrix} + N_f z = \begin{pmatrix} -r_1 \\ r_1 \\ r_3 \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} z \\ &= \begin{pmatrix} -r_1 \\ r_1 - \frac{r_1 + K_2 r_3}{1+K_2} \\ r_3 + \frac{r_1 + K_2 r_3}{1+K_2} \end{pmatrix}. \end{aligned} \quad (8.261)$$

The null space K of N_f^T is now:

$$K = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}. \quad (8.262)$$

Thus

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{pmatrix} A \\ B+C \end{pmatrix}$$

$$\rightarrow w_1 = A; \quad w_2 = B+C. \quad (8.263)$$

For the complete transformation, the condition $\bar{x}_f = 0$ must be evaluated:

$$r_2 = 0 = B - K_2 C \rightarrow B = K_2 C. \quad (8.264)$$

Thus allowing the expression of the old variables in terms of the new variables:

$$A = w_1; \quad B = \frac{K_2 w_2}{1 + K_2}; \quad C = \frac{w_2}{1 + K_2}. \quad (8.265)$$

Inserting to yield the reduced model:

$$\begin{aligned} \begin{pmatrix} \dot{w}_1 \\ w_2 \end{pmatrix} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} -r_1 \\ r_1 \\ -r_3 \end{pmatrix} = \begin{pmatrix} -r_1 \\ r_1 - r_3 \end{pmatrix} \\ &= \begin{pmatrix} -k_1 w_1 \\ k_1 w_1 - k_3 \frac{w_2}{1 + K_2} \end{pmatrix}. \end{aligned} \quad (8.266)$$

EXERCISE 8.3 Metabolic control analysis.

In this exercise we make clear the difference between local and global properties in Metabolic control analysis. Given is a branch point where an incoming flux is divided into two branches. The following differential equation will be used:

$$\dot{X} = r_1 - r_2 - r_3. \quad (8.267)$$

Use the following kinetics:

$$r_2 = k_2 X \quad \text{and} \quad r_3 = k_3 X. \quad (8.268)$$

a Use the formulas given in the section to derive the matrix of flux control coefficients \mathbf{C}^J and the matrix of concentration control coefficients \mathbf{C}^S . What is the dimension of the matrices?

b Calculate the influence of parameter k_2 in reaction rate r_2 on the steady flux J_3 using the derived matrices.

c Now verify your calculation by solving the system analytically. Do the following steps: Determine the steady state for the given set of parameters. Then perturb the system by parameter $k'_2 = k_2(1 + \Delta)$. Calculate the "local" change in rate r_2 based on the change of parameter k_2 . Then, calculate the new steady state and determine the change in flux J_3 which we denote as J'_3 . For the calculation take lim for $\Delta \rightarrow 0$.

SOLUTION:

To derive the equations for the flux and concentration control coefficients we can apply the following general equation:

$$\begin{pmatrix} \mathbf{C}^J \\ \mathbf{C}^S \end{pmatrix} \begin{pmatrix} K & \mathbf{E} \end{pmatrix} = \begin{pmatrix} K & \mathbf{0} \\ \mathbf{0} & -I \end{pmatrix}. \quad (8.269)$$

From this equation we can extract two equations:

$$\mathbf{C}^J \begin{pmatrix} K & \mathbf{E} \end{pmatrix} = \begin{pmatrix} K & \mathbf{0} \end{pmatrix} \quad (8.270)$$

and

$$\mathbf{C}^S \begin{pmatrix} K & \mathbf{E} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & -I \end{pmatrix}. \quad (8.271)$$

In the example, K is the null space of the stoichiometric matrix $N = (1 \ -1 \ -1)$. Matrix N has one row and three columns; therefore, the null space has two vectors:

$$K = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (8.272)$$

The null space represents two ways through the system either via r_1 and r_2 or via r_1 and r_3 . Taking into account that r_1 is the input flux, \mathbf{E} is a column vector with one zero and two further entries representing the derivatives of the reaction rates with respect to concentration X :

$$\mathbf{E} = \begin{pmatrix} 0 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix} = \begin{pmatrix} 0 \\ k_2 \\ k_3 \end{pmatrix}. \quad (8.273)$$

Matrix K together with \mathbf{E} build a 3×3 matrix that is not singular. For the flux control coefficients we get therefore:

$$\mathbf{C}^J = \begin{pmatrix} K & \mathbf{0} \end{pmatrix} \begin{pmatrix} K & \mathbf{E} \end{pmatrix}^{-1} \quad (8.274)$$

which results after some algebraic calculations:

$$\mathbf{C}^J = \frac{1}{k_2 + k_3} \begin{pmatrix} k_2 + k_3 & 0 & -k_2 \\ k_2 & k_3 & -k_2 \\ k_3 & -k_2 & k_2 \end{pmatrix}. \quad (8.275)$$

We calculate the change of rate 2 by parameter k_2 at steady state. That is:

$$\frac{dr_2}{dk_2} = X|_{ss}. \quad (8.276)$$

To determine the influence of k_2 on steady state flux J_3 , element $C_{3,2}^J$ must be considered. We finally get:

$$\frac{dJ_3}{dr_2} = \frac{-k_3}{k_2 + k_3} X|_{ss}. \quad (8.277)$$

Now we calculate the complete analytical solution. For a fixed value of $r_1 = u$ as input, the steady state value for X is:

$$X|_{ss} = \frac{u}{k_2 + k_3}, \quad (8.278)$$

and the steady state rates are therefore:

$$\begin{aligned} J_2 &= k_2 X|_{ss} = \frac{k_2 u}{k_2 + k_3} \\ J_3 &= k_3 X|_{ss} = \frac{k_3 u}{k_2 + k_3}. \end{aligned} \quad (8.279)$$

Now we perturb the system and calculate the new steady state for parameter $k'_2 = k_2(1 + \Delta)$. The values for the state variable X and fluxes J_2 and J_3 for the perturbed parameter are:

$$\begin{aligned} X'|_{ss} &= \frac{u}{k_2(1 + \Delta) + k_3} \\ J_2 &= k_2(1 + \Delta) X'|_{ss} = \frac{k_2(1 + \Delta) u}{k_2(1 + \Delta) + k_3} \\ J_3 &= k_3 X'|_{ss} = \frac{k_3 u}{k_2(1 + \Delta) + k_3}. \end{aligned} \quad (8.280)$$

Next, we calculate the difference in the third rate, ΔJ_3 :

$$\begin{aligned} \Delta J_3 &= J'_3 - J_3 = \frac{k_3 u}{k_2(1 + \Delta) + k_3} - \frac{k_3 u}{k_2 + k_3} \\ &= -\frac{\Delta k_2 k_3 u}{(k_2(1 + \Delta) + k_3)(k_2 + k_3)}, \end{aligned} \quad (8.281)$$

and we get for the difference quotient with $X|_{ss}$ from above:

$$\frac{\Delta J_3}{\Delta k_2} = -\frac{k_3}{k_2(1 + \Delta) + k_3} X|_{ss}. \quad (8.282)$$

Finally we take lim for $\Delta \rightarrow 0$:

$$\frac{dJ_3}{dk_2} = \lim_{\Delta \rightarrow 0} \frac{\Delta J_3}{\Delta k_2} = -\frac{k_3 u}{k_2 + k_3} X|_{ss}. \quad (8.283)$$

Now we have to calculate the local change of rate r_2 by the perturbation in parameter k_2 :

$$r'_2 = k_2(1 + \Delta) X|_{ss}. \quad (8.284)$$

Note here the difference: For the local change we consider the first steady state; we get for the difference:

$$\Delta r_2 = r'_2 - r_2 = k_2(1 + \Delta) X|_{ss} - k_2 X|_{ss} = k_2 \Delta X|_{ss}, \quad (8.285)$$

and for the difference quotient:

$$\frac{\Delta r_2}{\Delta k_2} = X|_{ss} \rightarrow \frac{dr_2}{dk_2} = X|_{ss}. \quad (8.286)$$

Now we have both parts available and we see that the control coefficient that is the ratio dJ_3/dr_2 is the same as above: $-k_3/(k_2 + k_3)$.

EXERCISE 8.4 Structured kinetic modeling.

Use the same structure of the reaction system as in the relevant chapter (two metabolites A and B , and reactions r_1, r_2, r_3). Instead of the activation of r_2 by metabolite B use an inhibition kinetics of the form $K_I/(K_I + B)$. Calculate the Jacobian of the system and perform a numerical simulation by varying two parameters. What do you observe?

Bibliography

- [1] Z. P. Gerdtzen et al. Non-linear reduction for kinetic models of metabolic reaction networks. *Metab. Eng.* 6, 2004, 140-154.
- [2] J. W. Hearne. Sensitivity analysis of parameter combinations. *Appl. Math. Modelling* 9, 1985, 106–108.
- [3] R. Steuer et al. Robust signal processing in living cells. *PLoS Comp. Biol.* 7(11): e1002218, 2011.
- [4] M. A. Savageau. Biochemical System Analysis: Some mathematical properties of the rate law for the component enzymatic reactions. *J. Theor. Biol.* 25, 1969, 365-369.
- [5] H. Kacser and J.A. Burns. The control of flux. *Symp. Soc. Exp. Biol.* 27, 1973, 65-104.
- [6] R. Heinrich and S. Schuster. *The control of cellular systems*. Springer Verlag, 1996.
- [7] J.-H. S. Hofmeyr and A. Cornish-Bowden. Co-response analysis: a new experimental strategy for metabolic control analysis. *J. Theor. Biol.* 182, 1996, 371-380.

- [8] M.A. Savageau. *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*. Addison-Wesley, 1976.
- [9] R. Steuer et al. Structural kinetic modeling of metabolic networks. *PNAS* 103, 2006, 11868-11873.
- [10] H. Conzelmann et al. A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks. *BMC Bioinformatics* 7, 2006, 34.
- [11] Ruijter et al. Overexpression of phosphofructokinase and pyruvate kinase in citric acid-producing *Aspergillus niger*. *Biochimica et Biophysica Acta (BBA)*, 1334(2) 1997, 317–326.

Chapter 9

Aspects of Control Theory

In this section, theoretical control theory aspects are considered. First, structural characteristics are treated like observability and stability. Questions of stability are of interest when multiple steady states are present, that is, for one given input value more than one steady state of the state variables can be calculated. The presence of multiple steady states was already shown in one of the previous chapters and shall be highlighted again in light of the structures of biochemical networks. Feedback and feed-forward loops have multiple tasks in engineering sciences: For instance to reach a set point that is given from outside the system or to quickly compensate for disturbances. This will be shown later in an example of bacterial chemotaxis; for biochemical networks it is often not possible to clearly identify parts representing the controlled subsystem and the controller itself. Aspects of robust control come into play when uncertainties are to be taken into account.

9.1 Observability

What one understands from the property of observability is the possibility that the state variables \underline{x} of a system can be reconstructed from measured output variables \underline{y} . In most cases the number of measurable states is far smaller than the number of all state variables. This can only be carried out when there is a formally known correlation between the state and output variables. This implies that a sufficiently precise mathematical model must be present in order to carry out such an analysis. The problem can be solved in two steps. Next, observability will be determined and defined, and when possible, the state variables will be reconstructed. Therefore, an observer (or a filter) must be designed that enables the estimation of the course of the state variables over time. This method is of important significance for systems biology as the development of suitable measurement methods and their use is time costly (e.g., using cDNA chips). Further, such an aid tool enables a better understanding of cellular processes, even if some of the components can not be measured directly. If all state variables could be reconstructed it would be possible to efficiently control the system. Another application is relevant

for model reduction: Variables that are not observable are in some cases not of interest and thus, can be eliminated from the model. The intention of this section is not to give a complete mathematical presentation of the topic but to illustrate the main properties using simple examples as shown in Figure 9.1A. The reaction kinetics are assumed to be irreversible and are of first order.

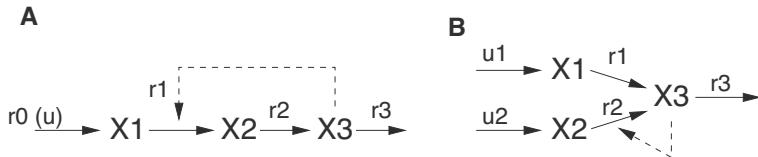


FIGURE 9.1: **A:** Linear signal path with feedback. The feedback enables observing the system with only a measurement of X_1 . **B:** Signal path of a small network to examine the structural observability.

Let us consider a system with three components, X_1 , X_2 and X_3 , for which the differential equations for the concentrations read:

$$\begin{aligned}\dot{X}_1 &= r_0 - r_1 = r_0 - f(X_1, X_3) \\ \dot{X}_2 &= r_1 - r_2 = k_1 f(X_1, X_3) - k_2 X_2 \\ \dot{X}_3 &= r_2 - r_3 = k_2 X_2 - k_3 X_3\end{aligned}\quad (9.1)$$

which can be written in a compact matrix form for the simplified case $f(X_1, X_3) = k_1 X_1$:

$$\begin{aligned}\underline{\dot{x}} &= A \underline{x} + \underline{b} u \\ \underline{y} &= C \underline{x}.\end{aligned}\quad (9.2)$$

Matrix A and vector \underline{b} are:

$$A = \begin{bmatrix} -k_1 & 0 & 0 \\ k_1 & -k_2 & 0 \\ 0 & k_2 & -k_3 \end{bmatrix}, \quad \underline{b} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}. \quad (9.3)$$

The measured variables are assigned to matrix C . System input u is the rate r_0 . For linear systems like the one given here an observability analysis is carried out by calculating the rank of a matrix P that is given in the following form:

$$P = [C^T \ A^T C^T \ \dots \ (A^T)^{n-1} C^T] = [C^T \ A^T C^T \ (A^T)^2 C^T]. \quad (9.4)$$

Matrix P is the observability matrix. In the case that the rank of P is smaller than $n = 3$, not all the states can be reconstructed. For a system without feedback ($r_1 = k_1 X_1$) and in which a measurement of X_2 yields the rank of P

to be 2, it is not possible to reconstruct all the state variables. Measurement of X_3 shows that P is of full rank and hence all the states can be reconstructed. For linear pathways the explanation is clear: The information that can be obtained from the first component in the pathway doesn't contain any information from the downstream part of the pathway; on the other hand, the last component in the pathway contains all information from the upstream part. Linear metabolic pathways are frequently controlled ($r_1 = f(X_1, X_3)$) via allosteric inhibition through feedback by the first enzyme in the pathway. If the feedback is included in the calculation, then the rank of P — when X_1 measured — is (almost) 3 and the system is observable (for the example above the term for f must be linearized). The reason for this is that the feedback transfers information from X_3 to the beginning of the path, or, in other words, X_1 contains information of all elements of the control loop. Therefore, a measurement of X_1 suffices to reproduce all the state variables.

The concept of observability was expanded so that it is independent of the numerical values of the parameters. This is where the structural observability analysis enters the picture. Structural observability analyzes the structure of the matrices S_A and S_C , where S_A and S_C possess the same dimension as A and C in Equation (9.2). However, every entry contains a * instead of the numerical values. The system is structurally observable when (i) the output/measurement variables along with all the state variables can be represented as a graph with nodes that directly or indirectly connect to each other; and (ii) the structural rank of the matrix S_P is n :

$$S_P = \begin{pmatrix} S_A \\ S_C \end{pmatrix}. \quad (9.5)$$

In order to determine the structural rank of matrix S_P , n columns with at least a single * entry must be found and labeled. For biochemical networks, this condition is almost fulfilled, as the components typically have a direct influence on themselves. To represent the dynamic system graphically, entries from the Jacobi matrix of a system given with ordinary differential equations will be used. The entries indicate whether an element j has influence on the element i or not.

EXAMPLE 9.1 Simple Reaction System.

A simple example is shown in Figure 9.1B. Components X_1 and X_2 are connected to a third component X_3 . If only X_3 can be measured, matrix S_P reads:

$$S_P = \begin{pmatrix} * & 0 & 0 \\ 0 & * & * \\ * & * & * \\ \dots & \dots & \dots \\ 0 & 0 & * \end{pmatrix}. \quad (9.6)$$

The conditions for this example are fulfilled as X_3 can be reached from X_1 and X_2 ($S_p(3,1) = S_p(3,2) = *$). Furthermore, the structural order is equal to 3 as the diagonal elements of the upper parts of S_P have a single entry. If it is assumed that reaction r_2 in Figure 9.1 is given as:

$$r_2 = k_2 \frac{X_2}{K_2 + X_2} X_3 \quad (9.7)$$

and if $X_2 \gg K_2$ is valid, r_2 is simplified to:

$$r_2 = k_2 X_3. \quad (9.8)$$

Then, S_P in this case is:

$$S_P = \begin{pmatrix} * & 0 & 0 \\ 0 & 0 & * \\ * & 0 & * \\ \dots & \dots & \dots \\ 0 & 0 & * \end{pmatrix} \quad (9.9)$$

and the structural rank is only 2 and the system is in this case not structurally observable. The structural considerations guarantee that the system is observable for *almost* all parameters' values (required condition but not sufficient). A counterexample is as follows: Both metabolites X_1 and X_2 are running on the same time scale, that is, the entries in the systems matrix, $\frac{dx_1}{dx_1} = a_{11} = -1$ and $\frac{dx_2}{dx_2} = a_{22} = -1$, are the same. The other entries in the system matrix A and matrix C^T are given as:

$$A = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & a_{23} \\ a_{31} & a_{32} & -3 \end{pmatrix}; \quad C^T = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \quad (9.10)$$

Matrix P is then:

$$P = \begin{bmatrix} 0 & a_{31} & -4a_{31} \\ 0 & a_{32} & -4a_{32} \\ 1 & -3 & (9 + a_{23}a_{32}) \end{bmatrix}. \quad (9.11)$$

One can see by close inspection of the first two rows matrix P has rank = 2; that is, the system is not completely observable.

Biochemical networks are often analyzed in bio-reactor systems and, therefore, the environmental variables like biomass/optical density and substrate are taken into account as state variables. It is shown that the basic structure discussed above has some interesting implications for observability. The

equations that describe the growth of a culture in a batch experiment in a bio-reactor have a special structure that enables the reconstruction of the state variables. Equations for biomass B , substrate S and internal metabolic products M_i read (see above):

$$\begin{aligned}\dot{B} &= \mu B = Y r_{up} B \\ \dot{S} &= -r_{up} B \\ \dot{M}_i &= \sum_j \gamma_{ij} r_{ij} - \mu M_i = \sum_j \gamma_{ij} r_{ij} - Y r_{up} M_i\end{aligned}\quad (9.12)$$

where r_{up} is the substrate uptake rate, γ_{ij} the stoichiometric coefficients, r_{ij} the kinetic rates, and Y the yield coefficient. The substrate uptake rate r_{up} is in general dependent on the substrate concentration. If growth is assumed to be dependent on the uptake rate $\mu = Y r_{up}$, a graph based on the Jacobian matrix can be drawn similar to that in Figure 9.2. The solid arrows mark the links between the state variables. All internal metabolic products can be used to reconstruct the biomass and the substrate, as both state variables can be linked to any metabolic product (dashed arrow 1).

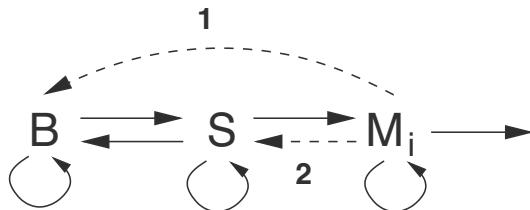


FIGURE 9.2: General structure of a cellular model which describes the uptake and metabolism of substrates. For the solid lines, it is assumed that the uptake rate depends only on the substrate. When one or more metabolic products M_i are involved in the uptake process, for example, $r_{up} = f(S, M_i)$, the graph with dashed lines (1) and (2) is expanded.

Furthermore, if the uptake rate is dependent on the intracellular metabolic product (dashed arrow 2) like for example ATP or one of the components of the phosphotransferase system (PTS), then this metabolic product influences the biomass concentration and therefore all metabolic products that have a direct or an indirect connection. Thus, the system is observable with respect to the involved metabolites only by measuring biomass or substrate concentration. From the observability analysis, it is possible to support the reconstruction of biochemical networks. A comparison between a simulation with a mathematical model and an actual measurement can be helpful to identify previously unknown connections.

9.2 Monotone Systems

The most important task of control engineering is to design a controller that ensures a system consisting of a controller and a controlled system is stable. Basically, a living system exhibits a stable behavior (including attractors like limit cycles); on the other hand, a component could accumulate leading to the destruction of an organism. Hence, this suggests a simple behavior in biochemical reaction networks. Nevertheless, two or more steady states can be observed due to the non linearities in biochemical systems. In addition, such phenomenon has been found to be used in nature for signal processing as described above. A system with multiple steady states can have a reversible/irreversible switching behavior which is used in decision making e.g., differentiation or cell division (examples can be found also in the signal transduction chapter). In general, not all steady states can be observed experimentally as they can be unstable. The analysis of such systems which show such an interesting behavior can help in inferring the network structure and enable new possibilities of active intervention in those systems.

The dynamics of signal pathways and various functionalities were already discussed. An important concept which will be used in control engineering is the analysis of a whole system from the characteristics of its individual components. In control theory, an open control loop consisting of a controller and a controlled system will be analyzed and the behavior of the closed loop will be determined from the characteristics of the individual components. This is the classical way of designing a controller so as to meet the requirements pertaining to disturbances or set point specifications.

9.2.1 Monotony of networks

Sontag and coworkers considered a special class of monotone systems and showed that under certain conditions, such systems will not exhibit oscillatory or chaotic behavior [1]. Monotone systems can be non linear but yet possess useful characteristics in mathematical sense: If the stimulation (or other conditions) is modified, for example, increased: $u_2 > u_1$, the system will react with a higher output value $y_2(t) > y_1(t)$ (or lower value, both cases occur but never simultaneously). One therefore expects if a higher input is applied then the output would be higher (or lower, respectively). A similar behavior is also applicable for a variation of the initial values $y_2(t = 0) > y_1(t = 0)$. Those intuitive properties characterize monotone systems. Monotony (input/output monotony for single input single output systems only) can be guaranteed when two conditions are met. These conditions are related to a representation of the system under consideration as a graph (see chapter on modeling strategies).

Condition (i): For the Jacobian matrix $J|_{x=\bar{x}, u=\bar{u}}$ of the linearized system at the equilibrium position \bar{x} , \bar{u} , the signs of all entries (entries in the main diagonal play no role) and the signs of the derivatives in vector \underline{b} ($\underline{b} = \partial f / \partial u|_{x=\bar{x}, u=\bar{u}}$) are uniquely specified with respect to the sign.

Condition (ii): Cycles (or loops) in the resulting graph have positive signs. The sign of a cycle is determined by multiplying the sign of the individual arrows. In other words, the signs of the arrows that are between any two arbitrary nodes must be the same.

The property of monotony is of special importance for biochemical systems as it renders the analysis easy to perform, since information regarding relationships between the elements (e.g., arrows with + for activation and - for inhibition) often found as graphs in relevant publications. Let us analyze a simple network with two components. The property of monotony can here be easily demonstrated for a linear two dimensional system where stability is presumed. The Jacobian matrix A has four entries which describe the interaction of the components involved as shown in Figure 9.3.

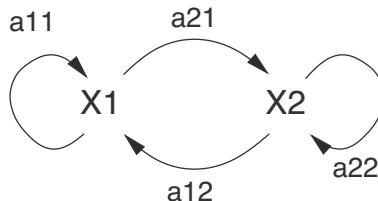


FIGURE 9.3: Two dimensional system.

The following equations are valid in general:

$$\begin{aligned}\dot{X}_1 &= a_{12} X_1 + a_{11} X_2 \\ \dot{X}_2 &= a_{21} X_1 + a_{22} X_2.\end{aligned}\quad (9.13)$$

The following characteristic equation results for the system:

$$\lambda^2 + c_1 \lambda + c_0 \stackrel{!}{=} 0 \quad (9.14)$$

where on the basis of the stability condition, the following equation is valid:

$$c_1 = -\text{trace}(A) \geq 0, \quad c_0 = \det(A) \geq 0. \quad (9.15)$$

The solution of the characteristic equation leads to:

$$\lambda_{1,2} = -\frac{c_1}{2} \pm \frac{1}{2} \sqrt{c_1^2 - 4c_0}. \quad (9.16)$$

The term under the root is crucial for the dynamic behavior. One calculates:

$$c_1^2 - 4 c_0 = (a_{11} - a_{22})^2 + 4 a_{12} a_{21}. \quad (9.17)$$

The first summand is positive; for the second summand, the signs play a role. Should a monotone system be available, one can test whether the cycles overall have a positive sign. This is the case for the present system when a_{12} and a_{21} possess the same sign (see in Figure 9.3 above that there is just one single cycle in the system). In this case the second summand is then positive as well and all eigenvalues are negative and real.

EXAMPLE 9.2 Lactose uptake.

Using the example of lactose uptake, the procedure will be shown for a larger system (Figure 9.4 left).

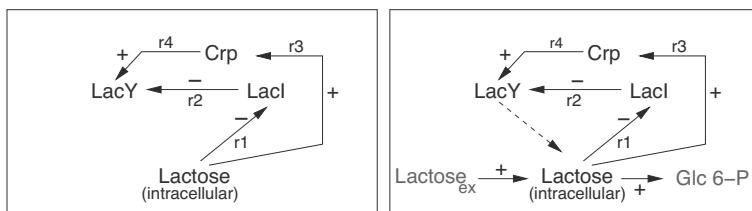


FIGURE 9.4: Scheme of lactose induction, extended to include the influence of the global Regulator Crp. The open loop system is shown on the left side and the closed system on the right side.

As already described earlier, intracellular lactose deactivates the repressor LacI (r_1) and activates the transcription factor Crp (r_3). On lactose the growth rate is slower when compared to glucose; therefore this transcription factor is activated. The repressor inhibits the synthesis of the lactose transporter (r_2) while the synthesis is activated by Crp (r_4). First, a dynamical model is introduced that extends the example in previous chapters. However, this previous scheme is a simplified version and neglects the detailed interaction of the repressor with the promoter binding site. In the following a detailed scheme for the repressor interaction is presented and the consequences are discussed.

For the moment the activation of LacY synthesis by Crp is neglected and is considered later on. For the reaction rates it is assumed as before that 2 molecules of lactose are needed to deactivate the repressor (r_1); the interaction of the repressor with the DNA binding site is now described by two reactions (r_{2a}, r_{2b}) and the expression efficiency is calculated. The reaction scheme is summarized in Figure 9.5.

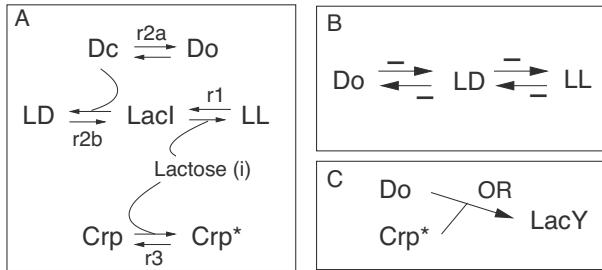


FIGURE 9.5: **A** Detailed reaction scheme for expression of LacY. The reaction scheme comprises four reactions. **B** Subgraph for the interaction of LacI with the DNA binding site. **C** Interaction of activated Crp and the open DNA binding site.

Deactivation of the repressor

$$\text{LacI by lactose (Lac}_{in}\text{)} \quad r_1 = k_1 \text{LacI Lac}_{in}^2 - k_{-1} LL$$

Open complex formation

$$r_{2a} = k'_2 D_c - k'_{-2} D_o$$

Binding of the repressor to DNA $r_{2b} = k_2 \text{LacI } D_c - k_{-2} LD$. (9.18)

In the scheme so far, conservation equations can be detected:

$$D_0 = D_c + D_o + LD \quad (9.19)$$

$$Laci_0 = LacI + LD + LL. \quad (9.20)$$

As two conservation equations are available the number of state variables can be reduced by two. As shown in the exercise the selection of the state variables that are considered further on as differential equations is a crucial point in the approach. Here, we proceed with D_o , LD and LL , and the differential equations read as follows:

$$\dot{D}_o = r_{2a}(D_0, D_o, LD) \quad (9.21)$$

$$\dot{LD} = r_{2b}(D_0, D_o, LD, LL) \quad (9.22)$$

$$\dot{LL} = r_1(D_0, D_o, LD, LL). \quad (9.23)$$

Now it is shown that both conditions (i) and (ii) are fulfilled. Therefore the Jacobian matrix of this subsystem is calculated (only the entries of the sec-

ondary diagonal are given):

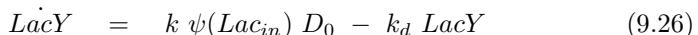
$$J = \begin{pmatrix} * & \frac{\partial r_{2a}}{\partial LD} & 0 \\ \frac{\partial r_{2b}}{\partial D_o} & * & \frac{\partial r_{2b}}{\partial LL} \\ 0 & \frac{\partial r_1}{\partial LD} & * \end{pmatrix} =$$

$$\begin{pmatrix} * & -k_2 & 0 \\ -k'_2(LacI_0 - LD - LL) & * & -k'_2(D_0 - D_o - LD) \\ 0 & -k_1 Lac_{in}^2 & * \end{pmatrix}. \quad (9.24)$$

Close inspection of the entries reveals that all entries are negative and a change of sign is not possible. Therefore, condition (i) is fulfilled. A scheme of the reduced graph is given above in Figure 9.5 in plot **B**. As can be seen, no additional cycle could be detected based only on the interaction of LacI with the DNA binding site. To activate Crp, a further reversible reaction is taken into account:



To calculate the gene expression efficiency, a simple OR connection between the two parts is used (Figure 9.5 in plot **C**) and the differential equation for the protein reads:



where ψ describes the expression efficiency as a function of internal lactose. The OR gate is valid, since Crp acts as a non essential activator. Now Figure 9.4 is considered again. Two interactions contribute to the synthesis of LacY. In the model, both parts are combined as described. In the open loop whose input is lactose and output is LacY, two paths from the lactose to the transporter result, and the sign of the cycle is positive.

9.2.2 Multiple stationary states

The property of monotony can now be used especially for the determination of multiple stationary states. Here, the multi-stability of the closed loop can be ensured when (i) the system in the open loop is monotonic and a unique signal response curve $S(u)$ exists and (ii) the output of the system (LacY here) is fed back through a monotonic increasing function g (that is a

linear relationship or a sigmoid kinetics) to the input (lactose here). More generally, g could also be a monotone system. The stationary states of the closed loop can then be determined from the intersection points of the functions $S(u)$ and g^{-1} and stability is characterized by conditions:

Stable steady state: $S' < g'^{-1}$, unstable: $S' > g'^{-1}$.	(9.27)
---	--------

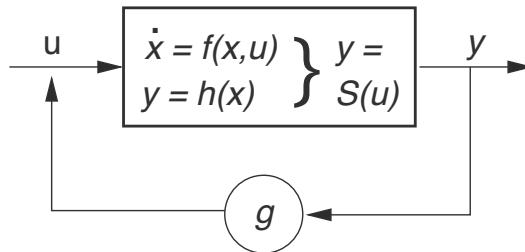


FIGURE 9.6: Schematic for analysis of monotonic systems: The open loop is described by a stationary input output response curve $S(u)$. The system is fed back by a function $u = g(y)$. The behavior can be determined from the intersection points of the functions S and g^{-1} .

EXAMPLE 9.3 Lactose uptake (continued)

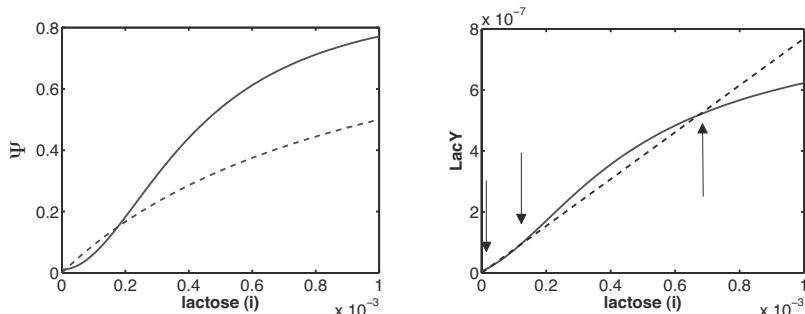


FIGURE 9.7: Left: Stationary relationship between intracellular lactose, D_o and Crp^* (active Crp), respectively. Right: stationary characteristics of the open loop $S(u)$ (solid) and the function g^{-1} (dashed). Three intersection points can be detected (see arrows). From slopes of the intersection points one determines that there are two stable and one unstable steady state.

The graphs on the left of Figure 9.7 show the relationship between the input lactose and output (D_o and Crp^*). The system is closed by a stationary

feedback:

$$Lac_{in} = k_l LacY. \quad (9.28)$$

The significance of this is that the internal lactose amount is proportional to the enzyme amount $LacY$. The right of Figure 9.7 shows the stationary values of the open loop $S(u)$ and the feedback g^{-1} .

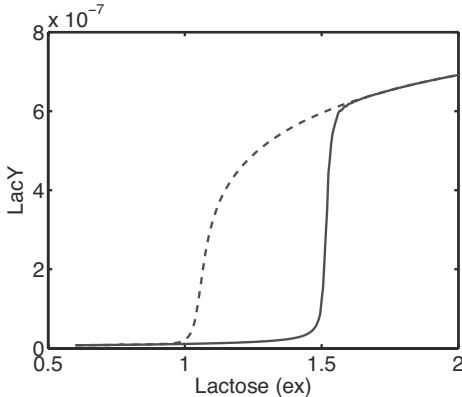


FIGURE 9.8: System behavior when extracellular lactose affects the input, and therefore the slope of function g^{-1} is changing. For a small range of the input signal two stable steady states exist. The system exhibits hysteresis.

Influencing the system now as shown in Figure 9.4 with extracellular lactose Lac_{ex} , the equation of the feedback is modified:

$$Lac_{in} = k'_l LacY Lac_{ex}. \quad (9.29)$$

Since now the slope of g (and of g^{-1}) is modified by different values of Lac_{ex} , the number of intersection points will change accordingly. Simulating the system for increasing values of the external lactose, one finds that for a high concentration the system switches on; that is, a high expression of $LacY$ is observed. Reducing the input signal, the expression initially remains at a high level. The system will switch off first at a low concentration. The example shows clearly that non linearities in cellular system are required to realize a well defined functionality: In the example, the nonlinearity is used to ensure that the system avoids a frequent on/off switching if a fluctuation of the stimulant is present in the environment of the cells.

9.2.3 From nonmonotone systems to monotone systems

In many cases, the structure at hand can not be analyzed using the introduced methodology due to the system's nonmonotone behavior. It is sometimes possible to split the system into two parts which then can be analyzed as a feedback system as described earlier.

EXAMPLE 9.4 Non monotone system.

Figure 9.9 shows a network which shows a non monotone structure (plot in the first row). The following system of differential equations is applicable here:

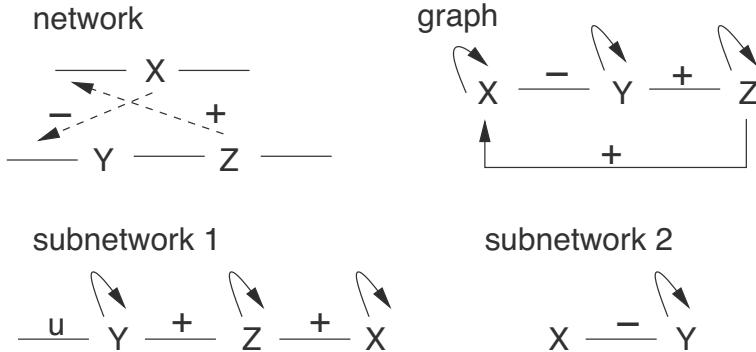


FIGURE 9.9: Network, graph and subsystems of the example case.

$$\begin{aligned}\dot{X} &= -k X + Z \\ \dot{Y} &= \frac{1}{1+X} - k Y \\ \dot{Z} &= k Y - Y.\end{aligned}\quad (9.30)$$

The resulting Jacobian matrix of the system with steady state X_{ss} is:

$$J = \begin{pmatrix} -k & 0 & 1 \\ \frac{-1}{1+X_{ss}} & -k & 0 \\ 0 & k & -1 \end{pmatrix}. \quad (9.31)$$

One can see that the signs are well defined; there is no sign change. Considering the graphs in the previous figures, one identifies that a feedback loop is present, and that the loop has a negative sign. Therefore a non monotone system is present. To solve the problem, one splits the system into two components as shown in the bottom of the figure. This leads to the following system of

equations:

subsystem 1	subsystem 2
$\dot{Y}_1 = u - k Y_1$	$\dot{Y}_2 = \frac{1}{1 + X_2} - k Y_2$
$\dot{Z}_1 = k Y_1 - Z_1$	
$\dot{X}_1 = Z_1 - k X_1$	(9.32)

For subsystem 1, an analysis will be performed in order to determine the response curve $X_1 = S(u)$. The steady state is given by:

$$Y_1 = \frac{u}{k}, \quad Z_1 = u, \quad X_1 = S(u) = \frac{u}{k}. \quad (9.33)$$

For the second subsystem, the input quantity X_2 and the steady state are given by:

$$Y_2 = \frac{1}{k(1 + X_2)}. \quad (9.34)$$

To connect both systems, the equation for subsystem 2 must be converted:

$$X_2 = \frac{1 - k Y_2}{k Y_2}. \quad (9.35)$$

Figure 9.10 shows both response curves and the interconnection of the subsystems.

Formally, setting quantities u , Y_2 , X_1 and X_2 to be equal and calculating the intersection point of the curves leads to:

$$X = \frac{u}{k} = \frac{1 - k u}{k u}. \quad (9.36)$$

A quadratic equation for the case of $k = 1$ appears:

$$u^2 + u - 1 = 0 \quad \longrightarrow \quad u_{ss} = -\frac{1}{2} + \frac{1}{2}\sqrt{5}. \quad (9.37)$$

The steady state is stable as there is only a single intersection point. A calculation to verify the stability can be carried for the whole system, where the eigenvalues of the Jacobian matrix are calculated. For $k = 1$ they read: $\lambda_1 = -1.7$, $\lambda_{2,3} = -0.64 \pm 0.63 i$; therefore the whole system is stable. The example makes clear that the feedback need not be a statistical function; instead, it could be a dynamic monotone system.

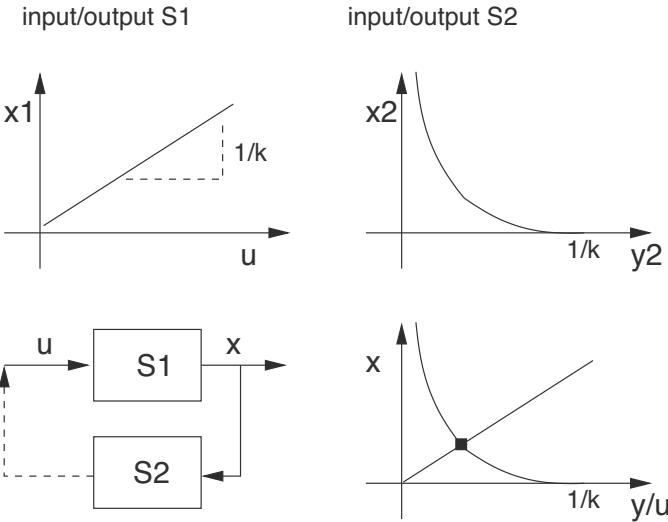


FIGURE 9.10: Characteristic lines and the interconnection of both subsystems.

9.3 Integral Feedback

An interesting biochemical example for feedback can be found in bacterial chemotaxis. An interesting observation is — as already described earlier — that the stationary output (the phosphorylated form of one of the proteins of the system) does not depend on the input which is in this case an attractant for the bacterium, for example, substrates like glucose or amino acids. Using the method of Doyle and coworkers, the system can be represented in a form of closed control loop with an integral feedback [2].

Integral feedback is frequently used in technical systems so as to compensate for disturbances and maintain the system's operation at a desired set-point. As can be seen from Figure 9.11, the system with integral feedback reacts differently to variations in the set-point or with regard to the disturbances. Now we are interested also in the dynamics of the system. Therefore, to keep things not too complicated, it is common to linearize the model in order to use methods of linear control theory. Laplace transformation is often used to transform a system of differential equations into a system of algebraic equations. The transformed system now functions in the frequency domain, i.e., the system is analyzed using frequency ω with regard to the amplitude gain and phase shift of a given sinusoidal input function. The fundamentals of Laplace transformation are given in the Appendix. The following basic observations can be made when considering a linear system: If a system is

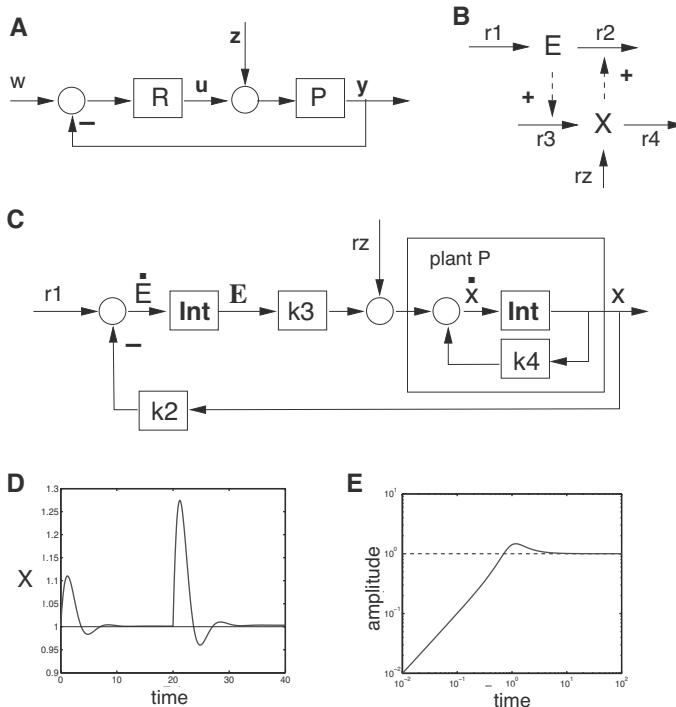


FIGURE 9.11: **A** General model of a closed control loop with integral feedback. **B** Reaction network. **C** Scheme of a reaction network that represents a closed loop system. **Int** signifies integration. When degradation of E proceeds near saturation, the concentration of E becomes almost independent of itself. **D** Simulation using two different values of r_z utilizing equations (9.42). For the second disturbance, the adaption is not perfect. **E** Course of the amplification of S over the frequency range (Equation (9.40)). The areas under and above the dashed line are equal (note the logarithmic representation).

stimulated with a frequency ω , the system's response is characterized via: (i) the same frequency ω , resulting in a phase shift of the input signal and (ii) amplification or damping of the input amplitude. The response of the system with respect to changes of the set-point value or with respect to a disturbance can be described using a separate transfer function that results from the Laplace transformation. The transfer functions result from the characteristics of the controller and the controlled system. From an engineering point of view, they have to fulfill certain requirements: the set-point values must be reached quickly and precisely, and the system must be stabilized as quickly as possible in case of disturbances.

For a linear system with a system matrix A , scalar b and output vector c (single input u , single output y), the transfer function P (for example an

algebraic relationship between input and output in the frequency domain) of the controlled system can be calculated as follows:

$$Y = \underline{c}^T (sI - A)^{-1} b U = P U. \quad (9.38)$$

Using the transfer function of the controller R , the output of the closed control loop Y (Figure 9.11) can now be determined from two components, the transfer function for the set-point W and the transfer function for the disturbance Z :

$$Y = \underbrace{\frac{PR}{1+PR}}_{G_1} W + \underbrace{\frac{P}{1+PR}}_{G_2} Z. \quad (9.39)$$

Both transfer functions consist of individual transfer functions of the controller and controlled system. It can be shown the transfer function S plays an important role for the dynamics of the closed control loop:

$$S = \frac{1}{1+PR}. \quad (9.40)$$

There exists a conservation relationship if S is considered for all frequencies ω : The integral over all frequencies is zero. This signifies that some well defined frequency range is suppressed while another range is amplified. As S is present in both transfer functions G_1 and G_2 , one has to find a compromise with the two demands described above.

EXAMPLE 9.5 Simple network with integral feedback.

Shown in Figure 9.11 is a cellular network that shows a near ideal adaptation: Enzyme E catalyzes the synthesis of the metabolic product X by connecting X to the enzyme E 's degradation. According to the figure and assuming Michaelis-Menten kinetics for the protein degradation, the non linear equations for the system read:

$$\begin{aligned} \dot{E} &= r_1 - k_2 \frac{X \cdot E}{E + K_e} \\ \dot{X} &= k_3 E - k_4 X + r_z \end{aligned} \quad (9.41)$$

with the rate r_z being a disturbance affecting X . As can be seen from Figure 9.11, the network can be represented as a control loop with integral feedback. Here r_1 is the set-point value and r_z is a disturbance, although the degradation of the enzyme is preset to proceed in saturated mode, i.e., the decomposition of E is independent of itself/ its concentration. This simplifies Equation (9.41) to:

$$\begin{aligned} \dot{E} &= r_1 - k_2 X \\ \dot{X} &= k_3 E - k_4 X + r_z. \end{aligned} \quad (9.42)$$

Depending on the strength of the disturbance, we observe an almost perfect adaption (Figure 9.11). The Figure shows as well the so-called Bode plot of $S(\omega)$. Based on the conservation law, a part of the curve must be larger than 1 (dotted line). Accordingly, depending on the input frequency, disturbances at low frequencies are suppressed and are amplified at higher frequencies. Those characteristics are described by the term "robust yet fragile" nature of cellular systems with integral feedback [3]. As the input function in general is no pure sinusoidal function but instead a step, the expected time temporal behavior can be seen as a mixture of the response over all frequencies. In the simulation in Figure 9.11, this can be seen in the time course of X , which shows an underswing.

9.3.1 Bacterial chemotaxis

As discussed earlier, bacterial chemotaxis is a suitable application example of integral feedback. What is understood from bacterial chemotaxis is the change of direction of motion of the bacteria via substance concentration gradients. The bacteria move toward an attractant (for example sugar or some amino acids) and move away from repulsive substances (for example metallic ions). For *E.coli* one distinguishes between tumble motion and run motion. In case of the tumble motion, the direction is randomly varied, while for the run motion the direction is fixed. If an *E.coli* culture is stimulated using an attractant, one observes that the tumble frequency is reduced, but it then returns to its original value. This observation is called adaptation or adaptive behaviour.

The biochemical fundamentals will be concisely addressed here and the focus is on the analysis of the system. MCPs (methyl accepting proteins) are receptors that are responsible for the reception of stimuli in the cell wall. For the following consideration it is assumed that the receptors can be active or inactive. Active denotes that the output is switched on such a way that tumble motion is preferred. Further, it is known that the receptor exists in methylated and nonmethylated form. A simplified scheme shown in Figure 9.12 will now be analyzed.

Here we follow a model that is described in [4, 5] and is only slightly simplified. Considering only the receptor, a differential equation for the active methylated X_m^A and the inactive methylated X_m receptors can be obtained:

$$\begin{aligned}\dot{X}_m &= k_1 R f(X) - k_l X_m + k_{l'} X_m^A S \\ \dot{X}_m^A &= k_l \cdot X_m - k_{l'} X_m^A S - k_2 X_m^A B^P\end{aligned}\tag{9.43}$$

where $f(X)$ represents the influence of the nonmethylated receptor. It is assumed now that the enzyme CheR functions in saturation; hence one can set $f(X) = 1$. Introducing the total concentration of methylated receptors X_m^t :

$$X_m^t = X_m + X_m^A,\tag{9.44}$$

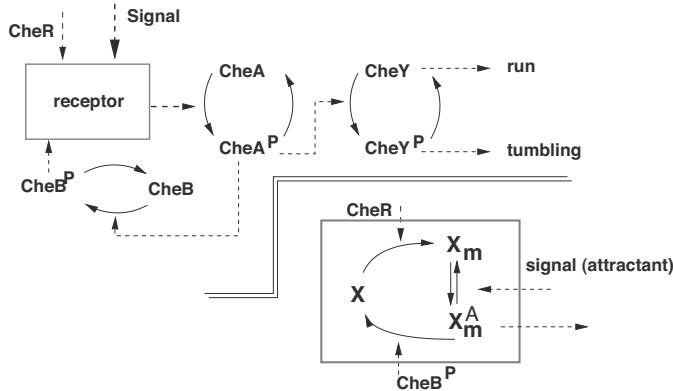


FIGURE 9.12: Molecular reaction diagram of the bacterial chemotaxis. For the model at the bottom, only the receptor dynamics are considered.

one obtains the equations:

$$\begin{aligned}\dot{X}_m^t &= k_1 R - k_2 B^P X_m^A \\ \dot{X}_m^A &= k_l X_m^t - (k_l S + k_2 B^P + k_l) X_m^A.\end{aligned}\quad (9.45)$$

The second equation is non linear and can be linearized. Introducing the following parameter which partially contains the stationary values of the system:

$$\begin{aligned}w &= k_1 \cdot R, \quad p_1 = k_2 \cdot B^P, \quad p_2 = k_l \\ p_3 &= k_2 \cdot B^P + k_l + k'_l \cdot S^{st}, \quad p_4 = k'_l \cdot X_m^{Ast} = k'_l \cdot \frac{k_1 \cdot R}{k_2 \cdot B^P},\end{aligned}\quad (9.46)$$

one obtains:

$$\begin{aligned}\dot{X}_m^t &= w - p_1 \cdot X_m^A \\ \dot{X}_m^A &= p_2 \cdot X_m^+ - p_3 \cdot X_m^A - p_4 \cdot z,\end{aligned}\quad (9.47)$$

where z represents the signal that in control engineering terminology is considered as a disturbance. Representing both equations as a block diagram is shown in Figure 9.13.

Calculating in the Laplace domain as earlier, two transfer functions for output X_m^A with respect to the set-point (W) and the disturbance (Z) are obtained:

$$X_m^A = \underbrace{\frac{p_2}{s^2 + p_3 s + p_1 \cdot p_2}}_{:=G_1, PT_2 \text{ behavior}} W - \underbrace{\frac{s \cdot p_4}{s^2 + p_3 s + p_1 \cdot p_2}}_{:=G_2, DT_2 \text{ behavior}} Z. \quad (9.48)$$

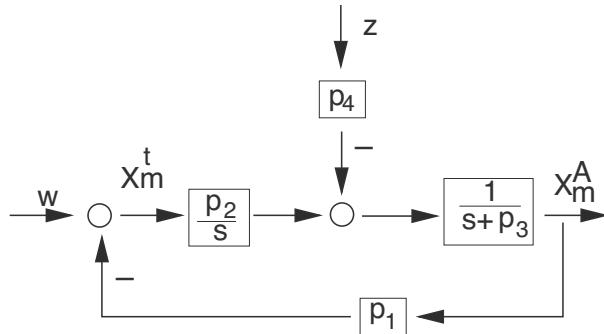


FIGURE 9.13: Block diagram of the control loop. Both differential equations can be represented in the Laplace domain as a coupling of a controller with a controlled system. The disturbance affects the controlled system.

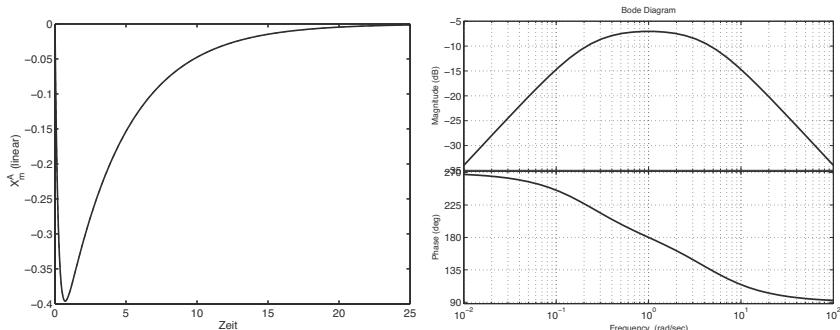


FIGURE 9.14: Step response and frequency response of the DT₂ transfer function.

An analysis of the equations shows that in the stationary case, the disturbance is completely compensated for. For the disturbance behavior, one determines a cascade setup consisting of a differentiator and a filter:

$$\begin{aligned} G'_2 &= -s \cdot p_4 \\ G''_2 &= \frac{1}{s^2 + p_3 \cdot s + p_1 \cdot p_2}. \end{aligned} \quad (9.49)$$

Based on the equations some statements can be made regarding the parameters: As the signal must be transferred, a sufficiently large maximal value is to be chosen:

$$p_4 \gg 1. \quad (9.50)$$

The system is not allowed to oscillate; therefore the characteristic equation is

not allowed to have eigenvalues with complex conjugates:

$$s_{1/2} = -\frac{p_3}{2} \pm \sqrt{\frac{p_3^2}{4} - p_1 \cdot p_2} \quad \rightarrow \quad \frac{p_3^2}{4} > p_1 \cdot p_2. \quad (9.51)$$

The system should not be too slow to adapt; thus, the eigenvalues are not allowed to lie far from each other. Figure 9.14 shows the step and frequency responses for an optimal set of parameters. The example shows that control engineering analysis proposes a number of constraints for kinetic parameters; this facilitates estimating the parameters based on concrete measurements.

9.4 Robust Control

An interesting application of control engineering concepts in systems biology is robust control. Here, in the design of the control loop parameter uncertainties as well as structural uncertainties can be considered. Starting point for the analysis is a linear model description of the system. Having a nonlinear model, the linear form can be obtained by linearization of model equations. Furthermore, as used in the last chapter, the system is Laplace transformed. It is now sought to put the system into the form shown in Figure 9.15. The uncertainties are summed up in block Δ . In this block uncertainties in the parameters or in the network structure are summarized. If one could set a limit for the uncertainties, for example, by using a scaling scheme, according to theoretical results one obtains a condition for the stability of the closed loop.

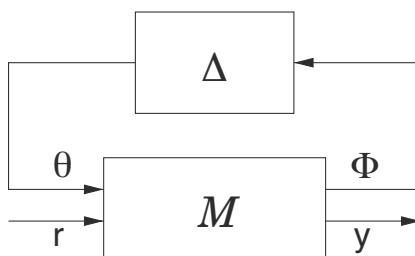


FIGURE 9.15: The $M-\Delta$ structure for considering uncertainties. Variables r and y are the system's input and output, respectively. The description of the uncertainties is via variables Φ and Θ .

Transfer functions in M depend on the structure of the control loop, choice of measurements and the control input. The approach for analyzing the system proceeds in analogy to the method used in classical controller design: Based on the results from the open loop the behavior of the closed system is inferred.

A closed loop here denotes that the uncertainty Δ is not connected to the overall system shown in the Figure. For the open loop (not closed via Δ), one could determine individual transfer functions:

$$\begin{bmatrix} \Phi \\ y \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} \Theta \\ r \end{bmatrix}. \quad (9.52)$$

To determine the dynamics of the closed loop, the term $\Theta = \Delta \Phi$ is inserted in the previous equation to obtain:

$$\Phi = G M_{12} r \quad (9.53)$$

$$y = (M_{22} + M_{21} \Delta G M_{12}) r \quad (9.54)$$

where the transfer function G is:

$$G = (I - M_{11} \Delta)^{-1}. \quad (9.55)$$

A controller within M shall now stabilize the system so that one would expect the individual transfer function to be stable. The stability problem is then reduced to investigate the stability of G . The Small Gain Theorem is here of relevance.

Small Gain Theorem: Assuming no structural constraints for Δ , and the transfer function M_{11} is stable with a maximal gain $|M_{11}| < \gamma_M$; when the following is valid for the gain of Δ : $|\Delta| \leq \gamma_\Delta$, then the closed loop is stable with G : $|G| < 1$, if $\gamma_M \gamma_\Delta < 1$.

Hence γ_M and γ_Δ represent an upper limit, or, put differently, the maximal gains. The Small Gain Theorem delivers only a sufficient condition. For a SISO (single input, single output) system, the theorem implies that the amplitude gains of M_{11} and Δ must lay below given limits. The difficulty in applying the theorem in systems biology is attributed to formulating the system in a way that it reflects the structure of Figure 9.15.

EXAMPLE 9.6 Lactose uptake

The approach shall be shown again using the example of lactose metabolism. Consider the signal flow diagram in Figure 9.16 showing the lactose induction. L represents the inducer of lactose that deactivates the repressor via the previously discussed non linear characteristics and hence activating the enzymatic synthesis of the protein Y representing protein LacY. An increased concentration of the enzyme leads to higher values of the inducer. In the given system form, a robust control structure can not be deduced. To proceed further, it is assumed that the sigmoid characteristics of Figure 9.7 can be approximated using non linear characteristics as in Figure 9.16.

Furthermore, it is assumed that the Hill coefficient n of the characteristics is known from measurements. Therefore, the system's unknowns in our example are the values of the upper and lower limits L_l and L_u of the intracellular

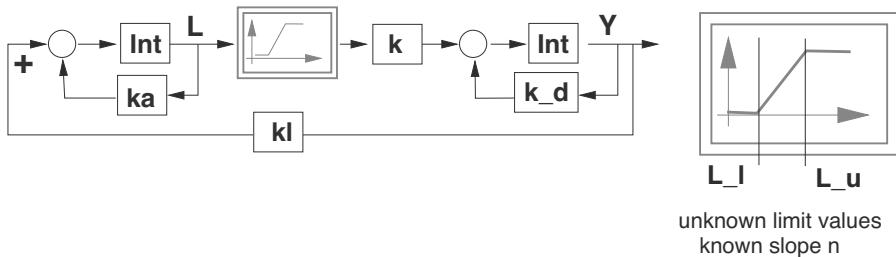


FIGURE 9.16: Scheme of lactose uptake in a signal flow diagram. Intracellular lactose (L) is the input in the non linear block. The output of the block is the inactive repressor $LacI$ that is responsible for the synthesis of Y ($LacY$). Y is in turn responsible for the transportation of extracellular lactose from outside to the inside, and it increases the lactose concentration.

lactose concentration where the switching occurs (see Figure). In contrast to the earlier model, regulator Crp is not taken into account. However, in addition to the system before, dynamics for the intracellular lactose is taken into account. The system with the non linear characteristics may now be formally transformed so that it matches the M- Δ structure which is the crucial part of the approach. The known information is to be extracted from the non linear characteristics (in the example the Hill coefficient n), and to be included in the description of the linear part. This is illustrated in Figure 9.17.

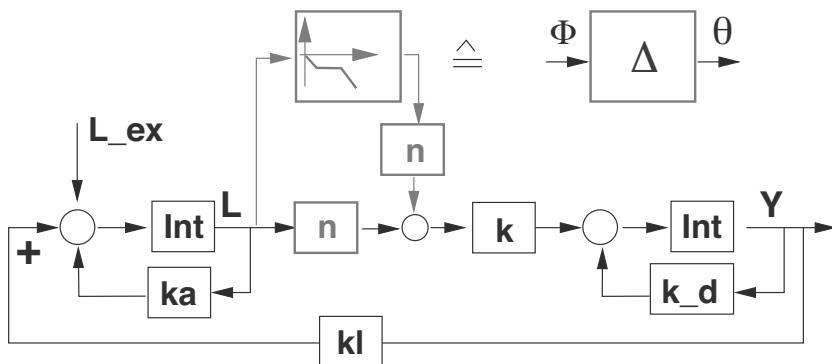


FIGURE 9.17: Schematic of lactose uptake in a signal flow diagram of the transformed non linear characteristics. The new characteristics correspond to the uncertainty description in the system. An additional input L_{ex} corresponds to the extracellular lactose concentration.

The characteristics from Figure 9.16 can be given as follows (to keep it simple for illustration $LacI$ is the inactive form of the repressor and not the

active one):

$$LacI = \begin{cases} n(L_u - L_l) & : L \geq L_u \\ n(L - L_l) & : L_l < L < L_u \\ 0 & : L \leq L_l. \end{cases}$$

The characteristics can be rewritten:

$$\begin{aligned} n \cdot L + \begin{cases} n(L_u - L_l) - n \cdot L & : L \geq L_u \\ -n \cdot L_l & : L_l < L < L_u \\ -n \cdot L & : L \leq L_l \end{cases} \\ n \cdot L + n \begin{cases} L_u - L_l - L & : L \geq L_u \\ -L_l & : L_l < L < L_u \\ -L & : L \leq L_l. \end{cases} \end{aligned} \quad (9.56)$$

The new characteristics according to Equation (9.56) can be found in Figure 9.17. It distinguishes itself (in magnitude) with a maximal signal gain of 1 (the new characteristics proceeds next with an increase -1) and with a constant range between the values L_l and L_u . The differential equations that describe the newly formulated system read (where the variables k_l , k and k_d are identical to the example above, k_a describes the degradation of Lac_{in}):

$$\begin{aligned} \dot{L} &= k_l Y + Lac_{ex} - k_a L \\ \dot{Y} &= k n (L + \Theta) - k_d Y \end{aligned} \quad (9.57)$$

where the term $n(L + \Theta)$ results from the nonlinear characteristics. Note that the feedback from Y to intracellular lactose L taking into account the influence of external lactose Lac_{ex} cannot be written as above as a product; instead we use the linear form by writing down a sum of the corresponding terms. Transforming the linear parts of the network using the Laplace transformation, in terms of transfer functions P_1 and P_2 one obtains:

$$\begin{aligned} L &= \Phi = P_1 Y + P'_1 Lac_{ex} \\ Y &= P_2 (\Phi + \Theta). \end{aligned} \quad (9.58)$$

The transfer function can be written as:

$$P_1 = \frac{k_l}{s + k_a}, \quad P'_1 = \frac{1}{s + k_a}, \quad P_2 = \frac{n k}{s + k_d}. \quad (9.59)$$

Inserting the transfer functions into each other, after some algebraic manipu-

lation we obtain:

$$\Phi = \underbrace{\frac{P_1 P_2}{1 - P_1 P_2}}_{M_{11}} \Theta + \underbrace{\frac{P'_1}{1 - P_1 P_2}}_{M_{12}} Lac_{ex} \quad (9.60)$$

$$Y = \underbrace{\frac{P_2}{1 - P_1 P_2}}_{M_{21}} \Theta + \underbrace{\frac{P'_1 P_2}{1 - P_1 P_2}}_{M_{22}} Lac_{ex} \quad (9.61)$$

and in matrix form (compare Figure 9.18):

$$\begin{pmatrix} \Phi \\ Y \end{pmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \cdot \begin{pmatrix} \Theta \\ Lac_{ex} \end{pmatrix}. \quad (9.62)$$

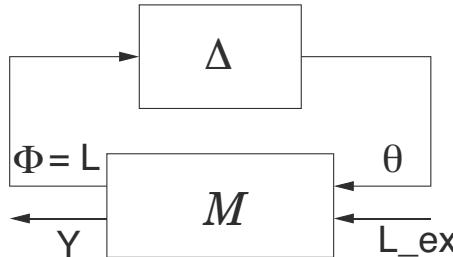


FIGURE 9.18: The M- Δ structure for the example of lactose uptake.

Of interest is now the analysis of M_{11} . One obtains:

$$M_{11} = \frac{n k_l k}{s^2 + (k_a + k_d) s + k_a k_d - n k_l k}. \quad (9.63)$$

According to the signs rule for having a stable behavior the following condition is obtained: $k_a k_d > n k_l k$. For the stationary values one obtains:

$$M_{11}(s \rightarrow 0) = \frac{n k_l k}{k_a k_d - n k_l k}. \quad (9.64)$$

This system is considered to be a PT₂ system, whose gain drops as the frequency is increased. The maximal gain is obtained for small frequencies and this yields the condition:

$$|M_{11}(s = 0)| < 1 \longrightarrow k_a k_d > 2 n k_l k. \quad (9.65)$$

The calculation shows a key condition, that the product of the synthesis rate multiplied by 2 n must be smaller than the product of the degradation rates. The system can then have for the uncertainty block an arbitrary model with a gain factor < 1 and the system is stable.

Exercises

EXERCISE 9.1 Monotone systems.

For monotone systems, it is required that the sign in the Jacobian matrix is fixed. Consider the following system:

$$A \xrightleftharpoons{r_1, I_1} B \xrightleftharpoons{r_2, I_2} C. \quad (9.66)$$

There are two independent modifiers I_1 and I_2 . The reaction equations are given as follows:

$$\begin{aligned} r_1 &= k_1 (I_1 A - K_1 B) \\ r_2 &= k_2 (I_2 B - K_2 C). \end{aligned} \quad (9.67)$$

The metabolites are related by a conservation equation:

$$N = A + B + C. \quad (9.68)$$

Based on the conservation equation, the system could be reduced by one equation. What would be the best choice with respect to a further analysis on monotony?

SOLUTION:

A prerequisite for checking monotone systems is that the sign of the Jacobian is fixed. For the given system we start by setting up the differential equations of the system:

$$\begin{aligned} \dot{A} &= -r_1 \\ \dot{B} &= r_1 - r_2 \\ \dot{C} &= r_2. \end{aligned} \quad (9.69)$$

In the first case, we eliminate A and we get:

$$\begin{aligned} \dot{B} &= r_1 - r_2 \\ &= k_1 (I_1 (N - B - C) - K_1 B) - k_2 (I_2 B - K_2 C) \\ &= k_1 I_1 N - (k_1 I_1 + k_1 K_1 + k_2 I_2) B - (k_1 I_1 - k_2 K_2) C \\ \dot{C} &= r_2 = k_2 (I_2 B - K_2 C). \end{aligned} \quad (9.70)$$

The entries in the Jacobian are the terms in parenthesis before B and C . We note that the entry for C in the equation of B might change the sign if the value of I_1 changes. Now we consider a second case where $B = N - A - C$ and we get:

$$\begin{aligned}\dot{A} &= -k_1(I_1 A - K_1(N - A - C)) \\ &= -k_1(I_1 + K_1)A - k_1 K_1 C \\ \dot{C} &= k_2(I_2(N - A - C) - K_2 C) \\ &= k_2 I_2 N - k_2 I_2 A - (k_2 I_2 + K_2)C.\end{aligned}\tag{9.71}$$

Now the situation is different: the terms before A and C have fixed entries with respect to the sign. To summarize our findings in the second case, we see the following dependencies:

$$\begin{aligned}A &\xrightarrow{I_2, -} C \\ C &\xrightarrow{-} A.\end{aligned}\tag{9.72}$$

EXERCISE 9.2 Robust control.

Perform simulation studies to verify the conditions for robustness in the text. What do you observe for the steady solution for different values of external lactose?

Bibliography

- [1] E. D. Sontag. Some new directions in control theory inspired by systems biology. *Systems Biology* 1, 2004, 9-18.
- [2] T.-M. Yi et al. Robust perfect adaption in bacterial chemotaxis through integral feedback control. *PNAS* 97, 2000, 4649-4653.
- [3] M. E. Csete and J. C. Doyle. Reverse engineering of biological complexity. *Science* 295, 2002, 1664-1669.
- [4] U. Alon. *An Introduction to Systems Biology*. Chapman & Hall/CRC, 2007
- [5] N. Baarkai and S. Leibler. Robustness in simple biochemical networks. *Nature* 387, 1997, 913-917.

This page intentionally left blank

Chapter 10

Motifs in Cellular Networks

An analysis of various biochemical networks in recent years has revealed the existence of several recurrent patterns which are more frequent than others; that is, these patterns occur more frequently in real networks than in artificial networks with random connections. These patterns are referred to as motifs. Figure 10.1 illustrates some recurrent patterns identified in the transcriptional network of *E. coli* [1].

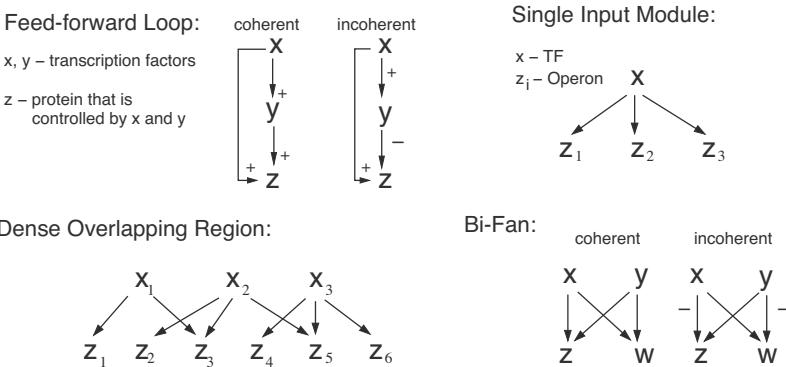


FIGURE 10.1: Motifs in the regulatory network of *E. coli*.

The network under study comprises 424 nodes and 519 edges. For example, 40 feed-forward loops have been discovered in this particular network; a network with a random connection structure should theoretically comprise only 7 ± 3 feed-forward loops. 203 occurrences of the Bi-Fan motif were detected in the real network as compared to the expected 47 ± 12 occurrences in a random network [2]. The Single Input Module describes a simple connection of one transcriptional factor X with several operons Z_i . If multiple operons Z_i are affected by more than one transcription factor simultaneously, the motif is termed Dense Overlapping Region; however, there is no single, unambiguous pattern for this motif.

Cellular systems in general are characterized by the fact that they very quickly can adapt to new environmental conditions. To realize this, a powerful “apparatus” of signal transduction units and actuators as discussed in the last chapters is established. Changes in environmental conditions lead to different growth rates μ of the organisms and the question arises as to which

growth rate may be achieved and in which way the organism realizes this. The theoretical prediction of growth rates still represents a big challenge in theoretical studies; in many cases, it can be said that cells grow after a specific change, for example, by switching off genes; however, the exact numerical value is hardly predictable. On the other hand it is observed that certain intracellular state variables change little or not at all if environmental conditions are changed, when a mutation is inserted, or a protein is highly overexpressed. The observation that certain network properties are not sensitive but robust was already shown in a number of examples. Such network properties are realized with specific interconnections between the involved components. Although the term “motif” as said above is used often for pattern that appear frequently in networks, the term should also be used in a broader sense to cover network circuits with a specific functionality. To guarantee such a functionality such networks are also often robust in the sense that small perturbations do not change the input/output characteristics.

Below, some motifs with their functionalities will be presented and their properties are described. A typical functionality represents a switch behavior, as described in Figure 10.2.

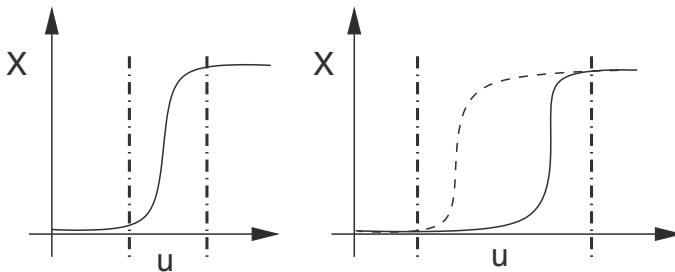


FIGURE 10.2: Switch: Sigmoid curve (left) and bistable system (right).

The left plot shows a sigmoid curve with a large slope in the inflection point. The input u can be an external stimulus, but also an intracellular compound. In this case the question raises which processes influence the concentration of u , and how is it ensured that u has no big fluctuations. The dashed area in the figure indicates the range of u so that the output X could be on or off. If u is always smaller (or larger), then the system is either always off (or always on).

The right plot shows a bistable system (see also other chapters). Here the output “jumps” if a certain value of the input u is exceeded. The reason for this is that the structure of the system and the kinetic parameters allow several solutions (stable or unstable) for the steady state — in the plot the solid line shows the steady state for increasing values of u while the dashed line shows the steady state solution for decreasing values of u . Out of this region there is only one (stable) solution, which is then achieved. Again, it is

important that values of u are adjusted in the specified area; otherwise this behavior cannot be observed.

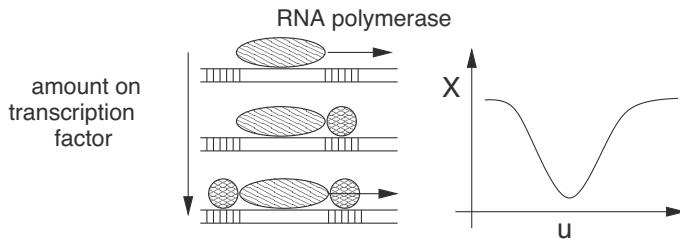


FIGURE 10.3: Biphasic behavior: A DNA binding site binds RNA polymerase and a transcription factor (left). Transcription only starts if none or both binding sites are occupied with transcription factors. For low and high values of the transcription factor the system is turned on; otherwise it is off.

Figure 10.3 shows a biphasic system that shows different regions of activation and inhibition. This can be realized if the binding sites possess different affinities for the transcription factor. In the case shown, for high and low values of the transcription factor the system is on; otherwise it is off. Now we switch again to the feed-forward motif. This motif appears in genetic, metabolic and signaling networks and therefore represents one of the most interesting circuits.

10.1 Feed-forward loop (FFL)

The feed-forward loop is the most frequently occurring motif in the transcriptional network of *E. coli*. As can be seen in Figure 10.4, there exist two variants: the coherent and the incoherent loop. Coherence implies that the sign parity (number of edges with a negative sign) of the direct path from X to Z equals the sign parity of the path from X via Y to Z (minus times minus yields plus). The figure illustrates the 8 possible cases.

To investigate the structural and dynamic characteristics of a FFL, we focus on the coherent type 1 (top row of the figure, leftmost) and on the incoherent type 1 (bottom row of the figure, leftmost) as described in [3]. It is assumed that in each case an inducer I_X and I_Y is required to activate the respective transcription factors. Hence, the activity of the transcription factors is described by the following relations: $X^* = X$ if $I_X = 1$ and $Y^* = Y$, if $I_Y = 1$. There are two cases: In order to activate transcription factor Z , the two transcription factors X and Y can be connected by either an AND or an OR gate. The AND gate requires both proteins to be present in order to

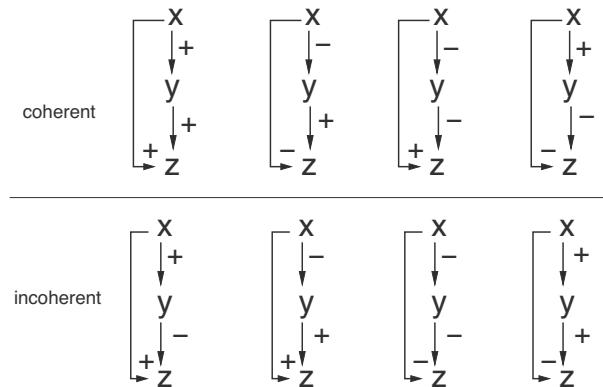


FIGURE 10.4: Top row: Coherent feed-forward loop motifs. Bottom row: Incoherent feed-forward loop motifs

initiate transcription, whereas the presence of one of the factors is sufficient in case of the OR gate (for a detailed description and derivation of the equations, see chapter on polymerization). Figure 10.5 illustrates a mechanistic view of these interactions.

Plot **A** illustrates the AND gate. The two DNA binding sites of the two regulators do not overlap. However, both transcription factors are necessary for the readout of the genetic information. In case **B**, which represents the OR gate, it is assumed that the two factors compete for the same binding site. Due to the competition for one binding site, X and Y act as mutual inhibitors, as can be seen in the corresponding equations. In the simulation study presented below, the AND and OR gate models are contrasted with a third model. Here, the two transcription factors affect transcription by a combined AND/OR gate (case **C**).

The model equations are summarized in the following. Protein X is regarded as a (constant) input signal so that one obtains two differential equations to describe the dynamics of Y and Z :

AND gate

$$\begin{aligned}\dot{Y} &= k_{bY} + k_Y f_1^u(X^*) - k_{dY} Y \\ \dot{Z} &= k_{bZ} + k_Z f_1^u(X^*) f_2^u(Y^*) - k_{dZ} Z.\end{aligned}\quad (10.1)$$

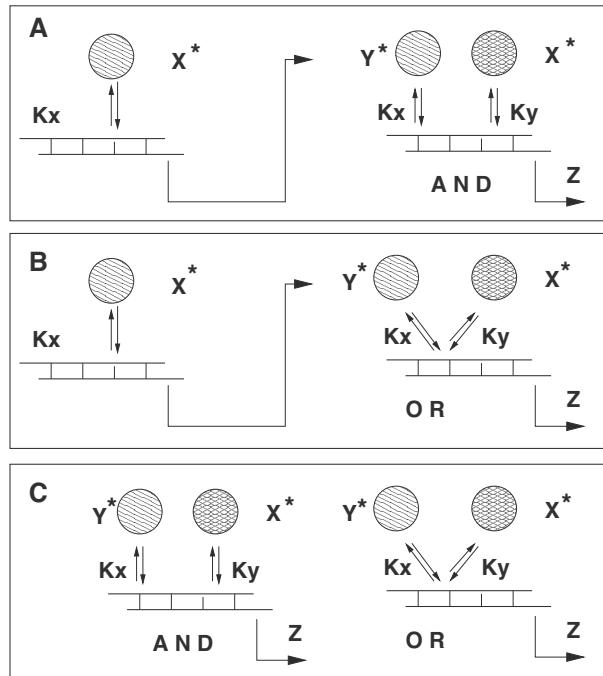


FIGURE 10.5: **A** FFL with an AND gate. **B** FFL with an OR gate. **C** Model without FFL.

OR gate

$$\begin{aligned}\dot{Y} &= k_{bY} + k_Y f_1^u(X^*) - k_{dY} Y \\ \dot{Z} &= k_{bZ} + k_Z (f_1^0(X^*, K_X, K_Y, Y^*) + f_2^0(Y^*, K_X, K_Y, X^*)) - k_{dZ} Z.\end{aligned}\quad (10.2)$$

The corresponding kinetics are given next. For the AND gate:

$$f_1^u = \frac{\left(\frac{X^*}{K_X}\right)^{n_1}}{1 + \left(\frac{X^*}{K_X}\right)^{n_1}}, \quad f_2^u = \frac{\left(\frac{Y^*}{K_Y}\right)^{n_2}}{1 + \left(\frac{Y^*}{K_Y}\right)^{n_2}}; \quad (10.3)$$

for the OR gate:

$$f_1^0 = \frac{\left(\frac{X^*}{K_X}\right)^{n_1}}{1 + \left(\frac{X^*}{K_X}\right)^{n_1} + \left(\frac{Y^*}{K_Y}\right)^{n_2}}, \quad f_2^0 = \frac{\left(\frac{Y^*}{K_Y}\right)^{n_2}}{1 + \left(\frac{X^*}{K_X}\right)^{n_1} + \left(\frac{Y^*}{K_Y}\right)^{n_2}}. \quad (10.4)$$

The inhibition of the incoherent loop can be formulated as follows:

$$f^0 = \frac{1}{1 + \left(\frac{X^*}{K_X}\right)^{n_1} + \left(\frac{Y^*}{K_Y}\right)^{n_2}}. \quad (10.5)$$

10.1.1 Logic switching pattern

The logic switching pattern for the coherent case is listed in the following table. If X is inactive ($I_X = 0 \rightarrow X^* = 0$) then Z can not be formed; if Y is inactive ($I_Y = 0 \rightarrow Y^* = 0$) then Z can be formed only when X is active and an OR gate is present.

$$\begin{aligned} X^* = 1, \quad Y^* = 1 &\rightarrow Z \text{ active} \\ X^* = 1, \quad Y^* = 0 &\rightarrow Z \text{ active (OR)}, \quad Z \text{ inactive (AND)} \\ X^* = 0, \quad Y^* = 1 &\rightarrow Z \text{ inactive} \\ X^* = 0, \quad Y^* = 0 &\rightarrow Z \text{ inactive}. \end{aligned} \quad (10.6)$$

In the incoherent case the following table applies (only the AND gate is considered):

$$\begin{aligned} X^* = 1, \quad Y^* = 1 &\rightarrow Z \text{ inactive} \\ X^* = 1, \quad Y^* = 0 &\rightarrow Z \text{ active} \\ X^* = 0, \quad Y^* = 1 &\rightarrow Z \text{ inactive} \\ X^* = 0, \quad Y^* = 0 &\rightarrow Z \text{ inactive}. \end{aligned} \quad (10.7)$$

In summary, the following logical operations arise: Coherent FFL with an AND gate: $X^* \wedge Y^*$; Coherent FFL with an OR gate: X^* ; and Incoherent FFL with an AND gate: $X^* \wedge \text{NOT}Y^*$.

10.1.2 Dynamic behavior

In the following figures, simulations of the previous cases are shown. Moreover, Figure 10.5 displays the simulations of a reference model in which the component Z is directly activated by X and Y . In order to obtain comparable results, the different models were simulated using the same set of kinetic parameters.

The left panel of Figure 10.6 shows the time courses of Y and Z of the AND gate, as well as the time response of Z of the reference model. At time $t = 0$, the input signal X was chosen as $X(0) = 1$ and at $t = 5$ (indicated by the vertical line in the figure) X was set to $X(5) = 0$ (which corresponds to switching off the system). For the chosen set of parameters, the time course of Z of the AND FFL model is delayed in comparison to Z of the reference model. The parameters were chosen such that the stationary points of Z are

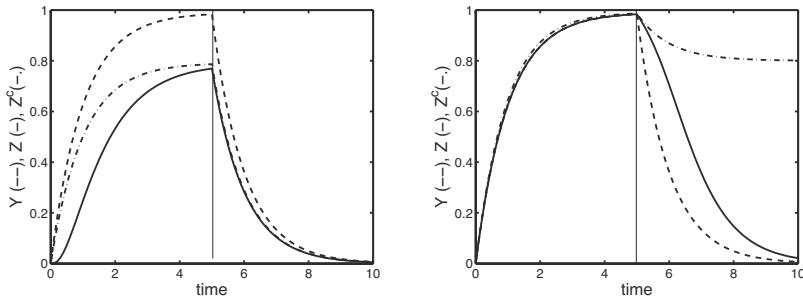


FIGURE 10.6: Left: Simulation study of a coherent FFL with an AND gate (Z : solid line, Y : dashed line, Z^c of reference system: dash-dotted line).

the same for both models. Both systems display the same behavior after being switched off at $t = 5$: Both Z curves decrease exponentially to zero. Consider the OR gate, which is shown on the right side of the figure to illustrate the differences. As either X or Y is required for activation, the switch-on dynamics are the same. However, a difference can be seen after the system is switched off. For the reference system, the second activator is still present, hence there is only a slight change in the concentration of Z . The Y variable decreases exponentially as X is switched off. The effect of switching off for the Z variable with an OR gate is delayed, as Y is still present and a synthesis can still take place. Only when there is hardly any Y left, the Z variable then returns to the initial value.

Figure 10.7 shows a simulation for the incoherent case in which Y acts as an inhibitor. Only the AND gate is considered here, where the basal level of the synthesis of Y is increased in the right figure. As can be seen, this structure acts as a pulse generator: Z is formed first and then the inhibition is reduced further by Y . The magnitude of the pulse is determined by the basal synthesis level, which is higher in the right plot.

10.2 Feed-forward Loops in Metabolic Networks

The possible structures of feed-forward loops in metabolic networks look somewhat different. As can be seen from Figure 10.8, components can be linked by fluxes or signals, where one or two signals are possible. If two signals are present, there are two linear paths; in the other case, the components are coupled by divergent or convergent elements. The network shown in **A** represents a structure that is found in glycolysis in *E. coli* and represents

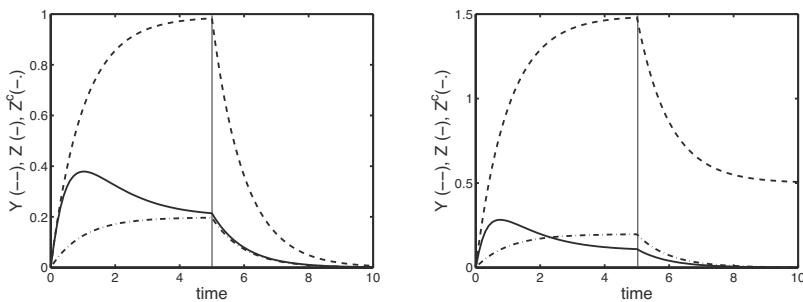


FIGURE 10.7: Left: Simulation of the AND gate of the incoherent FFL (Z : solid line, Y : dashed line, Z^c of the reference system: dash-dotted line). Right: Simulation of the AND gate of the incoherent FFL (Z : solid line, Y : dashed line, Z^c of the reference system: dash-dotted line).

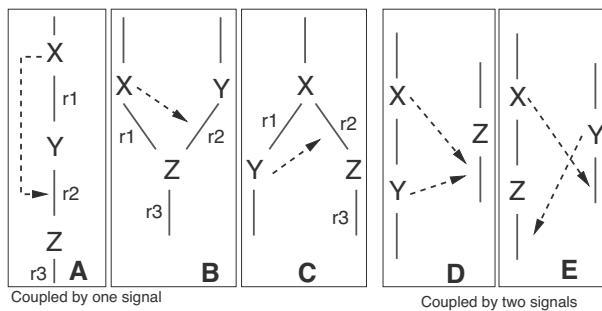


FIGURE 10.8: Feed-forward loop motifs in metabolic networks. **A-C** variants with two flux interconnections and a single signal interconnection. **D-E** variants with a single flux interconnection and two signal interconnections.

an interesting case study for the interrelation between control of substrate uptake, metabolism and signaling.

10.2.1 Carbohydrate uptake and control

The bacterium *Escherichia coli* has numerous uptake systems which allow the uptake of various substrates from the medium [4, 5]. An uptake system describes the single proteins with their regulation which can be rather complex in some cases. For carbohydrates, the uptake systems comprise an integral membrane protein, which is responsible for the actual uptake, as well as other proteins which cause a modification of the substrate, e.g., phosphorylation. The transportation systems are usually specific and therefore have a rather small substrate spectrum. Therefore, almost all substrates have an individual system for uptake. Since it would be economically nonsensical if the cell provided all systems at all times, the systems are only provided, if the specific substrate is present in the medium. If a substrate is provided in a batch-

culture, a constant growth rate μ is reached during exponential growth. The growth rate varies strongly for different carbohydrates which means that the cell can use some sugars better than others. From molecular biological research, it is known that a regulatory protein is significantly important for the synthesis of a transport system and ensures that the according proteins are expressed if they are needed. Since this regulator is involved in gene expression of a high number of genes, it is referred to as a global regulator.

The global transcription factor Crp is involved in the initiation of the transcription of a multitude of catabolic genes. The activation of the transcription factor depends on the rate of phosphorylation of the component EIIA of the phosphotransferase system (PTS). Experimental research has shown that there is a connection between the growth rate of *E. coli* and the degree of phosphorylation of EIIA. The relation can be plotted as a characteristic curve: At high growth rates, the degree of phosphorylation is low, while at low growth rates it is high (Figure 10.9). Accordingly Crp is activated only if the growth rate is low and is inactive if the growth rate is high; note, however, that there will be no switching; instead a gradual relationship is observed.

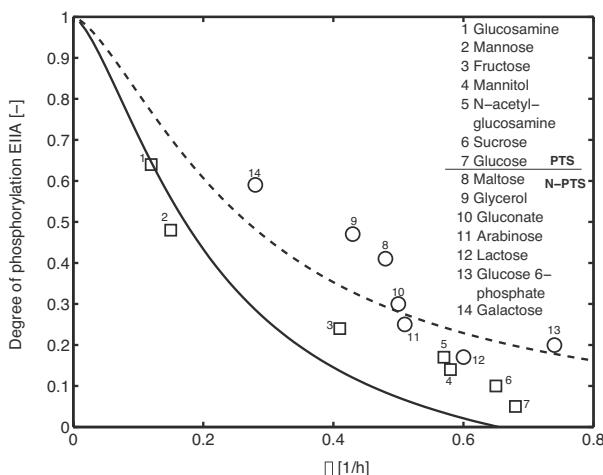


FIGURE 10.9: Comparison of experimental data with simulations. PTS sugars are plotted with a solid line; nonPTS sugars are plotted with a dotted line [7].

With help of a mathematical model we will try to better understand the system; moreover, it will be demonstrated that the system is structurally robust: Changes in parameters or the structure of the model have only a small influence on the behavior of the system. In the following, the connection between the specific growth rate μ and the output of the PTS shall be illustrated. Therefore, a simplified reaction scheme as shown in Figure 10.10 is used [6, 7]. As described in Biological Fundamentals the PTS is a transport and a signal-

ing system at the same time. To distinguish transport activities, we denote carbohydrates taken up by the PTS as PTS carbohydrates and consequently the PTS is active; the other case is the reverse, with nonPTS carbohydrate and inactive PTS, respectively. In the case of an inactive PTS, for example, when the cells grow on lactose, glycerin or glucose 6-phosphate, the rate of the reversible PTS-reaction r_{pts} must equal 0 since the phosphoryl group is not further consumed.

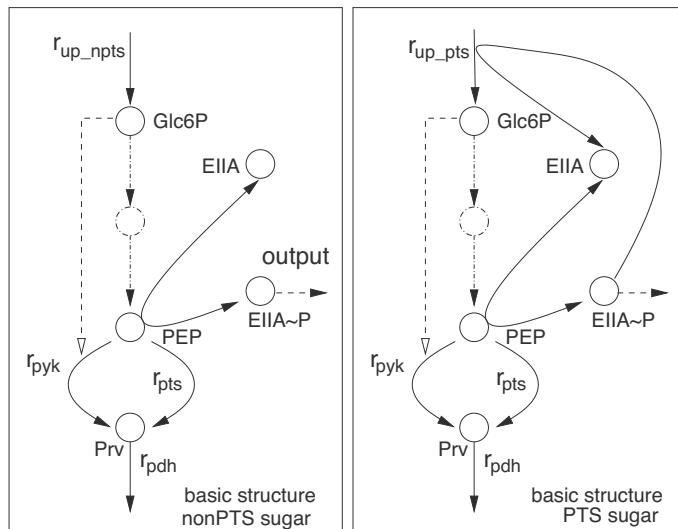


FIGURE 10.10: Basic structure of the model. The PTS proteins are only represented by the component EIIA. Some reactions of the glycolysis are merged since their outflow to biosynthesis is marginal. Only such cases are considered, where substrates enter the cell via glucose-6-phosphate. However, this restriction has no consequences on the general case. A feed-forward loop activates the pyruvate-kinase (in the figure, glucose-6-phosphate is shown as a metabolite, which serves as an activator; the actual activator is fructose-1,6-bisphosphate). Left: Structure for nonPTS carbohydrates; right: Structure for PTS sugars.

Assuming a simple reaction with mass action kinetics we get:

$$r_{pts} = k_{pts} (PEP \cdot EIIA - K_{PTS} Prv \cdot EIIAP) \stackrel{!}{=} 0 \quad (10.8)$$

with Prv standing for pyruvate and PEP for phosphoenolpyruvate. As indicated in the scheme, for the PTS, only component $EIIAP$ is considered as output:

$$EIIA = EIIA_0 - EIIAP. \quad (10.9)$$

Thus, the relation between the ratio of PEP/pyruvate and the degree of phosphorylation of the component EIIA is:

$$EIIAP = EIIA_0 \frac{\frac{PEP}{Prv}}{K_{PTS} + \frac{PEP}{Prv}}. \quad (10.10)$$

In the case of an active PTS the uptake rate of the sugar $r_{up/pts}$ is equal to the rate through the PTS. In this case we get:

$$EIIAP = \frac{EIIA_0 \frac{PEP}{Prv} - \frac{r_{up/pts}}{k_{pts}}}{K_{PTS} + \frac{PEP}{Prv}}. \quad (10.11)$$

Both equations show a similar structure, which is central for further considerations: To obtain the experimentally observed low degree of phosphorylation for high growth rates, the PEP/pyruvate ratio must be kept lower for high growth rates (Figure 10.11 upper row). From the equation for EIIAP a constraint follows, that is, that the derivative with respect to the growth rate must be negative. If we do so we get:

$$\frac{dEIIAP}{d\mu} < 0 \rightarrow \frac{dPEP(\mu)}{d\mu} < \frac{dPrv(\mu)}{d\mu} c \quad (10.12)$$

with a constant positive term c .

In order to realize the desired behavior, several possibilities, how PEP and pyruvate have to behave as function of the growth rate, arise. The possible curves are shown in Figure 10.11. It can be seen easily that if both functions increase monotonically with respect to the growth rate, as in the plot lower left, the behavior is quite sensitive: Small changes or disturbances of the system can lead to strong changes and, in extreme cases, an inversion of the ratio. By contrast, the case shown at the plot lower right of Figure 10.11 results in a robust structure: Since the slope of PEP is always negative, the condition above is always fulfilled.

In the living cell the desired robust structure can be realized by means of the feed-forward control, which is shown in Figure 10.10. A high growth rate and therefore a high flux through glycolysis also requires a high flux through the pyruvate kinase. However, if the PEP concentration is to decrease with an increasing growth rate, this must be compensated since the rate could not be maintained otherwise. The components of the upper half of glycolysis can now act as a signal and cause a high rate by activating the pyruvate kinase since the concentrations increase with an rising growth rate. In fact, it is already known for a long time from tests with isolated pyruvate kinase that the enzyme is strongly activated by fructose 1,6-bisphosphate (in Figure 10.10, fructose 1,6-bisphosphate is not explicitly shown, but is replaced by glucose-6-phosphate as a substitute).

In the following we use a mathematical model to verify our considerations. A simple, stationary model, which includes the feed-forward loop, can be

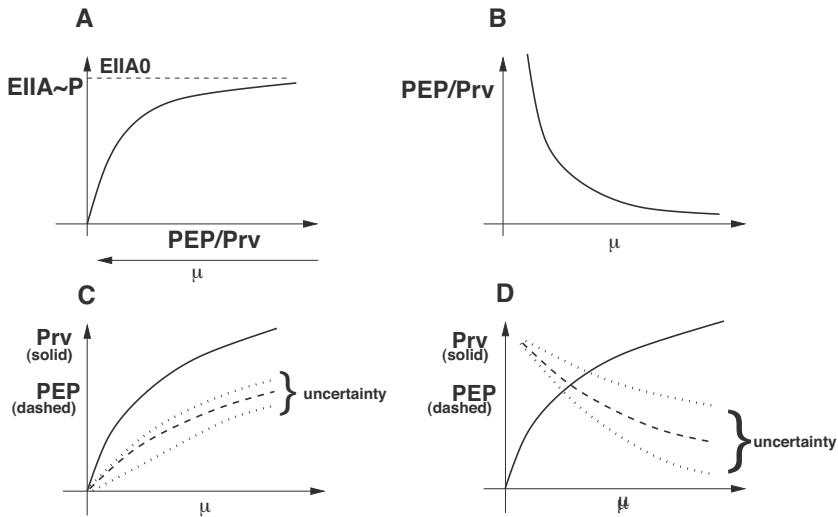


FIGURE 10.11: Upper left: According to Equation (10.10) a high PEP/pyruvate ratio and therefore a high rate of phosphorylation arises for low growth rates. Upper right: PEP/pyruvate ratio as a function of the growth rate. Lower left: PEP and pyruvate as a function of the growth rate. Sensitive structure since uncertainties easily lead to strong shifts in the ratio. Lower right: PEP and pyruvate as a function of the growth rate in a controlled network. Robust structure, since uncertainties barely lead to shifts in the ratio.

formulated for the case that a nonPTS-sugar is taken up (rate $r_{up/pts}$). The differential equations for metabolite glucose-6-phosphate, PEP and pyruvate are:

$$\begin{aligned} \text{Glc-6-P: } G\dot{6}P &= 0 = r_{up/pts} - r_{gly} \\ \text{PEP: } P\dot{E}P &= 0 = 2r_{gly} - r_{pyk} - r_{pts} \\ \text{pyruvate: } P\dot{r}v &= 0 = r_{pyk} + r_{pts} - r_{pdh} \end{aligned} \quad (10.13)$$

with the kinetic expressions:

$$r_{gly} = k_{gly} G6P^\gamma; r_{pyk} = k_{pyk} G6P^\alpha PEP^\delta; r_{pdh} = k_{pdh} Prv^\nu, \quad (10.14)$$

with α representing the influence of the feed-forward activation. Model input is rate $r_{up/pts}$ representing the growth rate. For these power law kinetics,

analytic expression for the steadystates are easily derived:

$$\begin{aligned} G6P &= \left(\frac{r_{up/npts}}{k_{gly}} \right)^{\frac{1}{\gamma}}, \quad PEP = \left(\frac{2}{k_{pyk}} \right)^{\frac{1}{\delta}} \left(\frac{r_{up/npts}}{k_{gly}} \right)^{\frac{\gamma-\alpha}{\gamma\delta}} \\ Prv &= \left(\frac{2 r_{up/npts}}{k_{pdh}} \right)^{\frac{1}{\nu}}. \end{aligned} \quad (10.15)$$

Here we see that for $\alpha > \gamma$ PEP is decreasing in dependence on the input. In the nonPTS case with a flux vector given by $(1, 2, 2, 0)^T$ the steady-state output response is thus of the form

$$EIIA^P = \frac{EIIA_0}{1 + K_{pts} \frac{Prv}{PEP}} = \frac{EIIA_0}{1 + \text{const.} r_{up/npts}^{(1/\nu) - (\gamma - \alpha)/(\gamma\delta)}}. \quad (10.16)$$

Here the condition to have a negative slope reads:

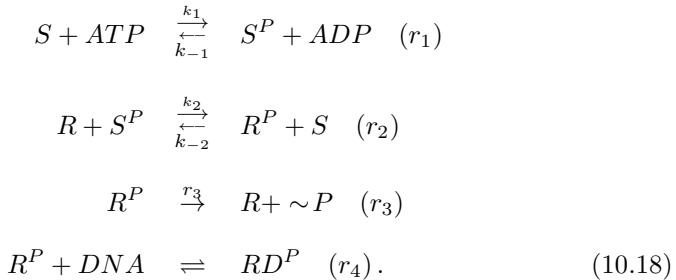
$$\frac{\delta}{\nu} + \frac{\alpha}{\gamma} \geq 1. \quad (10.17)$$

Here, we clearly see the influence of the kinetic parameter of the feed-forward activation α . A large value of this parameter guarantees that the condition of a negative slope is fulfilled; hence the FFL structure in metabolism increases robustness.

10.3 FFL in Signaling Systems: Two-component Signal Transduction

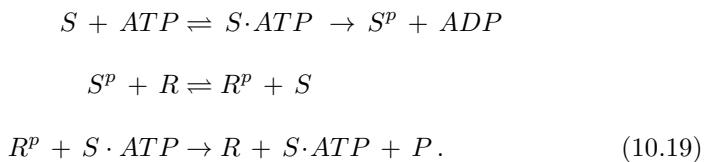
Many bacterial systems possess a simple signal transduction system: The two-component system (see Biological Fundamentals). A sensor kinase S is phosphorylated by ATP where the kinase acts as an enzyme which phosphorylates itself. In a second step, the phosphoryl group is transferred to the response regulator R . The third step ensures that the system can be switched off again (dephosphorylation). Here the phosphate group $\sim P$ is released. The fourth step describes the binding for the regulator to the DNA binding site. The binding of the regulator R to the DNA allows starting transcription of

the respective genes. The following reaction scheme can be derived [8]:



Since it is not exactly known how the stimulus influences the system, $k_1 = k_1(I)$ can be taken as a function of the stimulus I in the model. Interesting variations of the model arise from considering rate r_3 . If a phosphatase Ph is involved, one obtains for mass action kinetics $r_3 = k_3 \cdot R^P \cdot Ph$; however, it is also known that the sensor kinase has the ability to dephosphorylate the response regulator. The bifunctionality of the sensor has an influence on the signal response curve (see below).

In current publications [9, 10] the two-component system was structurally analyzed and characteristic properties were revealed. Thereby, the influence of the total amounts of the sensor and the regulator play an important role. Even though both models are formulated differently, they can be transformed into the same structure. The second reaction scheme given now includes an intermediate complex consisting of sensor and ATP, and can be formulated as follows:



The system is now shown in Figure 10.12, where S^* now represents a specific form of the sensor S (for example with or without ATP). Note, that it is important to see that the first reaction — the activation of the sensor — is irreversible in the model. Based on the shown scheme, the respective equations shall be used to determine the input/output behavior.

We obtain the following two differential equations for S^P and R^P :

$$\begin{aligned}
 \dot{S}^P &= r_1 - r_2 \\
 \dot{R}^P &= r_2 - r_3.
 \end{aligned} \tag{10.20}$$

In the steady state $r_1 = r_2 = r_3$. If this relation is used to calculate output



FIGURE 10.12: General structure of the two-component system. In the illustration, a forward loop is shown, which is necessary to yield a robust behavior with respect to the total amount of sensor and regulator.

R^P we get for mass action kinetics:

$$k_1(I)S^* = k_3 S^* R^P \quad (10.21)$$

$$\longrightarrow R^P = \frac{k_1(I)}{k_3} \neq f(S_0, R_0). \quad (10.22)$$

Therefore, one obtains the interesting result that the degree of phosphorylation of the regulatory protein (the output of the system) is independent of the total amount of sensor and regulator; this property is again referred to as robustness. The robustness was verified experimentally (see literature from above). Note, however, that the kinetic order of S^* in r_1 and r_3 must be the same.

Now we compare two model structures taking into account that the newly synthesized protein P has an influence on the two-component system. The model is extended by two differential equations for mRNA and protein. The results are shown in Figure 10.13. The solid lines show the results for phosphorylated sensor and phosphorylated regulator. The dashed lines show a model variant where the rate of dephosphorylation r_3 is modified taking into account the unphosphorylated sensor as bifunctional enzyme and the protein P that acts as an activator; in this case the feedback is realized by increasing the rate of dephosphorylation:

$$r_3 = k_3 R^P S f(P) = k_3 R^P S \frac{P}{K_I + P}. \quad (10.23)$$

The modifications have a strong influence on the stimulus response curve. The degree of phosphorylation is higher and the system shows a switch characteristic.

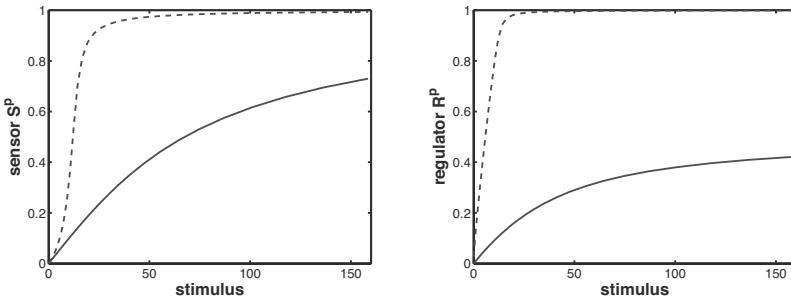


FIGURE 10.13: Signal response curve of the two-component system: left sensor S^p and right regulator R^p in dependence on a stimulus. The solid lines show results for the standard model. The dashed lines result from two model variations in the rate of dephosphorylation r_3 .

10.4 Further Signaling Motifs

10.4.1 Quorum sensing system

Quorum sensing describes the ability of many bacteria to measure their cell density. This ability allows for a cell density-dependent activation of metabolic pathways and the expression of certain operons. These pathways are often linked to the production of biofilms or competence (ability to take up external DNA). These processes are usually unimportant for a single cell so they are only activated if the cell population has reached a critical size. Quorum sensing is transmitted through signaling molecules, so called autoinducers, which are produced inside the cell and are then excreted into the medium. Sometimes these molecules are modified during this process. The cell possesses systems that can measure the concentration of substances in the medium and then trigger a signal response. The response includes control of transcription of operons, which are involved in the production of biofilms or competence.

Figure 10.14 shows a simplified scheme of quorum sensing. An autoinducer A is synthesized and then released into the medium (rate r_a). A sensing system measures the concentration and activates its own synthesis (rate r_s). For model formulation, we assume that the biomass is an input parameter and that the autoinducer is subject to degradation.

Model equations for the components A and A_{ex} are as follows with constant biomass B :

$$\begin{aligned}\dot{A} &= r_s - r_a = k_s \frac{A_{ex}}{A_{ex} + K} - k_a A \\ \dot{A}_{ex} &= r_a B - k_d A_{ex} = k_a A B - k_d A_{ex}.\end{aligned}\tag{10.24}$$

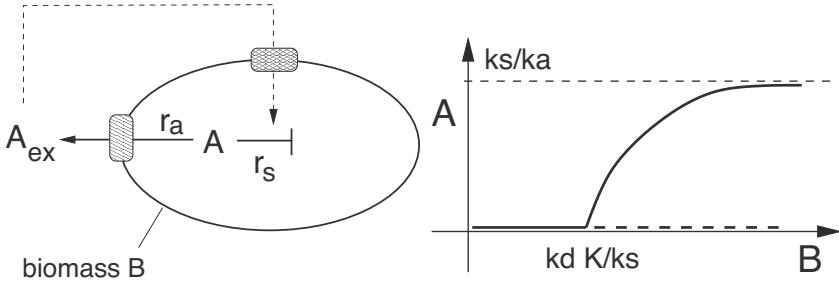


FIGURE 10.14: Left: Scheme of the quorum sensing system. The autoinducer A is excreted into the medium. Synthesis of A is positively influenced by the amount of A_{ex} . Right: Bifurcation diagram. Input is biomass B .

The steady state of the equation system is calculated. Equating to zero and rearranging the equations yields the two steady states:

$$\text{Steady state 1: } (0, 0)$$

$$\text{Steady state 2: } \left(\frac{k_s B - k_d K}{k_a B}, \frac{k_s B_S - k_d K}{k_d} \right). \quad (10.25)$$

It can be seen that the second steady state exists only under certain conditions: If the degradation rate is too high, this steady state does not exist. The relation:

$$B > \frac{k_d K}{k_s} \quad (10.26)$$

has to be satisfied. If $B = \frac{k_d K}{k_s}$, both steady states coincide. This point is referred to as a bifurcation point. The stability of the system can be examined by means of linearization. The Jacobian matrix is as follows:

$$J = \begin{pmatrix} -k_a & \frac{k_s K}{(K + A_{ex})^2} \\ k_a B & -k_d D \end{pmatrix}. \quad (10.27)$$

The characteristic polynomial is then:

$$\lambda^2 + (k_a + k_d) \lambda + k_a k_d - \frac{k_s k_a K B}{(K + A_{ex})^2} = 0 \quad (10.28)$$

The stability shall only be examined by means of the sign rule; that is, all signs must be positive. It is sufficient to consider the signs of the third and fourth term of the sum by inserting the respective values for A_{ex} . The first steady state is (only the third and fourth term are considered):

$$A_{ex} = 0 : k_a k_d - \frac{k_s k_a B}{K} > 0 \rightarrow B < \frac{k_d K}{k_s} \quad (10.29)$$

and the second steady state:

$$A_{ex} = \frac{k_s B - k_d K}{k_d} : k_a k_d - \frac{k_a K k_d^2}{k_s B} > 0 \rightarrow B > \frac{k_d K}{k_s}. \quad (10.30)$$

Therefore, the second stationary point, if it exists (see above), is stable. If the condition for the existence of the second steady state is not met, that is, only the first steady state exists, it is also stable. From a physiological point of view, the system is switched on if the biomass reaches a critical value (bifurcation point) that depends on the kinetic parameter k_s , K and k_d (Figure 10.14). The output A reaches its maximal value for high values of biomass.

10.4.2 Transfer of phosphate groups in a cascade

Phosphate groups are transferred in cascades in many systems. Examples are mainly known from eucaryotes: The MAP (mitogen-activated protein) kinase cascade is a signaling pathway that plays an important role in mitosis and cell division and can be found in different cell types. From a theoretical point of view, the pathway is very interesting due to its different variations. Here a general but simple cascade is introduced [11]. Figure 10.15 shows the scheme of signal transduction. The contemplation starts with kinase K_3^P , which affects the transcription of several genes. Kinase K_3^P is influenced in its activity by the kinase kinase K_2^P which is again modified by K_1^P (kinase kinase kinase). The input is a stimulus R , which activates synthesis of K_1^P , whereby the signal is transmitted into the cell by membrane-bound receptors. The following differential equations describe the dynamics of the system:

$$\begin{aligned} \frac{dK_1^P}{dt} &= k_1 \cdot K_1 \cdot R - k_{-1} \cdot K_1^P = k_1 \cdot K_{10} \cdot R \cdot \left(1 - \frac{K_1^P}{K_{10}}\right) - k_{-1} \cdot K_1^P \\ \frac{dK_2^P}{dt} &= k_2 \cdot K_1^P \cdot K_2 - k_{-2} \cdot K_2^P = k_2 \cdot K_{20} \cdot K_1^P \cdot \left(1 - \frac{K_2^P}{K_{20}}\right) - k_{-2} \cdot K_2^P \\ \frac{dK_3^P}{dt} &= k_3 \cdot K_2^P \cdot K_3 - k_{-3} \cdot K_3^P = k_3 \cdot K_{30} \cdot K_2^P \cdot \left(1 - \frac{K_3^P}{K_{30}}\right) - k_{-3} \cdot K_3^P. \end{aligned} \quad (10.31)$$

The steady states can be calculated successively since there is no feedback. The results are:

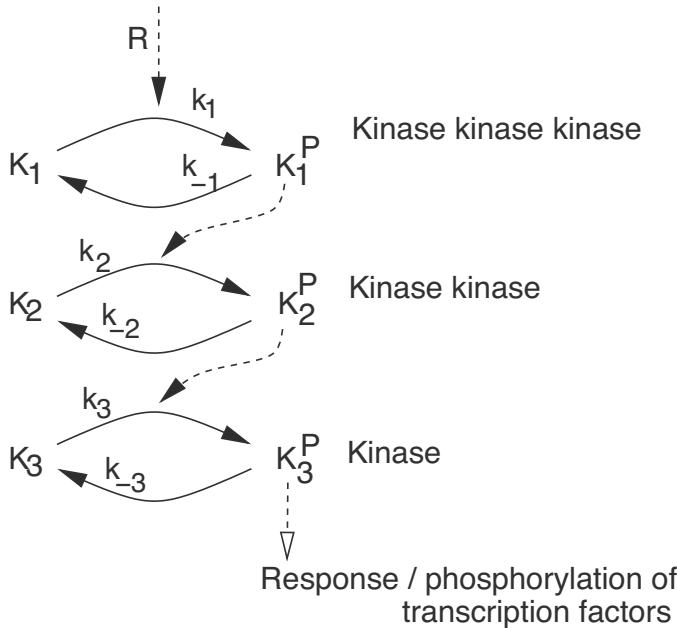


FIGURE 10.15: Scheme of the MAP kinase cascade.

Level 1

$$\begin{aligned}
 0 &= k_1 \cdot K_{10} \cdot R \cdot \left(1 - \frac{K_1^P}{K_{10}}\right) - k_{-1} \cdot K_1^P \\
 k_1 \cdot K_{10} \cdot R &= K_1^P \cdot (k_1 \cdot R + k_{-1}) \\
 K_1^P &= \frac{k_1 \cdot K_{10} \cdot R}{k_1 \cdot R + k_{-1}} = \frac{K_{10} \cdot R}{R + K_{D1}} \\
 \text{with } K_{D1} &= \frac{k_{-1}}{k_1};
 \end{aligned} \tag{10.32}$$

Level 2

$$\begin{aligned}
 k_2 \cdot K_{20} \cdot K_1^P &= (k_2 \cdot K_1^P + k_{-2}) \cdot K_2^P \\
 K_2^P &= \frac{k_2 \cdot K_{20} \cdot K_1^P}{k_2 \cdot K_1^P + k_{-2}} = \frac{K_{20} \cdot K_1^P}{K_1^P + K_{D2}} \\
 \text{with } K_{D2} &= \frac{k_{-2}}{k_2};
 \end{aligned} \tag{10.33}$$

Level 3

$$K_3^P = \frac{K_{30} \cdot K_2^P}{K_2^P + K_{D3}} \quad \text{with } K_{D3} = \frac{k_{-3}}{k_3}. \tag{10.34}$$

Under certain conditions the signal is amplified, namely if:

$$K_3^P = \frac{K_{30} \cdot K_2^P}{K_2^P + K_{D3}} > K_2^P \quad \longrightarrow \quad K_{30} > K_2^P + K_{D3}.$$

Iteratively inserting in the equations above yields the following relation for the whole cascade:

$$K_3^P = \frac{K_{30} \cdot R}{R \cdot \left(1 + \frac{K_{D3}}{K_{20}} + \frac{K_{D3} \cdot K_{D2}}{K_{10} \cdot K_{20}}\right) + \frac{K_{D3} \cdot K_{D2} \cdot K_{D1}}{K_{10} \cdot K_{20}}} \quad (10.35)$$

Here, as a result, the equation has the same structure as a Michaelis-Menten kinetics; the maximal value and the half-saturation parameter are given by a combination of the kinetic parameter of the individual steps. The MAP-kinase cascade was analyzed in depth in many publications. The negative feedback from the lowermost level to the first level plays an important role [12]. It ensures that the system can also oscillate. Since the kinases can be phosphorylated either once or twice at every level, one obtains large Hill coefficients for each level and the cascade subsequently shows a switch-like behavior.

Exercises

EXERCISE 10.1 Carbohydrate uptake.

For carbohydrate uptake it was shown that the kinetic parameters determine the slope of the function $EIIA(\mu)$. Use the tools from Metabolic Control Analysis to verify the results.

Bibliography

- [1] S. S. Shen-Orr et al. Network motifs in the transcriptional regulation network of Escherichia coli. *Nature Genetics* 31, 2002, 64-68.
- [2] R. Milo et al. Network motifs: simple building blocks of complex networks. *Science* 298, 2002, 824-827.
- [3] S. Mangan and U. Alon, Structure and function of the feed-forward loop network motif. *PNAS* 100, 2003, 11980-11985.
- [4] F. Neidhardt. *Physiology of the bacterial cell: a molecular approach*. Sinauer Associates, 1990.

- [5] P. W. Postma et al. Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. *Microbiol. Rev.* 57, 1993, 543-594.
- [6] A. Kremling et al. Analysis of global control of *Escherichia coli* carbohydrate uptake. *BMC Systems Biology* 1, 2007, 42.
- [7] A. Kremling et al. A feed-forward loop guarantees robust behavior in *Escherichia coli* carbohydrate uptake. *Bioinformatics* 24, 2008, 704-710.
- [8] A. Kremling et al. Analysis of two-component signal transduction by mathematical modeling using the KdpD/KdpE system of *Escherichia coli*. *BioSystems* 78, 2004, 23-37.
- [9] M. Goulian. Robust control in bacterial regulatory circuits. *Current Opinion in Microbiology* 7, 2004, 198-202.
- [10] G. Shinar et al. Input-output robustness in simple bacterial signaling systems. *PNAS* 104, 2007, 19931-19935
- [11] R. Heinrich et al.. Mathematical models of protein kinase signal transduction, *Molecular Cell* 9, 2002, 957-970.
- [12] J. Saez-Rodriguez et al. Modular analysis of signal transduction networks. *IEEE Control Systems Magazine* 24, 2004, 35-52.

This page intentionally left blank

Part IV

Analysis of Cellular Networks

This page intentionally left blank

Chapter 11

Metabolic Engineering

Having analyzed smaller networks that were also called motifs we now switch to cellular models, that is, models that describe a large number of individual reactions. One important application of such models is in biotechnology. So we focus first on such types of networks.

11.1 Reconstruction of Metabolic Network

Reconstruction of a metabolic network is a systematic record of information on the network's components and its interactions. Interactions can be described in different ways: (1) direct allocation of components, i.e., substrate and products that are transformed by enzymes, or information about genes and corresponding gene products, including regulation of gene expressions (e.g., transcription factors, sigma factors). (2) “Omics”-technologies give information about the conditions under which a gene is expressed, the rate at which an enzyme functions if uptake and production rates of metabolites into the cell are known. Consequently, a reconstruction can be seen as primary information that can be used and analyzed. Today there are various databases available to compare and collect this information. For this book, information from the Ecocyc database is widely used [1]. Additionally, one can use original publications and extract information by hand or by computer. In general, there are links in web-based databases that can help find information on selected genes, components, metabolic pathways or enzymes. Reconstructed networks are used for metabolic engineering, studies about bacterial evolution, analysis of networks, characterization of phenotype behavior and finding new relationships in molecular networks [2].

In Metabolic Engineering the networks help to make cells more efficient in the sense of a production system, that is, to analyze their potential (which bacteria can make the product best?) and intervention possibilities (which intervention leads to higher yield/productivity?). Intervention possibilities are, for example, optimizing the conditions for enzyme activities, eliminating feedback loops and by-product formation, eliminating overexpressed enzymes, and increasing export activities.

11.2 Tasks and Problem Definition

Metabolic Engineering aims in general to modify cellular networks to improve the production of interesting biotechnical components in quantity and quality. In the following, the focus will be on cellular networks given in form of stoichiometric reaction networks. According to the definition for reaction equations (see chapter modeling), stoichiometric coefficients are negative on the left side and positive on the right side. Reaction networks are characterized by a stoichiometric matrix, which includes the stoichiometric coefficients of all reactions, and a corresponding rate vector. This enables a general form of balance equations for all intracellular metabolites. The balance equations are in the form of differential equations and are as follows:

$$\dot{\underline{c}} = N \underline{r}(\underline{c}) - \mu \underline{c}. \quad (11.1)$$

In this equation \underline{c} ($n \times 1$) is the vector of the concentration of intracellular metabolites, N ($n \times q$) is the stoichiometric matrix, \underline{r} ($q \times 1$) is the rate vector that is dependent on the concentration of the components and on the kinetic parameters and μ (scalar) is the specific growth rate. Consequently, the network consists of n components and q biochemical reactions. The focus is normally on such parts of the metabolic network with high fluxes; that is, central metabolism is considered where the main fluxes are distributed into different other parts of the network. Often, one is not interested in the dynamics of the system, but rather in the steady states. Here, flux maps are interesting as they present quantitatively all rates in the network for a given growth rate or environmental condition. In these cases the equation above is simplified and the following basic equation is given:

$$\underline{0} = N \underline{r}. \quad (11.2)$$

This is an algebraic equation where rows correspond to components and columns to the rates. Due to the fact that some rates can be calculated or that the limits of some rates are known from measurements of enzyme activities, it is possible to formulate a number of problems that are basic to determine flux maps. The problems become more complex when network modifications, for example, genetic modifications like mutations, are taken into account.

Overview on problem formulations

The stoichiometric matrix N is an important variable in the analysis of networks. Biochemical reactions that are presented in this matrix are recorded in databases for many organisms, mostly bacteria. This is known as network reconstruction, when genetical and physiological data are integrated and made accessible.

- The stoichiometric matrix N presents the total network capacity. This comprises natural products that the cells can produce under distinguished conditions. The matrix helps to estimate the amount of monomers and precursors needed for the macromolecular composition of the cell. These fluxes must be considered in equations for the components of central metabolism since metabolites are often precursors and drain to monomers cannot be neglected.
- A flux map can be determined with the help of the stoichiometric matrix N . Quantitative values for the rate that fulfills the equation above can be determined. To do so, an intracellular network and measurements of the components in the medium serve as a starting point. The situation is shown in Figure 11.1.

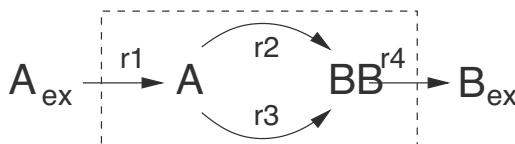


FIGURE 11.1: Simple network with four reactions, two external substrates/products and two intracellular metabolites A and B.

By considering all metabolites (external A_{ex} , B_{ex} and internal A , B) in the system, the following matrix N_T containing the stoichiometric coefficients for all reactions results:

$$N_T = \begin{pmatrix} r_1 & r_2 & r_3 & r_4 \\ 1 & -1 & -1 & 0 \\ 0 & 1 & 1 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{matrix} A \\ B \\ A_{ex} \\ B_{ex} \end{matrix} . \quad (11.3)$$

For many studies, only the intracellular metabolites are important, since it is assumed that the environment does not necessarily reach a steady state. Consequently, the total matrix can be subdivided (after changing the columns) as follows:

$$\underline{r} \quad r_{1,4} \\ N_T = \begin{pmatrix} N_{int} & | & N_{ex1} \\ 0 & | & N_{ex2} \end{pmatrix} \quad (11.4)$$

where \underline{r} includes two intracellular reactions r_2 and r_3 . Matrix N_{int} only

considers the dependency of the internal metabolites, while N_{ex1} describes the stoichiometry of transport reactions into the cell, and N_{ex2} describes the stoichiometry of transport rates from the point of view of external metabolites. External metabolites are not considered for the following analysis, only the matrix of the internal metabolites N , which is composed of N_{int} and N_{ex1} :

$$N = \begin{pmatrix} -1 & -1 & | & 1 & 0 \\ 1 & 1 & | & 0 & -1 \end{pmatrix}. \quad (11.5)$$

- The stoichiometric matrix can be seen as an invariant of the metabolic network. It is often sparse (has many zeros) and also contains reactions that present other “artificial” reactions like transport by diffusion, growth and dilution by growth.
- In an analysis essential and nonessential enzymes/proteins for a certain growth situation can be determined. Essential enzymes are active in all situations; turning off the enzyme does not lead to a solution when calculating the fluxes. Nonessential enzymes indicate that there is an alternative way that serves as a bypass.
- By integrating other data such as microarrays or proteome data, it is easier to evaluate each pathway’s importance for growth and product formation.
- By analysis, it is possible to determine metabolic pathways or enzyme groups that are always active together, meaning that they are correlated with each other. This suggests a control scheme, for example, a common transcription factor.
- Metabolic Engineering aims to optimize cellular performance. In simulation studies, changes in flux maps caused by network changes can be estimated by using matrix N .
- The stoichiometric matrix can be simplified by including data, such as knowledge about transcription factors that activate metabolic pathways under certain conditions.
- Simulation studies enable an estimation of how the fluxes change under different environmental conditions. This allows to determine how robust the fluxes are, as well as the network’s sensitivity when small changes of environmental conditions lead to greater changes in flux maps.

11.3 Subspaces of Matrix N

Firstly, the most important characteristics of the solution of the equation system above will be summarized. This will enable a solution of the problems described above [3]. Matrix N is characterized by four fundamental subspaces that will be discussed in the following. The four subspaces are: the (right) null space $Null(N)$ (named K), the left null space $lNull(N)$, as well as the row subspace $Row(N)$ and column subspace $Col(N)$

- The column subspace indicates temporal changes of variables (dynamic flux maps) and, thus, enables the formation of pool variables. These can have very fast dynamics that might not be within the considered time horizon.
- The null space indicates the stationary flux maps; it requires to determine the rates \underline{r} in steady state in such a way that $\dot{\underline{c}} = 0$.
- The column space indicates the weight of each rate in the velocity vector and can be used for model reduction.
- The left null space describes conservation conditions and, consequently, also time invariants of the system.

Normally matrix N is not fully ranked as the number of reactions q is greater than the number of participating components n . The rank of the stoichiometric matrix is then smaller than n and is fixed by r . The following relationships are important for subspaces in biochemical networks:

$$\begin{aligned} \dim(rank(N)) &= \dim(Col(N)) = \dim(Row(N)) = r \leq n \\ \dim(Null(N)) + \dim(Row(N)) &= q \\ \dim(lNull(N)) + \dim(Col(N)) &= n. \end{aligned} \quad (11.6)$$

All vectors in subspace K of N are summarized in one matrix that has the dimension $q-r$. The vectors form a vector space that can be described by basis vectors. Thus, linear combinations of each vector are also in the null space: $\underline{k} = K \cdot \underline{a}$ with a vector \underline{a} . The question is which basis vectors should be chosen for the null space, which can also be interpreted physiologically. Normally, the stoichiometry of the network under consideration is given so that the fluxes are expected to have positive values. Consequently, only those vectors with positive signs are taken into consideration as basis vectors. Figure 11.2, plot A, shows a simple network.

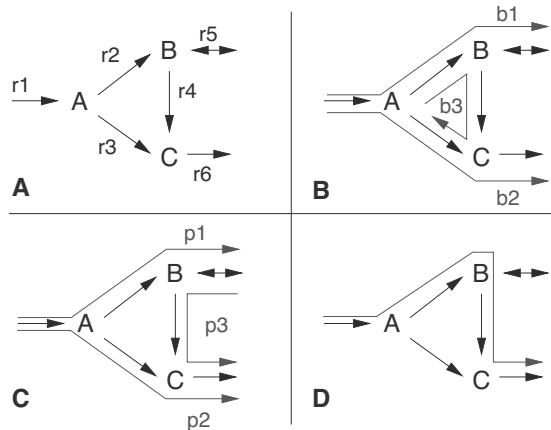


FIGURE 11.2: Example with 3 components and 6 reactions. **A** Presentation of the network. **B** Vectors of the null space. **C** Convex basis representation; reactions only run in the direction given by the scheme. Vector p_3 is a linear combination of the vectors \underline{b}_i with $p_3 = -\underline{b}_1 + \underline{b}_2 + \underline{b}_3$. **D** A further possibility to achieve a flux through the network with positive entries.

The network in the scheme has 6 reactions and 3 components. The null space can be illustrated with three vectors. By using software, one obtains the vectors shown in plot **B**.

$$\underline{b}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \underline{b}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \underline{b}_3 = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}. \quad (11.7)$$

Vector b_3 presents reaction r_3 in the opposite direction (expected direction is from A to C). The “Extreme Pathways” concept by B. O. Palsson searches for a set of basis vectors so that all flux maps can be presented as positive linear combinations of the basis vectors (shown in plot **C**). A different linear combination can be used to obtain a vector shown in plot **D**: $b^* = b_2 + b_3$.

The null space enables an identification of structural and functional aspects:

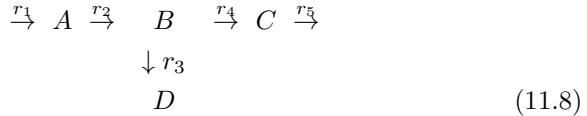
1. Coupled reactions:

Subset of reactions that always occur together and have the same pro-

portions of reaction rates in flux maps. An indication can be found by observing the rows in K and looking for dependencies.

2. Blocked reactions:

Do not occur in any flux maps with rates other than zero. Indications can be found in rows with zeros in K . In the network



the vector K becomes: $K = [1 \ 1 \ 0 \ 1 \ 1]$. In the third row there is a 0, which means that this rate is fixed for all conditions. This could be a hint to an inconsistent network structure.

3. Conservation relation:

The left null space is a time invariant of the system and can be interpreted as conservation relation. A conservation equation is the weighted sum of a metabolite concentration that is always constant. This can be written by using vector l ($n \times 1$ vector) as:

$$\underline{l}^T \cdot \underline{c}(t) = \text{const.} \quad (11.9)$$

The following network is considered as an example:



The stoichiometric matrix and the transposed matrix become:

$$N = \begin{pmatrix} -1 & 1 \\ -1 & -1 \\ 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad N^T = \begin{pmatrix} -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix}. \quad (11.11)$$

The dimension of the left null space is determined:

$$\dim(lN\text{Null}(N)) = n - \text{rank}(N) = 4 - 2 = 2, \quad (11.12)$$

and vector \underline{l} is sought:

$$\begin{pmatrix} -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix} \cdot \underline{l} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (11.13)$$

Note that we use $N^T \underline{l}$ instead of $\underline{l}^T N$. With that, two vectors can be given, which forms the matrix of the left null space L :

$$L = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}. \quad (11.14)$$

Thus, metabolites B and C as well as A and D represent conserved entities. The conservation equations allowing reducing the system. In the example one can calculate A and B and then determine C and D by using the two conservation relations.

11.3.1 Singular value decomposition of matrix N

Each subspace's dimension and values can be best determined by using singular value decomposition of matrix N . Singular value decomposition of a matrix is one of the different ways to factorize a matrix. A summary of other factorizations can be found in the Appendix. Matrix N can be presented as follows:

$$N = U \Sigma V^T \quad \text{with} \quad U^T U = I_n, \quad V^T V = I_q \quad (11.15)$$

with U and V as orthogonal matrices (this means that all vectors are orthogonal to each other and have value 1 for their Euclidian length). Σ is a diagonal matrix $diag(\sigma_1, \sigma_2, \dots, \sigma_r)$ with dimension $n \times q$, where: $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$. All other elements in the matrix are 0. The σ_i are known as singular values; they indicate the row/column weight of U and V . The following illustration in Figure 11.3 gives an overview of the structure of each matrix.

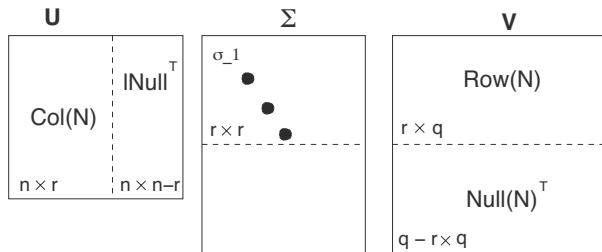


FIGURE 11.3: Singular value decomposition of the stoichiometric matrix N . The dimensions of the matrices are given.

By reformulating the basic equation above in

$$N V = U \Sigma \quad (11.16)$$

(where $V^T = V^{-1}$), it becomes clear that the stoichiometric matrix maps the right singular vectors (in matrix V) onto the left singular vectors (in matrix U), which are scaled with singular values. An example will explain this closer.

EXAMPLE 11.1 Singular values in a reversible reaction.

A simple elementary reversible reaction is considered, where substrates A and B are transformed.

$$A \rightleftharpoons B \quad \text{or broken down to} \quad \begin{array}{ll} A \rightarrow B & (r_1) \\ B \rightarrow A & (r_2). \end{array} \quad (11.17)$$

The stoichiometric matrix N can be decomposed as follows:

$$N = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \frac{1}{\sqrt{2}}. \quad (11.18)$$

According to the schema above, the subspaces have the following characteristics: The rank of matrix N is 1. As 2 reactions with 2 components are considered, all subspaces have dimension 1. The subspaces can be given as below (the values are scaled for better illustration):

$$\begin{aligned} Col(N) &= \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad lNull(N) = \begin{pmatrix} 1 & 1 \end{pmatrix} \\ Row(N) &= \begin{pmatrix} 1 & -1 \end{pmatrix}, \quad Null(N) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}; \end{aligned} \quad (11.19)$$

the only singular value is 2.

Figure 11.4 shows the subspaces in this case graphically. In the left figure both reaction rates r_1 and r_2 are plotted against each other. The bisecting line presents the subspace. Here both reactions have the same value. Row subspace ($Row(N)$) is orthogonal to the subspace $Null(N)$. Using ellipses, the two further cases $r_1 > r_2$ and $r_1 < r_2$ are shown. In the left figure, corresponding cases are shown in the coordination system \dot{A}, \dot{B} . If both rates are equally large, the system is in steady state and the velocity is zero. For $r_1 > r_2$, the system shifts so that A decreases (A is used) and B increases (B is produced). The movements of both conditions and therefore their velocity are no longer dependent on each other, but described by the column space.

Evaluation of Equation (11.16) leads to the following relation:

$$N V = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \quad U \Sigma = \begin{pmatrix} -2 & 0 \\ 2 & 0 \end{pmatrix}. \quad (11.20)$$

The null space vector is in the second column in V , which matrix N maps onto the null vector in the second column of $U \Sigma$.

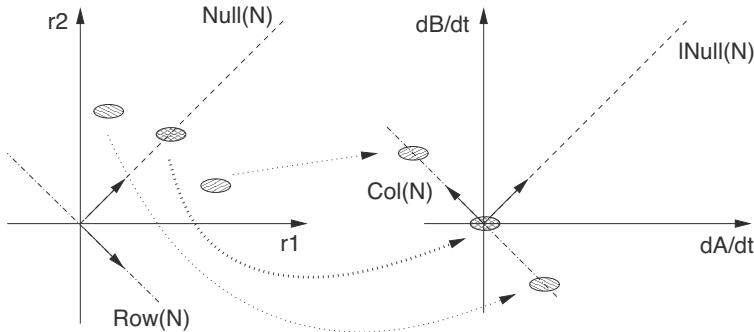


FIGURE 11.4: Illustration of subspaces with the example of a reversible reaction (based on B. O. Palsson: *Systems Biology – Properties of Reconstructed Networks*, chapter 8).

11.4 Methods to Determine Flux Distributions

11.4.1 Stationary flux distribution

The calculation of flux distributions is an important analysis when uptake and production rates are available. Here, values for \underline{r} are sought to satisfy:

$$\underline{0} = N \underline{r}. \quad (11.21)$$

For the stoichiometric matrix (N) in the example above (Equation (11.5)) a possible rate vector is $r^T = [1, 2, 3, 3]$ (observe the rate sequence r_2, r_3, r_1, r_4). This is due to:

$$N \underline{r} = \begin{pmatrix} -1 & -1 & 1 & 0 \\ 1 & 1 & 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \\ 3 \\ 3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \quad (11.22)$$

The vector is then a possible flux distribution. The flux distribution always results from linear combination of the null space vectors. In this case the null space is written as:

$$K = \begin{pmatrix} -1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, \quad (11.23)$$

and it can be shown that the flux vector above can be written as linear combination of null space vectors:

$$\underline{r} = 2 \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + 3 \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 3 \end{pmatrix}. \quad (11.24)$$

In metabolic flux analysis some fluxes in steady state in the network are measured and thus known. It is now necessary to reconstruct the inner flux distribution. Therefore, the rate vector is divided in known (k) and unknown (u) rates (u has p components):

$$0 = N \cdot \underline{r} = N_k \underline{r}_k + N_u \underline{r}_u \longrightarrow N_u \underline{r}_u = -N_k \underline{r}_k \quad (11.25)$$

and a solution for the equation system is sought. The vectors and matrices have the following dimensions: $N_u : n \times p$, $N_k : n \times (q-p)$, $\underline{r}_u : p \times 1$, $\underline{r}_k : (q-p) \times 1$. The solution of the system is dependent on the matrix rank N_u and measurement data. In the following a systematic approach is presented [4]. A general solution can be obtained as below:

$\underline{r}_u = -N_u^\# N_k \cdot \underline{r}_k + K_u \underline{a}$	with
$N_u^\#$	pseudoinverse of N_u (exists for all matrices)
K_u	null space of N_u (corresponds to the system's degree of freedom)
\underline{a}	random vector

(11.26)

The pseudoinverse (see also Appendix) and subspace can be calculated simply nowadays using software tools. Especially the composition of matrix N_u regarding number of rows and columns will be considered in the next part. Two cases can be differentiated:

- Determinacy

The system is determined if the number of equations is sufficient to calculate the unknown rates: $\text{rank}(N_u) = p$; it is under determined if $\text{rank}(N_u) < p$. The equations are not sufficient to solve the system.

- Redundancy

The system is not redundant if the rows in matrix N_u are not dependent and $\text{rank}(N_u) = n$. If $\text{rank}(N_u) < n$ the system is redundant. This means that using measurement information it is possible to test whether inconsistencies exist.

Combining Equation (11.26) and Equation (11.25), it is possible to obtain a further relationship:

$$0 = (N_k - N_u N_u^\# N_k) \underline{r}_k = R \underline{r}_k . \quad (11.27)$$

Matrix R is called a redundancy matrix with the dimension $n \times (q-p)$. It gives information about the rates in r_k that can be estimated in a redundant system. If the system is not redundant, then R is a null matrix. If there are values other than zero in the corresponding columns, then the respective rate can be estimated.

EXAMPLE 11.2 Metabolic network.

The network in Figure 11.5 with 7 metabolites and 9 reactions is considered. In the network ATP is formed in a reaction and used in a second reaction.

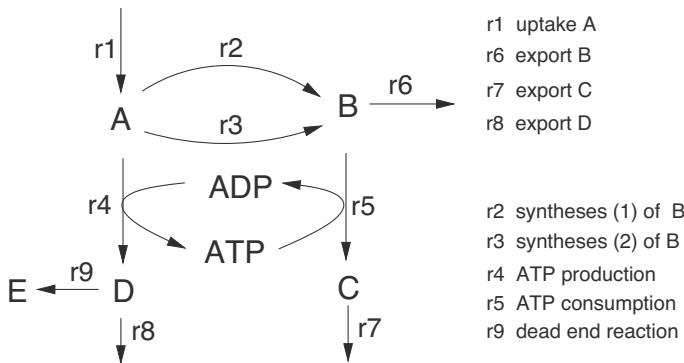


FIGURE 11.5: Network with nine reactions and seven metabolites.

By coupling both reactions interesting cases result, which will be considered in the following. The stoichiometric matrix is as below:

$$N = \begin{pmatrix} 1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} . \quad (11.28)$$

Reactions r_1 to r_9 can be found in the columns; in the rows there are

metabolite A to E . To begin with, the following values are determined: The matrix rank is $\text{rank}(N) = 6$. Thus, the dimension of the left null space is $\dim(lNull) = 1$ and the dimension of the null space is $\dim(Null) = 3$. In the null space a null row results:

$$K = \begin{pmatrix} 0 & 1 & 2 \\ -1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}. \quad (11.29)$$

Close inspection of K reveals that one reaction is fixed, which is reaction $r_9 = 0$. This is the dead-end reaction, which in this case always has a value zero. Furthermore, it is clear that reactions r_4, r_5, r_7 and r_8 are coupled. The corresponding values in the rows are correlated. From the left null space one can determine that the sum of ATP and ADP remains constant. Using the results so far, the network can be simplified. In the following some cases will be analyzed with different combinations of measured rates. A simpler network is considered, which eliminates the equations for ADP and E . Then matrix N has $n = 5$ rows and $q = 8$ columns. The matrix rank is $\text{rank}(N) = 5$.

Case 1: The network is determined and not redundant (Figure 11.6 on the left side): rates r_1, r_2 and r_6 are calculated. As 3 rates are measured, the number of unknown rates is $p = 5$ and the rank of N_u is $\text{rank}(N_u) = 5$. Thus, the system is determined and not redundant ($\text{rank}(N_u) = p = n$). All flux can be determined directly from measurements. The null space is empty; it is possible to invert directly: $N_u^\# = N_u^{-1}$. The solution is illustrated in Figure 11.6.

Case 2: The network is not determined and not redundant (Figure 11.6 in the center). To start out, only rate r_1 is calculated. Matrix N_u now has 5 rows and 7 columns and rank $\text{rank}(N_u) = 5$. Thus, the system is not determined and not redundant ($\text{rank}(N_u) < p = 7, \text{rank}(N_u) = n$). By observing the null space of matrix N_u , one can see that there are no null rows; this means that it is not possible to calculate any rate. We take r_7 as an additional measurement; the matrix N_u has now 5 rows and 6 columns, but the rank is still

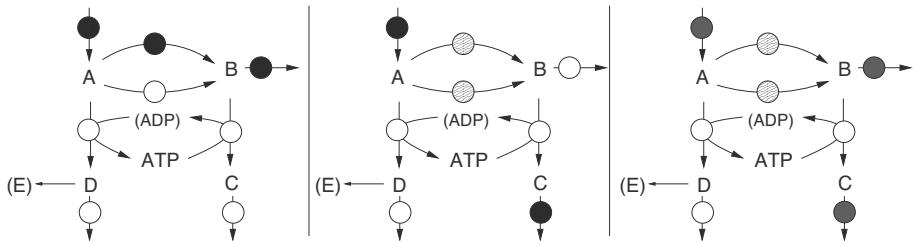


FIGURE 11.6: Classification of networks according to measured rates. On the left side: determined and not redundant network. In the center: not determined and not redundant network. The null space of N_u is: $K^T = [-1 \ 1 \ 0 \ 0 \ 0]$. On the right side: Not determined and redundant network; same null space as above. Filled black symbol: Rate is measured; filled gray: rate must be balanced; white symbol: rate can be determined; shaded symbol: rate cannot be determined.

$\text{rank}(N_u) = 5 < p = 6$. Consequently, the system continues to be not determined and not redundant ($\text{rank}(N_u) = n$); from observing the null space, it becomes clear that r_2 and r_3 cannot be determined. The other rates can be calculated, as Figure 11.6 shows.

Case 3: The network that is not determined and redundant (Figure 11.6 on the right side): rates r_1, r_6, r_7 are measured. Matrix N_u has 5 rows and 5 columns. However, the matrix has rank $\text{rank}(N_u) = 4 < p = 5$, $\text{rank}(N_u) = 4 < n = 5$. Thus, the network is not determined and redundant. The null space is the same as in the example above: rates r_2 and r_3 cannot be calculated. The three calculated rates can be balanced. In Figure 11.6 the flux distribution is illustrated. Matrix is R calculated as:

$$R = \begin{pmatrix} 0.14 & -0.14 & -0.29 \\ 0.14 & -0.14 & -0.29 \\ 0.28 & -0.28 & -0.58 \\ 0 & 0 & 0 \\ 0.14 & -0.14 & -0.29 \end{pmatrix}. \quad (11.30)$$

Matrix R has rank 1. With this, a reduced matrix R_r results, which has more columns than rows:

$$R_r = \begin{pmatrix} 0.14 & -0.14 & -0.29 \end{pmatrix}. \quad (11.31)$$

The system is overdetermined. The measured values in \underline{r}_k do not fulfill the equations exactly, so that further steps are necessary.

From the example above, two cases emerge that need further tools to continue the analysis: (i) There are too much measurements available for unique solution. To proceed further regression techniques can be applied to determine those values that describe the measured data best. (ii) There are too few measurements available to solve the system. This results in a number of degrees of freedom. To select one from many flux distributions an objective function is formulated. Both problems lead to special optimization tasks, which will be discussed in the following.

11.4.2 Redundant metabolic network

For redundant networks regression techniques can be applied. Here, the problem is to find a mathematical representation that allows taking into account also the measurement errors. Two approaches were introduced where the first one is a standard that does not allow including the errors while the second one does.

The first approach starts with the formulation of the problem as above. It is given:

$$\underline{0} = N_u \underline{r}_u + N_k \underline{r}_k \rightarrow N_u \underline{r}_u = -N_k \underline{r}_k. \quad (11.32)$$

In comparison with standard notation of the linear regression problem the left side represents matrix X and the unknown parameters \underline{p} , and the right side represents the measured data points \underline{Y}_M (equations given in the chapter on linear regression). The solution of this optimization problem leads to:

$$\underline{r}_u = - (N_u^T N_u)^{-1} N_u^T (-N_k \underline{r}_k). \quad (11.33)$$

Unfortunately, the measured values \underline{r}_k are represented only indirectly in \underline{Y}_M . Therefore measurement errors cannot be taken into account. As an alternative approach [6], the following equation set can be considered:

$$\begin{pmatrix} \underline{r}_k \\ \underline{0} \end{pmatrix} = \begin{pmatrix} I & \underline{0} \\ N' & \end{pmatrix} \begin{pmatrix} \underline{r}_k \\ \underline{r}_u \end{pmatrix} \quad (11.34)$$

where N' is the same as N but with a modified order of the columns as indicated by the vector on the right side and I is the identity matrix of dimension $q - p$. Now the matrix is subdivided in a different manner with four submatrices:

$$\begin{pmatrix} \underline{r}_k \\ \underline{0} \end{pmatrix} = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \begin{pmatrix} \underline{r}_I \\ \underline{r}_{II} \end{pmatrix}. \quad (11.35)$$

The arrangement can be explored by solving the second row. If metric T_{22} is not singular one gets:

$$\underline{r}_{II} = -T_{22}^{-1} T_{21} \underline{r}_I. \quad (11.36)$$

Inserting in the first row leads to:

$$\underline{r}_k = T_{11} \underline{r}_I + T_{12} \underline{r}_{II} = \mathbf{T} \underline{r}_I \quad (11.37)$$

with

$$\mathbf{T} = T_{11} - T_{12} T_{22}^{-1} T_{21}. \quad (11.38)$$

If we compare now this structure with the standard notation for the linear regression problem we get: $\underline{r}_k \equiv \underline{Y}_M$, $\mathbf{T} \equiv X$ and $\underline{r}_I \equiv \underline{p}$. Note that not only the unmeasured rates are determined but also the already measured rates are re-calibrated:

$$\begin{aligned} \hat{\underline{r}}_I &= (\mathbf{T}^T \Sigma^{-1} \mathbf{T}) \mathbf{T}^T \Sigma^{-1} \underline{r}_k \\ \hat{\underline{r}}_{II} &= -T_{22}^{-1} T_{21} \hat{\underline{r}}_I. \end{aligned} \quad (11.39)$$

In this way the measurement error can be considered in matrix Σ^{-1} .

EXAMPLE 11.3 *Simple network with four reactions and two metabolites (Figure 11.1).*

The stoichiometric matrix is

$$N = \begin{pmatrix} 1 & -1 & -1 & 0 \\ 0 & 1 & 1 & -1 \end{pmatrix}. \quad (11.40)$$

For the first approach, N must be subdivided. If rates r_1 , r_3 and r_4 are measured the respective submatrices are:

$$N_u = \begin{pmatrix} -1 \\ 1 \end{pmatrix}; \quad N_k = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \quad (11.41)$$

and the unknown rate r_2 can be estimated with Equation (11.33):

$$\hat{r}_2 = (N_u^T N_u)^{-1} N_u^T (-N_k \begin{pmatrix} r_1 \\ r_3 \\ r_4 \end{pmatrix}). \quad (11.42)$$

Inserting leads to:

$$\hat{r}_2 = \frac{r_1 + r_4}{2} - r_3. \quad (11.43)$$

Rate r_2 is calculated by the mean value of r_1 and r_4 minus the rate r_3 representing the second way to build component B .

Exploring the alternative method the following equation is obtained at first:

$$\begin{pmatrix} r_k \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & -1 \\ 0 & 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} r_I \\ r_{II} \end{pmatrix}. \quad (11.44)$$

In this case the submatrix T_{22} corresponds to the last element in the last row/last column or the last two columns and the last two rows. In both cases T_{22} is invertible. We follow the second possibility, and we obtain:

$$T_{22} = \begin{pmatrix} 0 & -1 \\ -1 & 1 \end{pmatrix}; \quad r_I = \begin{pmatrix} r_1 \\ r_3 \end{pmatrix}; \quad r_{II} = \begin{pmatrix} r_4 \\ r_2 \end{pmatrix}. \quad (11.45)$$

Now, let's assume that we can measure r_1 and r_3 twice as well as r_4 . Then, matrix Σ reads:

$$\Sigma = \begin{pmatrix} \sigma & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 2\sigma \end{pmatrix} \quad (11.46)$$

and we obtain for the estimation of all fluxes the following result:

$$\begin{aligned} \hat{r}_1 &= 2r_1/3 + r_4/3; & \hat{r}_3 &= r_3 \\ \hat{r}_4 &= 2r_1/3 + r_4/3; & \hat{r}_2 &= 2r_1/3 - r_3 + r_4/3. \end{aligned} \quad (11.47)$$

The influence of the measurement error can be seen in the rate \hat{r}_1 and \hat{r}_4 . Due to the high error in rate \hat{r}_4 the weight of \hat{r}_1 in the solution is double the value than for \hat{r}_4 . For equal errors the same result is obtained as for the simple procedure.

11.4.3 Optimality criteria

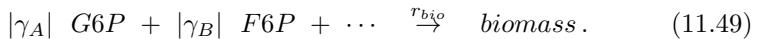
When there are not enough equations to solve the system, this means that the degrees of freedom lead to many solutions. In order to limit the range of feasible solutions, an objective function is defined and an extreme value is sought. The hypothesis behind this is that in the course of evolution, cells have developed optimally. However, the question as to what an objective function looks like and which principles are behind the observed growth pattern is more difficult to answer. A range of approaches can be derived from literature (for

an overview see [5]). Often, a scaling of fluxes is done in a way that all fluxes are scaled with respect to the uptake rates of the primary C source. Here r_{glc} is the glucose uptake rate used in the following relationships as glucose is used as substrate in most cases. The rate is seen as a constant.

- **Maximizing biomass production.**

$$\max \Phi = \frac{r_{bio}}{r_{glc}}. \quad (11.48)$$

It is assumed that due to evolution, cells evolved towards high growth rates. In order to calculate r_{bio} a stoichiometric equation is used that describes a flux from precursors to monomers. From composition of the macromolecules, the usage of monomers can be derived, which can in turn be used to calculate back the drain of precursors.



- **Maximize ATP-production.**

$$\max \Phi = \frac{\sum_i r_{ATPi}}{r_{glc}}. \quad (11.50)$$

As the most important energy provider, ATP also stands indirectly for high growth rates. Therefore, all rates that are generated by ATP are summed up.

- **Minimization of all intracellular rates.**

$$\min \Phi = \sum_{i=1}^q r_i^2. \quad (11.51)$$

This criterion is based on the assumption that enzymes are only produced in the needed amount. As there are many enzymes in one cell, the production must be made economically.

- Further criteria are combined from the listed objective functions:

$$\max \Phi = \frac{\sum_i r_{ATPi}}{\sum_{i=1}^q r_i^2} \quad (11.52)$$

and

$$\max \Phi = \frac{r_{bio}}{\sum_{i=1}^q r_i^2} \quad (11.53)$$

These two criteria are not linear and thus difficult to solve. In order to get a linear equation the following criterion is considered:

$$\max \Phi = \sum_i r_{ATP_i} - \epsilon \sum_{i=1}^q r_i^2 \quad (11.54)$$

and

$$\max \Phi = r_{bio} - \epsilon \sum_{i=1}^q r_i^2. \quad (11.55)$$

with a corresponding weight factor ϵ .

11.4.4 Linear optimization

In linear optimization, the objective function and equality constraints are described by linear functions (see above). The problem formulation is as follows:

$$\begin{aligned} \max f : \quad & \underline{c}^T \cdot \underline{r} \\ N \underline{r} = & -\underline{r}_k \\ \underline{r} \geq & 0. \end{aligned} \quad (11.56)$$

The measured rates \underline{r}_k are extracted from stoichiometry. In this standard formulation of a linear optimization problem, the last equation is a further constraint. In applications, the sign of variables r_i must be known.

The equality condition restricts the search for an optimal solution on a polytope (generalized polyhedron). It is important to note that the number of equations n is smaller than the number of variables q of the rate vector or else the equation system would have one unique solution and optimization would not be necessary anymore. The equation system has an unlimited number of solutions due to the degrees of freedom, which is illustrated in Figure 11.7, which shows a polytope. The equation system is given as:

$$N = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \text{ and } -\underline{r}_k = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \quad (11.57)$$

As one can see, the first row in N shows a plane with an angle of 45 degrees. This plane is intersected with the plane $x_2 = 0$. The intersecting set is the black line with two corners.

Above, a general solution for algebraic equation systems was given (Equation (11.26)) by calculating the pseudoinverse of N . However, we are not interested in a special solution but are interested in solving an optimization problem. The analysis is limited to the corners of the polytope. It is possible to show that an optimal point is one of the corners of the polytope. To understand the following we have to know that the corners are solutions of an

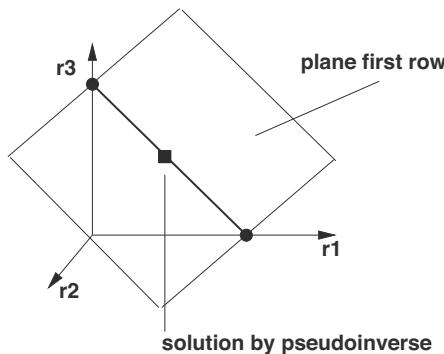


FIGURE 11.7: Equation system with two linear equations in \mathbb{R}^3 . The polyhedron has two corners. The line between the corners presents all possible solutions. The black square shows the solution of the pseudoinverse.

equation system given as:

$$N_b \underline{r}_b = -\underline{r}_k \rightarrow \underline{r}_b = -N_b^{-1} \underline{r}_k \quad (11.58)$$

with where matrix N is subdivided in such a way:

$$N = [N_b \ N_{nb}] \quad (11.59)$$

that N_b is quadratic and not singular; that is, we choose n columns from N and calculate the inverse. Furthermore it is required that all entries of \underline{r}_b are positive (see Equation 11.56). Note that the order of the fluxes has changed by this operation. The corners are the entries of \underline{r}_b while all other entries are zero. To verify if a corner represents the optimal solution, one inserts the equation for the corner into the objective function. Since the solution includes also the degree of freedom, one can easily check if the variables representing the degree of freedom can increase the value of the objective function or not. Inserting the corner into the objective function (with newly ordered entries also in vector $\underline{\gamma}$) we have to check if the entries of vector $\underline{\gamma}$:

$$\underline{\gamma} = \underline{c}_{nb} - \underline{c}_b N_b^{-1} N_{nb} \quad (11.60)$$

are smaller than zero. If this is the case the selected corner is the optimal one because every use of the degree of freedom makes the objective function smaller. Determination of flux distributions in underdetermined networks is termed Flux Balance Analysis (FBA). For more details on linear optimization, see the Appendix.

EXAMPLE 11.4 Simple network with four reactions and two metabolites (Figure 11.1).

For the example we assume that r_1 is the only measured rate of the system. Furthermore, the stoichiometry is slightly modified. In this way the formulation of the linear program reads:

$$\begin{aligned} \max f : \quad & \begin{pmatrix} a & 1 & 0 \end{pmatrix} \begin{pmatrix} r_2 \\ r_3 \\ r_4 \end{pmatrix} \\ & \begin{pmatrix} -1 & -1 & 0 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} r_2 \\ r_3 \\ r_4 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \\ \underline{r} & \geq 0. \end{aligned} \quad (11.61)$$

From matrix N we can construct 3 matrices with dimension 2×2 (first, second column, first, third column, and second, third column). With Equation (11.58) for 2 matrices (case 2 and 3), the inverse can be calculated and the two corners of the equation system are (in correct order):

$$\underline{r}_{c1} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}; \quad \underline{r}_{c2} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}. \quad (11.62)$$

For the third case the inverse cannot be determined. Using Equation (11.60) to analyze both corners, one obtains for γ (since we have only one degree of freedom, γ is a scalar):

$$\gamma_{c1} = 1 - a; \quad \gamma_{c2} = a - 1. \quad (11.63)$$

Since a represents the weight of reaction r_2 in the objective function, corner 1 is the optimal solution for a larger than one ($\gamma_{c1} < 0, \gamma_{c2} > 0$); in contrast, if a is smaller than 1, corner 2 is the optimal one.

11.5 Strain Optimization

The term strain optimization is often connected to Metabolic Engineering. The goal is to change the natural system so that improved flux distributions and more generalized growth parameters result.

- Including genetic information for heterologous proteins

- More effective substrate transformation (extending the substrate spectrum)
- Including more complete/extended metabolic pathways
- Increasing productivity by strengthening previous paths
- Improving cellular physiology
- Improving tolerance with low O_2 -concentration
- Improving inhibition of substrate excess
- Increasing stress tolerance

There are a number of approaches in literature that enable a theoretical consideration of this problem. These approaches can be divided into those that consider the whole cellular network or individual metabolic pathways. The team of Costas Maranas presented in one of their first works an approach to find those mutants that lead to optimized flux of a product [7]. To summarize the approach, the following topics have to be considered:

- Maximal growth rate: As in the former approach the growth rate, that is, the flux to biomass, should be maximal;
- Gene knockout strategy: Introduction of integer variables that are assigned to each flux and can adopt values zero (gene is made inactive) and one (gene is active/gene product is expressed).
- Formulating the problem as a bi-level-optimization problem.

In order to explain the approach, the standard problem is examined once again:

$$\max \Phi = \underline{c}_1^T \underline{r} \quad (11.64)$$

$$\text{s.t.} \quad N \underline{r} = -\underline{r}_k$$

$$\underline{r} \geq 0. \quad (11.65)$$

Unknown fluxes are considered in vector \underline{r} , and known/measured rates in vector \underline{r}_k . In the next step, binary variables e are introduced corresponding to each unknown rate, which allow a knockout of the genes and therefore the respective rates are zero:

$$e_j = \begin{cases} 1 & \text{when } r_j \text{ active} \\ 0 & \text{when } r_j \text{ inactive} \end{cases} \quad (11.66)$$

The following formulations ensure that when an integer variable e_j is zero, the corresponding flux disappears; otherwise the flux is in the given interval.

However, the opposite does not apply: If FBA determines that a rate is equal to zero, it does not result automatically that the corresponding integer variable is zero as well. The formulation is as:

$$r_j^{\min} e_j \leq r_j \leq r_j^{\max} e_j \quad \text{with cases:} \quad (11.67)$$

$$e_j = 1 : \quad r_j^{\min} \leq r_j \leq r_j^{\max}$$

$$e_j = 0 : \quad r_j = 0.$$

The bi-level-optimization problem is then formulated as follows, if vector \underline{c}_2 forms the secondary objective function:

$$\begin{aligned} & \max_{e_j} \underline{c}_2^T \underline{r} \\ \text{s.t. } & \max \underline{c}_1^T \underline{r} \\ \text{s.t. } & N \underline{r} = -\underline{r}_k \\ & r_j^{\min} e_j \leq r_j \leq r_j^{\max} e_j \\ & \sum (1 - e_j) \leq k. \end{aligned} \quad (11.68)$$

The last line ensures that the number of allowed mutations is not greater than number k . The unequal sign can be exchanged by an equal sign. This means that exactly k mutations are allowed.

Concerning the solution to the problem, the bi-level-optimization problem cannot be solved directly. For this reason, it is reformulated as a simple optimization problem where the strong duality is used. The inner problem is exchanged by an algebraic equation system. For the given inner problem:

$$\begin{aligned} \max \Phi &= \underline{c}_1^T \underline{r} \\ \text{s.t. } & N \underline{r} = -\underline{r}_k \\ & r_j \geq 0, \end{aligned} \quad (11.69)$$

the corresponding dual problem is (see Appendix):

$$\begin{aligned} \min \Phi' &= -\underline{r}_k^T \underline{y} \\ \text{s.t. } & N^T \underline{y} \geq \underline{c}_1. \end{aligned} \quad (11.70)$$

If the previous problem has an optimal solution, then both values of the objective functions correspond to each other: $\Phi = \Phi'$. By putting all information together, one obtains a MILP (mixed integer linear program) with equality

and inequality constraints. There is software available for this purpose. With vector \underline{x} :

$$\underline{x}^T = [\underline{r}^T \ \underline{y}^T \ \underline{e}^T] \quad (11.71)$$

one obtains the following equality constraint with blocks

$$\begin{pmatrix} N_{n \times q} & 0_{n \times n} & 0_{n \times q} \\ \underline{c}_1 & \underline{r}_k^T & 0 \end{pmatrix} \underline{x} = \begin{pmatrix} -\underline{r}_k \\ 0_{1 \times 1} \end{pmatrix} \quad (11.72)$$

and inequality constraints with the blocks:

$$\begin{pmatrix} -I_{q \times q} & 0_{q \times n} & diag(\underline{r}_{min}) \\ I_{q \times q} & 0_{q \times n} & -diag(\underline{r}_{max}) \\ 0_{q \times q} & -N_{q \times n}^T & 0_{q \times q} \\ 0_{1 \times q} & 0_{1 \times n} & \underline{i}_{1 \times q} \end{pmatrix} \underline{x} \leq \begin{pmatrix} 0_{2q \times 1} \\ -\underline{c}_1 \\ k - q \end{pmatrix}. \quad (11.73)$$

The index indicates matrix's dimensions.

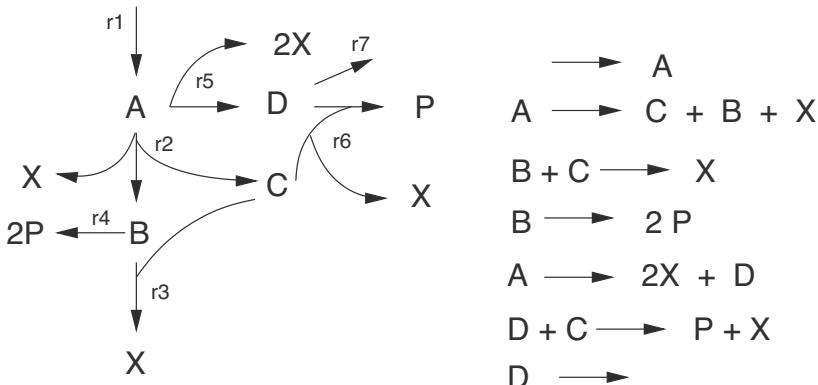


FIGURE 11.8: Network to optimize product P . The stoichiometry is given on the right side.

EXAMPLE 11.5 Network to demonstrate strain optimization.

The following small network in Figure 11.8 is considered with two objective functions:

$$\begin{aligned} c_1 &= r_2 + r_3 + 2r_5 + r_6 \\ c_2 &= 2r_4 + r_6 \end{aligned} \quad (11.74)$$

where c_1 describes the rates that contribute to biomass formation X while c_2 includes rates that contribute to the product formation P .

Solving the system with the approach introduced reveals that there are 3 pathways in total for biomass synthesis, where 1 mol influx in component A results in 2 mol of X . The outer problem searches from these pathways for the one that results in the maximal flux to product P . All flux maps are shown in Figure 11.9.

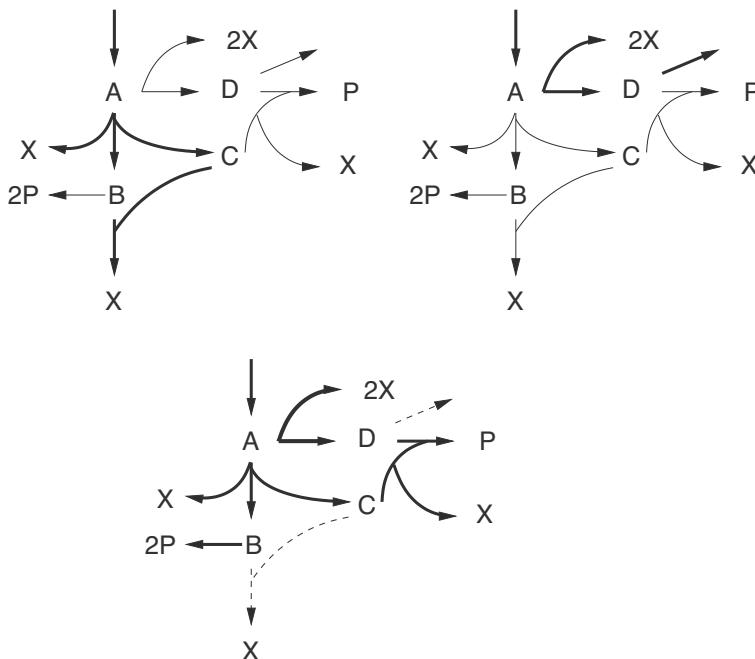


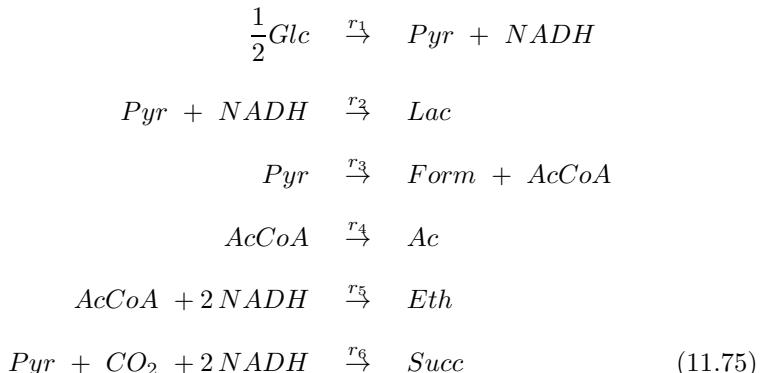
FIGURE 11.9: Upper left and right: Flux maps that result in maximized X but suboptimal production of P . Bottom: Flux map that results in maximized X and P .

Exercises

The intention of the exercises given here is to set up the correct equations to analyze the system and to determine the flux distributions through the network.

EXERCISE 11.1 Metabolic flux analysis.

Consider the following biochemical reaction network to produce succinic acid:



with extracellular compounds glucose, lactate, formate, acetate, ethanol, succinate and CO_2 . Intracellular metabolites are NADH, pyruvate and acetyl CoA.

a Write down the stoichiometric matrix N for the intracellular metabolites. What is the dimension of N ? How many rates do you have to measure at least, to calculate a unique flux distribution?

b The measured rates are glucose uptake, lactate and formate production. Is it possible to determine all rates definitely?

c The measured rates are formate, acetate and ethanol production. Is it now possible to determine all rates definitely?

d The measured rates are glucose uptake and lactate formation. Classify the system and determine the null space in this case. Interpret the result.

EXERCISE 11.2 Flux balance analysis - Linear program.

Consider a small network with one metabolite A and three rates (uptake rate r and excretion rates r_1 and r_2). Based on experimental knowledge the maximal values of rates r_1 and r_2 are known. The following linear program has to be solved:

$$\begin{aligned}
 \max \quad f &= a r_1 + r_2 \\
 s.t. \quad r - r_1 - r_2 &= 0 \\
 r_1 - 1 &\leq 0 \\
 r_2 - 2 &\leq 0. \tag{11.76}
 \end{aligned}$$

a By introduction of two additional variables r_{s1} and r_{s2} the two inequalities can be transformed into equations. Write down the problem in standard form.

b Determine all corners of the polytope and determine the optimal corner as a function of parameter a .

EXERCISE 11.3 *Strain optimization - Mixed integer linear program.*

For the network in the example for strain optimization give the complete matrices shown in Equations (11.72) and (11.73).

Bibliography

- [1] I. M. Keseler et al. EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic Acids Research* 39:D583-590, 2011.
- [2] A. M. Feist and B. O. Palsson. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nature Biotechnology* 26: 659, 2008.
- [3] B. O. Palsson. *Systems Biology – Properties of Reconstructed Networks*. Cambridge University Press, 2006.
- [4] S. Klamt et al. Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria. *Biotech. and Bioeng.* 77: 734-751, 2002.
- [5] R. Schuetz et al. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular Systems Biology* 3: 119, 2007.
- [6] S. P. Tsai and Y. H. Lee. Application of stoichiometric pathway stoichiometry to statistical analysis of bioreactor measurement data. *Biotech. and Bioeng.* 23: 713-715, 1988.
- [7] Burgard et al. OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotech. and Bioeng.* 84:647–657, 2003.

This page intentionally left blank

Chapter 12

Topological Characteristics

For large networks describing several interactions, a graph can be used to show the link between two components. In this chapter, methods for analyzing such networks are discussed [1]. They can be used in case the networks were reconstructed with the aforementioned methods. In this chapter, measures characterizing large networks are introduced. It is worth mentioning that technical and social systems possess characteristics similar to those of biochemical networks.

12.1 Network Measures

12.1.1 Connectivity and cluster measures

Connectivity: Denotes the number of edges k_j (or links) of a node j . In a directed graph, edges connected to a node j are either incoming (k_{inj} , head endpoint) or outgoing (k_{outj} , tail endpoint). For a known connectivity, the connectivity distribution $P(k)$ can be computed for each node. $P(k)$ is the probability that a chosen node has exactly k connections. In this case, the number of nodes with $k = 1, 2, \dots$ connections is divided by the total number of nodes. Figure 12.1 shows examples of distributions.

Cellular networks are characterized by a power law distribution: $P \approx k^{-\gamma}$ with γ is the connectivity exponent; typical values for γ are between 2 and 3. Distributions of this type indicate the existence of a few nodes with a large number of connections. These nodes are called hubs. Consequently, there are many nodes with only a few links in the network. Networks with a power law distribution are called scale free.

The connectivity of metabolic networks can also be calculated with a special representation of the stoichiometric matrix N . Thus, a matrix \hat{N} is defined. Each entry of \hat{N} is assigned a value of 1 if the corresponding entry of the matrix N is not equal to 0; otherwise it is assigned a value of 0. The matrix \hat{N} is also known as the binary form of N . The number of participating components in a reaction can be determined by adding up the entries of the corresponding columns. The sum of the entries in a column determines the components' degree of connectivity. Other quantities can be calculated from

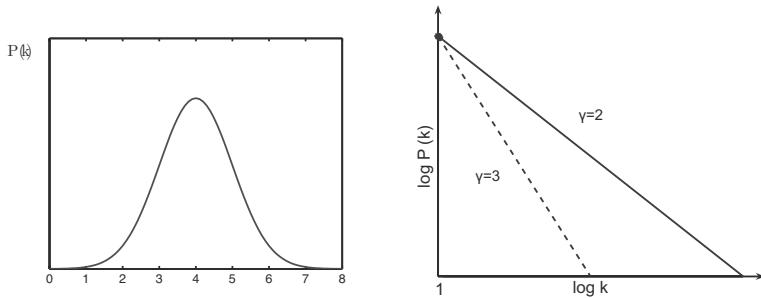


FIGURE 12.1: Connectivity distribution. It is necessary that: $\sum P(k) = 1$. On the left side: Networks characterized by mean and variance. On the right side: The power law distribution typical of cellular networks in a log-log diagram.

the matrices:

$$A_R \quad = \quad \hat{N}^T \hat{N} \quad \text{and} \quad (12.1)$$

$$A_K \quad = \quad \hat{N} \hat{N}^T. \quad (12.2)$$

One obtains the number of participating partners in a reaction via diagonal elements of the matrix A_R . The secondary diagonal of A_R contains the number of components that two reactions have in common. The diagonal elements of the matrix A_K indicate connectivity. The secondary diagonal elements of A_K indicate how many reactions two components have in common. Both matrices are symmetric.

EXAMPLE 12.1 Network with four components and five reactions.

Figure 12.2 shows the network where a simple stoichiometry (all $\gamma = 1$) is assumed.

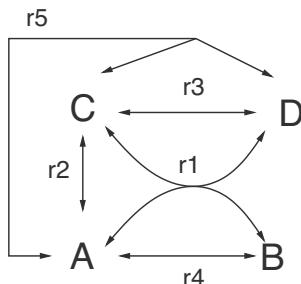


FIGURE 12.2: Network with four components and five reactions.

The matrices A_R (5×5) and A_K (4×4) are:

$$A_R = \begin{pmatrix} 4 & 2 & 2 & 2 & 3 \\ 2 & 2 & 1 & 1 & 2 \\ 2 & 1 & 2 & 0 & 2 \\ 2 & 1 & 0 & 2 & 1 \\ 3 & 2 & 2 & 1 & 3 \end{pmatrix}, \quad A_K = \begin{pmatrix} 4 & 2 & 3 & 2 \\ 2 & 2 & 1 & 1 \\ 3 & 1 & 4 & 3 \\ 2 & 1 & 3 & 3 \end{pmatrix}. \quad (12.3)$$

From A_K it can be determined that components A and C participate in 4 reactions (entries 1 and 3 of the main diagonal) and have 3 reactions (entries in row 1, column 3) in common (reactions r_1, r_2, r_5). From entry (1, 1) of the matrix A_R it can be determined that 4 components participate in the first reaction. Entry (1, 5) shows that there are three components participating in both r_1 and r_5 . From the property of connectivity, the following particular characteristics of cellular networks can be identified:

- Scale free

A randomly generated network (Figure 12.1 on the left) can be characterized by mean and variance. A network scale is created by using these values; scale free networks cannot be characterized by these entities. The mean value is irrelevant for the power law distribution as shown in Figure 12.1 (right side).

- Small world effect

In many networks the shortest path between any two nodes is surprisingly small. The shortest path is the minimal number of connections between two nodes. Generally, in a directed graph $l_{AB} \neq l_{BA}$ is valid. The average path length is the mean value $\langle l \rangle$ of all shortest paths of all connections. An investigation of scale free networks reveals that they are ultra small.

Clustering coefficient C_i : C_i determines the frequency with which the neighbors of a node i are connected. C_i is defined as:

$$C_i = \frac{2 n_i}{k(k-1)}, \quad (12.4)$$

where n_i is the number of connections among the neighbors of the node i and $\frac{k(k-1)}{2}$ is the total number of all possible links between the neighbors of node i .

EXAMPLE 12.2 Clustering coefficient C_i for various networks.

Examples I and II in Figure 12.3 represent extreme situations. The clustering coefficient of A is 0 (case I) and 1 (case II), respectively. In example

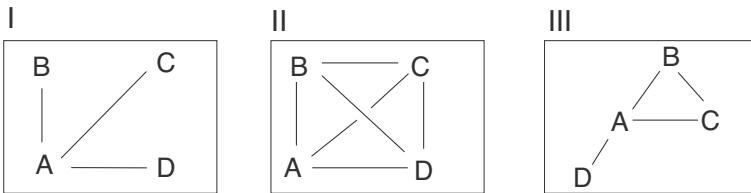


FIGURE 12.3: Examples of calculating the clustering coefficient C_i . The clustering coefficient for node A in example I is 0. The neighbors are not linked. In example II it is 1 for all nodes. For example III: $C_A = 1/3$ and $C_B = 1$.

III, A has 3 neighbors. The neighbors can have up to 3 connections with the other neighbors. The previous equation results in: $C_A = 1/3$ and $C_B = 1$.

The average clustering coefficient $C = \langle C_i \rangle$ indicates the general tendency of nodes forming a cluster. $C(k)$ can be assigned to nodes with k connections. If $C(k) \sim \frac{1}{k}$, the network is hierarchical (Figure 12.4). Therefore, hubs have lower clustering coefficients.

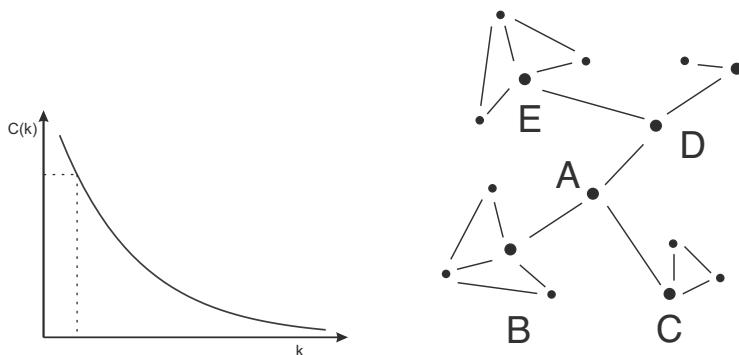


FIGURE 12.4: Left side: The mean clustering coefficient in a hierarchical network becomes smaller with an increasing number of connections. Right side: A hierarchical network.

12.2 Topological Overlap

While the clustering coefficient is a characteristic of each individual node, the topological overlap analyzes the nodes' behavior in pairs [2]. Thus, rela-

tionships and similarities between two nodes can be determined. The following relationship is applicable:

$$O(i,j) = \frac{Jn(i,j)}{\min(k_i, k_j)}, \quad (12.5)$$

where $Jn(i, j)$ is the number of nodes linked with both nodes i and j (plus 1 if i and j are directly linked with each other) and k_i, k_j is the number of the nodes' connections i, j . In the following example, both the clustering coefficient and the topological overlap are calculated.

EXAMPLE 12.3 *Topological overlap of a small network.*

Results of the topological overlap for the network in Figure 12.5 can be written in a matrix notation:

$$\begin{array}{ccccccccc}
O(i,j) = & & & & & & & & \\
& A & B & C & D & E & F & G & H & I & J & K \\
\\
A & 0 & \frac{2}{2} & 1 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (12.6) \\
B & \frac{2}{2} & 0 & 1 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
& \cdots & & & & & & & & & &
\end{array}$$

Nodes with similar entries in the matrix rows can be grouped together for analysis. This can be done in a cluster analysis. The individual cluster coefficients for the nodes are as follows: $A : 1$, $B : 1$, $C : 1/3$, $D : 0$, $E : 1/3$, $F : 1$, $G : 1/3$, $H : 1/3$, $I : 1/3$, $J : 2/3$, $K : 1$.

12.3 Formation of Scale Free Networks

As biochemical networks possess a scale free property, the question arises which mechanism leads to its formation. In the literature mentioned above, two essential processes are identified: (i) Growing networks: From an evolutionary perspective, biochemical networks are constantly changing and growing. This means that new components are formed in the network, e.g., by random duplication of a component during cell division or by newly established metabolic pathways. (ii) Preferential connections: Nodes with many connections are preferred for connecting with new elements. Such connections are similar to those of computer networks. A newly bought computer must be connected to a server (a hub) at the beginning and subsequently with a few other computers. Figure 12.6 illustrates this process.

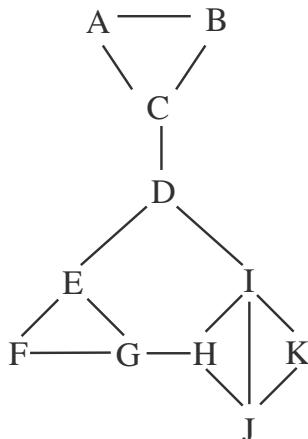


FIGURE 12.5: Example of topological overlap.

Plot **A** shows nodes with a different number of connections. Hence, the following results for the nodes:

Node 1 : 3 edges,	Node 2 : 5 edges,	Node 3 : 3 edges
Node 4 : 4 edges,	Node 5 : 3 edges,	Node 6 : 2 edges.

The network grows when a new node is added. **B** shows the probability of a connection to an existing node. A uniformly distributed random number is drawn from the interval $[0, 1]$. Hence, nodes with many connections have a higher probability of receiving the new connection. In the network one to five links are given; therefore, we divide the interval into 15 parts (sum from 1 to 5). For a node with 5 links the probability to get a new link is therefore $p = 5/15 = 1/3$ while for a node with only a single link $p = 1/15$. Node 1 in **C** has 3 connections. Thus, the probability of the node making a new connection is $p = 1/5$. In **D** node 5 was duplicated. All connections to other nodes were inherited. This also influences the connectivity of the network and hence also the connections of new nodes.

Bibliography

- [1] A.-L. Barabasi and Z. N. Oltvai. Network biology: understanding the cells functional organization, *Nat. Rev. Genetics* 5, 2004, 101-114.
- [2] E. Ravasz et al. Hierarchical organization of modularity in metabolic networks, *Science* 297, 2002, 1551-1555.

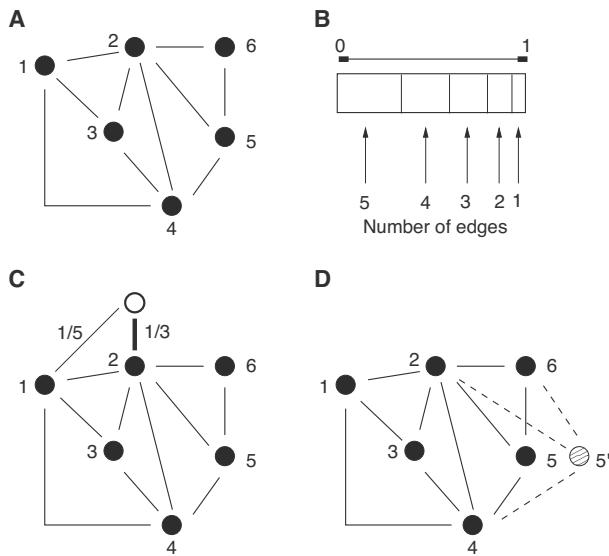


FIGURE 12.6: **A** Graph. The nodes have a different number of connections. **B** Depending on the number of nodes a new connection to the existing nodes is formed. The probability of a new connection depends on existing connections. **C** A connection of a new node (open circle) to node 2 has the highest probability while the probability to connect to node 1 is smaller. **D** During protein duplication, information about the connections is also duplicated.

This page intentionally left blank

Appendix

This page intentionally left blank

Appendix A

Collection of mathematical approaches

The Appendix summarizes same mathematical materials not fully covered in the main text.

Rules for vector derivatives

Given is a constant vector \underline{a} and a vector with variables \underline{x} . The scalar product is given by:

$$p = \underline{a}^T \underline{x}. \quad (\text{A.1})$$

The following rules are valid:

$$\frac{dp}{d\underline{x}} = \frac{d(\underline{a}^T \underline{x})}{d\underline{x}} = \underline{a} \quad \text{column vector} \quad (\text{A.2})$$

and

$$\frac{d(\underline{x}^T \underline{a})}{d\underline{x}} = \underline{a}. \quad (\text{A.3})$$

Having two products one gets after algebraic manipulations:

$$\frac{d(\underline{x}^T A \underline{x})}{d\underline{x}} = (A + A^T) \underline{x}. \quad (\text{A.4})$$

If A is symmetric:

$$\frac{d(\underline{x}^T A \underline{x})}{d\underline{x}} = 2A \underline{x}. \quad (\text{A.5})$$

Considering a column vector $f(\underline{x})$, for example, from balance equations with n rows, the derivative of the vector with respect to the column vector $\underline{x} = (x_1 \ x_2 \ \cdots \ x_m)^T$ with m entries results in a matrix, often also called a

Jacobian matrix that has dimension $n \times m$:

$$\frac{df(\underline{x})}{d\underline{x}} = \begin{pmatrix} \frac{df_1}{dx_1} & \frac{df_1}{dx_2} & \dots & \frac{df_1}{dx_m} \\ \frac{df_2}{dx_1} & \frac{df_2}{dx_2} & \dots & \frac{df_2}{dx_m} \\ \dots & \dots & \dots & \dots \\ \frac{df_n}{dx_1} & \frac{df_n}{dx_2} & \dots & \frac{df_n}{dx_m} \end{pmatrix}. \quad (\text{A.6})$$

Based on this definition, the following rules are valid for matrices:

$$\frac{d(A \underline{x})}{d\underline{x}} = A \quad (\text{A.7})$$

and

$$\frac{d(\underline{x}^T A)}{d\underline{x}} = A. \quad (\text{A.8})$$

In the first case, one gets a column vector and calculates the derivative; in the second case one gets a row. Therefore, in the first case, the derivative leads to a row vector corresponding to the number of elements. Let's make an example: Matrix A is given as:

$$A = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \rightarrow \frac{d(A \underline{x})}{d\underline{x}} = \frac{d}{d\underline{x}} \begin{pmatrix} x_1 + x_2 \\ 2x_1 - x_2 \end{pmatrix}. \quad (\text{A.9})$$

The derivative of a column vector with respect to a column vector results in a matrix where the first line shows the derivatives of the first entry with respect to the variables:

$$\frac{d}{d\underline{x}} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} = \frac{d}{d\underline{x}} \begin{pmatrix} x_1 + x_2 \\ 2x_1 - x_2 \end{pmatrix} = \begin{pmatrix} \frac{df_1}{dx_1} & \frac{df_1}{dx_2} \\ \frac{df_2}{dx_1} & \frac{df_2}{dx_2} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}. \quad (\text{A.10})$$

Now we consider

$$A = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \rightarrow \frac{d(\underline{x}^T A)}{d\underline{x}} = \frac{d}{d\underline{x}} \begin{pmatrix} x_1 + 2x_2 & x_1 - x_2 \end{pmatrix}. \quad (\text{A.11})$$

Here the derivatives are:

$$\frac{d}{d\underline{x}} \begin{pmatrix} f_1 & f_2 \end{pmatrix} = \frac{d}{d\underline{x}} \begin{pmatrix} x_1 + 2x_2 & x_1 - x_2 \end{pmatrix} = \begin{pmatrix} \frac{df_1}{dx_1} & \frac{df_2}{dx_1} \\ \frac{df_1}{dx_2} & \frac{df_2}{dx_2} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}. \quad (\text{A.12})$$

Linear Algebra

A central problem in the determination of flux maps is the solution space of linear equation systems given as:

$$A \underline{x} = \underline{b}. \quad (\text{A.13})$$

A special case is given when \underline{b} is the null vector. In this case, all vectors that fulfill the equation represent the null space of matrix A .

Rank of a matrix

The rank r of a matrix is the number of the linear independent rows/columns of the matrix (the rank with respect to the rows and the rank with respect to the columns is always the same). Linear independent means that you will find no values γ_1, γ_2 that solve the equation for the three vectors \underline{x}_i :

$$\gamma_1 \underline{x}_1 + \gamma_2 \underline{x}_2 = \underline{x}_3. \quad (\text{A.14})$$

Null space of a matrix

If a matrix A with n rows has full rank with respect to rows, that is $r = n$, the row vectors can be used as basis vectors of \mathbb{R}^n . If $r = n - 1$, a complete basis is not possible and one seeks an additional vector that can complement the $n - 1$ independent vectors to form a basis. This special vector is orthogonal to all row vectors and is called null space.

Example: A matrix with three rows is considered:

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{pmatrix}. \quad (\text{A.15})$$

The matrix has rank $r = 2$ (this can easily be seen, since the sum of the first and second row is the third row). The null space is:

$$K = \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}. \quad (\text{A.16})$$

K is orthogonal to all rows; that is, the scalar product gives zero. Using two rows and in addition vector K , all three vectors build a basis of \mathbb{R}^3 .

Now we consider an equation system representing a network and we are interested in the flux map. A special case emerges when considering only reactions inside the system but not the exchange fluxes with the environment A.1. In this case, the null space vectors represent cycles through the network.

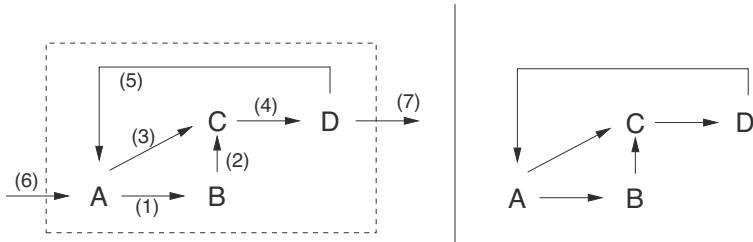


FIGURE A.1: Left: Open system with inflow and outflow. Right: Closed system.

For the network on the left side, the stoichiometric matrix reads:

$$N = \begin{pmatrix} -1 & 0 & -1 & 0 & 1 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & -1 \end{pmatrix}. \quad (\text{A.17})$$

The dimension of the null space is $q - \text{Rang}(N) = 7 - 4 = 3$ and the entries are as follows:

$$K = \begin{pmatrix} -1 & 1 & 1 \\ -1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}. \quad (\text{A.18})$$

The first vector describes a cycle with rates $-r_1, -r_2$ and r_3 ; the second one a cycle with rates r_1, r_2, r_4 and r_5 and the third one describes a flux distribution including the fluxes over systems' borders: r_1, r_2, r_4, r_6 and r_7 .

If one considers the reduced system on the right side with 5 rates, the dimension of the null space is $q' - \text{Rang}(N') = 5 - 3 = 2$ and the null space

in this case K' is:

$$K' = \begin{pmatrix} -1 & 1 \\ -1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}. \quad (\text{A.19})$$

One sees that only the two cycles remain. A special case is a feedback loop. Here it is required that all entries are positive. In the example only the second vector represents therefore a feedback loop. Note that a linear combination of the vectors (just the sum of the first and second column) also represents a feedback loop.

Matrix factorization

There are several possibilities to factorize a matrix A with N rows and K columns:

- (i) Singular value decomposition

$$A = U \cdot S \cdot V^T, \quad (\text{A.20})$$

where U is a $N \times N$, and V^T a $K \times K$ matrix. Matrix S ($N \times K$) contains the K singular values and has the structure:

$$S = \begin{bmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & 0 & \cdots \\ 0 & 0 & s_3 & \cdots \\ 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}, \quad (\text{A.21})$$

with $s_1 > s_2 > \cdots > 0$.

- (ii) QR factorization

$$A = Q \cdot R \quad (\text{A.22})$$

with

$$Q Q^T = I \quad (\text{A.23})$$

and R is a upper triangle matrix. Q is a $N \times N$ matrix and R is a $N \times K$ matrix. For $N > K$ the decomposition is unique.

(iii) LR factorization

The quadratic matrix A can be factorized as follows:

$$A = L \cdot R, \quad (\text{A.24})$$

with L a lower triangle matrix and R a upper triangle matrix.

(iv) Schur factorization

For square (real) matrices a Schur factorization looks like:

$$R = U^T \cdot A \cdot U \quad \text{with} \quad U^T U = I, \quad (\text{A.25})$$

where R is again a upper triangle matrix.

Pseudoinverse

We consider the problem of solving a linear equation system with the $n \times q$ matrix A :

$$A \underline{x} = \underline{b} \quad (\text{A.26})$$

and we distinguish two cases:

(i) Number of columns q is larger than the number of rows n . Matrix A has full row rank; that is, all rows are linear independent. Then

$$A \underline{x} = (A A^T) (A A^T)^{-1} \underline{b} = A \{A^T (A A^T)^{-1}\} \underline{b}. \quad (\text{A.27})$$

Since $A A^T$ has full rank we can take the inverse and we get by comparison:

$$\underline{x} = A^T (A A^T)^{-1} \underline{b} = A^\# \underline{b}. \quad (\text{A.28})$$

(ii) Number of rows is larger than the number of columns; for example, we have more measurements than unknowns. Matrix A has full column rank.

$$\underline{x} = (A^T A)^{-1} A^T \underline{b}. \quad (\text{A.29})$$

Statistics and probability theory

For some modeling aspects, for example, enzymatic reactions, the number of possible conformations plays an important role. Here, the binomial coefficient comes into play:

$$\binom{n}{k} = \frac{n!}{(n-k)! k!}. \quad (\text{A.30})$$

It gives the number of possible subsets with k elements from which a larger set of n elements can be obtained. Note that the order of the k elements is neglected here.

Summation of binomial coefficients

$$\sum_{i=0}^j \binom{a}{i} \binom{b}{j-i} = \binom{a+b}{j} + \cdots + \binom{a}{j} \binom{b}{0} = \binom{a+b}{j}. \quad (\text{A.31})$$

Random variables and distributions

A random variable is a real variable that can be assigned different values after an experiment. For a discrete variable, $P = p_j$ is the probability that the event x_i occurs:

$$P(X = x_i) = p_i. \quad (\text{A.32})$$

To determine p_i a number of experiments must be performed.

For continuous processes a density function f is used to describe the probabilities (probability density function **pdf**). It can be used to determine the probability that random variable X can be found in the interval a, b :

$$P(a \leq X < b) = \int_a^b f(x) dx = F(b) - F(a). \quad (\text{A.33})$$

The function $F(x)$ (cumulative density function **cdf**) is the integral over f .

$$F(x) = P(X < x). \quad (\text{A.34})$$

Characteristic values for distributions are the mean μ , variance σ^2 and the standard deviation σ .

Two important distributions are the normal distribution:

$$P_n = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}}, \quad (\text{A.35})$$

and the χ^2 distribution:

$$P_{\chi^2} = \sum_{i=1}^m x_i^2; \quad (\text{A.36})$$

here, x is normally distributed.

The following figures show the pdf and the cdf of the two distributions.

Rules for mean and variance

The rules are as follows:

$$\text{mean} \quad \bar{x} = E[x] \quad (\text{A.37})$$

$$\text{variance} \quad Var(x) = \sigma^2 = E[(x - \bar{x})^2]. \quad (\text{A.38})$$

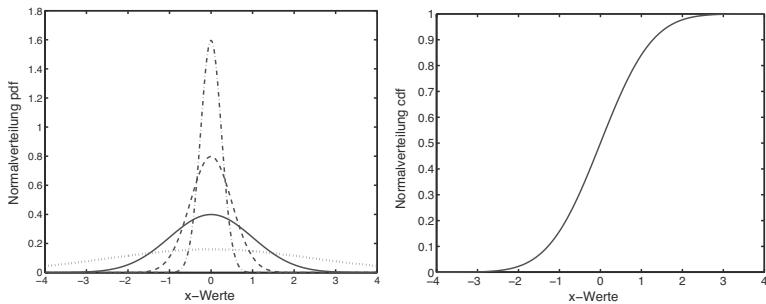


FIGURE A.2: Left: Pdf. Right: Cdf of a normal distribution.

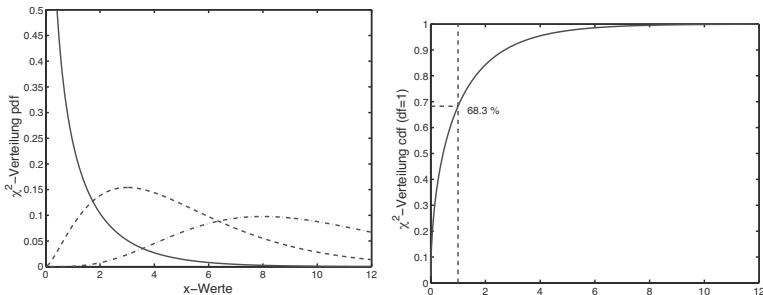


FIGURE A.3: Left: Pdf. Right: Cdf of a χ^2 distribution with $df = 1$.

For matrices and vectors:

$$\text{Var} [A \cdot x] = A \cdot \text{Var} [x] \cdot A^T. \quad (\text{A.39})$$

where a random variable x is the sum or difference of two other random variables a and b , one gets:

$$E[a \pm b] = E[a] \pm E[b] \quad (\text{A.40})$$

$$\begin{aligned} \text{Var}(a \pm b) &= E[(a \pm b - (\bar{a} \pm \bar{b}))^2] \\ &= \text{Var}(a) + \text{Var}(b) \pm 2 \text{Cov}(a, b) \\ &= \text{Var}(a) + \text{Var}(b) \pm 2(E[a b] - \bar{a} \bar{b}). \end{aligned} \quad (\text{A.41})$$

Conditional probability

For setting up model equations for stochastic systems, the understanding of conditional probabilities is a prerequisite. In contrast to the definition of a probability above, a conditional probability requires at least a two step

procedure. So, we are interested in an event A with a precondition that event B already happened. This is written as $P(A|B)$. To find an expression for the conditional probability the following relationship can be used:

$$P(A|B) = \frac{P(B \wedge A)}{P(B)}. \quad (\text{A.42})$$

This implies that we can use the conditional probability to calculate the probability that two events A and B occur:

$$P(B \wedge A) = P(A|B) P(B). \quad (\text{A.43})$$

Dynamic systems

To understand the behavior of cellular networks much better, the temporal dynamics of such systems is analyzed. Based on the non linear nature of the system the behavior is hardly predictable. Therefore, simplifications must be performed to allow a tractable analysis. One of such techniques is based on a linearization of the dynamical equations with respect to the state variables on a selected point of operation. The Taylor series is cut after the first derivative, so a linear system that is only valid in the neighborhood of the operating point is obtained.

Approximation of a function by a Taylor series

Every function that can be differentiated can be approximated by a Taylor series. The series has an infinite number of summands:

$$f(x) = \sum_{k=0}^{\infty} \frac{1}{k!} f^k(x_0) (x - x_0)^k \quad (\text{A.44})$$

with f^k as the k th derivative; writing all terms one gets:

$$f(x) = f(x_0) + \frac{df}{dx}(x - x_0) + \frac{1}{3!} \frac{d^2f}{dx^2}(x - x_0)^2 + \frac{1}{3!} \frac{d^3f}{dx^3}(x - x_0)^3 + \dots \quad (\text{A.45})$$

Cutting the series after the first derivative leads to a linearized equation:

$$f(x) = f(x_0) + \left. \frac{df}{dx} \right|_{x=x_0} (x - x_0). \quad (\text{A.46})$$

The derivatives must be calculated at points x_0 (operating point). For a system with two variables one obtains:

$$f(x) = f(x_{10}, x_{20}) + \frac{df}{dx_1}(x_1 - x_{10}) + \frac{df}{dx_2}(x_2 - x_{20}) \quad (\text{A.47})$$

where the operating point is characterized by $x_1 = x_{10}$ and $x_2 = x_{20}$.

In some cases, one is interested in the influence of a parameter x on the equation. In this case one calculates the alteration of the equation for a small Δx . Also in this case, a Taylor series can be used. With $x = x_0 + \Delta x$ one obtains:

$$f(x_0 + \Delta x) = f(x_0) + \frac{df}{dx} \Big|_{x=x_0} (x_0 + \Delta x - x_0) = f(x_0) + \frac{df}{dx} \Big|_{x=x_0} \Delta x. \quad (\text{A.48})$$

To be more flexible (without taking the operating point explicitly into account) the following equation is often used:

$$f(x + \Delta x) = f(x) + \frac{df}{dx} \Delta x. \quad (\text{A.49})$$

In this form, the equation is used for a sensitivity analysis.

A further special situation emerges when dynamical systems in the form of differential equations are considered. Here, one is interested in the behavior of the system near a steady state. If the system is given by

$$\dot{x} = f(x) \quad (\text{A.50})$$

a linearization with respect to the steady state x_0 (that is $f(x_0) = 0$) leads to a new equation that is valid in a neighborhood x' of the steady state:

$$\dot{x}' = \frac{df}{dx} \Big|_{x=x_0} x' = a x' \quad (\text{A.51})$$

with a as the derivative of f with respect to x and x_0 inserted. For $a < 0$ the steady state is a stable one. The system will come back to the steady state when it is disturbed while for $a > 0$ it is unstable.

Jacobian matrix

The Jacobian matrix is a special mathematical structure playing a central role in the analysis of dynamical systems. It is the result of a linearization of a nonlinear equation system given by:

$$\dot{\underline{x}} = \underline{f}(\underline{x}) \quad (\text{A.52})$$

with the following derivatives:

$$J = \begin{bmatrix} \frac{df_1}{dx_1} & \frac{df_1}{dx_2} & \dots \\ \frac{df_2}{dx_1} & \frac{df_2}{dx_2} & \dots \\ \frac{df_3}{dx_1} & \frac{df_3}{dx_2} & \dots \end{bmatrix}, \quad (\text{A.53})$$

that must be evaluated at steady state given by $\underline{x} = \underline{x}_0$.

Eigenvalues for systems of order 2

A system with two linear differential equations is given by:

$$\dot{\underline{x}} = A \underline{x} + \underline{b} u = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \underline{x} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} u, \quad (\text{A.54})$$

where A is the Jacobian matrix (see above). An analysis of the stability of the system requires the determination of the eigenvalues of matrix A . This can be done via the characteristic equation:

$$\lambda^2 - \text{trace}(A) \lambda + \det(A) = 0 \quad (\text{A.55})$$

with trace of matrix A : $\text{trace}(A) = a_{11} + a_{22}$ and determinant of A : $\det(A) = a_{11} a_{22} - a_{12} a_{21}$. The solution of this quadratic equation is given by:

$$\lambda_{1/2} = \frac{\text{trace}}{2} \pm \sqrt{\frac{\text{trace}^2}{4} - \det}. \quad (\text{A.56})$$

There are several cases that can be analyzed for the solution that are shown in Figure A.4.

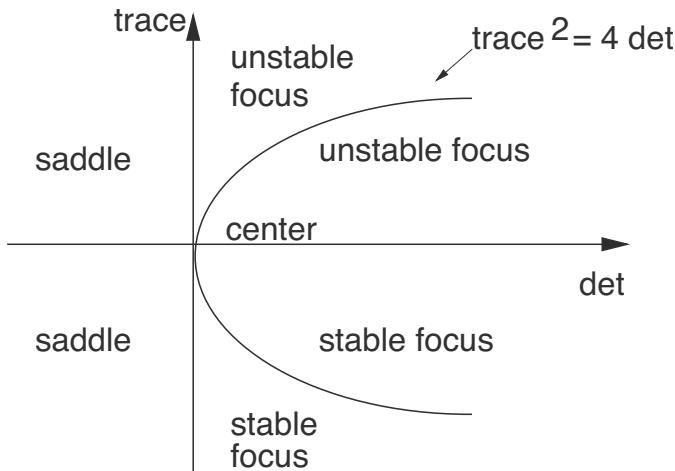


FIGURE A.4: Stability diagram for linear system of order 2.

Laplace transformation

Motivation: for many problems in engineering sciences one is interested in the behavior of systems that are coupled together from individual entities, for example, a controller and a controlled plant. To analyze the system, special

input functions like steps or pulses are used. For linear systems, the dynamic response of coupled sub-systems can easily be computed using Laplace transformation. Here, differential equations are replaced by algebraic equations.

The Laplace transformation is an integral transformation where a time continuous function $f(t)$ is transformed into the frequency domain $F(s)$. The formal definition is given by:

$$F(s) = \int_0^{\infty} e^{-st} f(t) dt \quad \text{with} \quad s = \sigma + i\omega. \quad (\text{A.57})$$

Derivatives and integral calculations can be done with the following formulas:

$$\text{Derivative} \quad \frac{dx}{dt} \longrightarrow sX - x(0^+) \quad (\text{A.58})$$

$$\text{Integration} \quad \int x dt \longrightarrow \frac{1}{s} X. \quad (\text{A.59})$$

Furthermore, input signals can be transformed also easily:

$$\text{step input} \quad f(t) = \begin{cases} 0 & \text{for } t < 0 \\ 1 & \text{for } t \geq 0 \end{cases} \longrightarrow \frac{1}{s} \quad (\text{A.60})$$

$$\text{puls} \quad f(t) = \begin{cases} 0 & \text{for } t < a \\ A & \text{for } a \geq t \geq b \\ 0 & \text{for } t > b \end{cases} \longrightarrow \frac{A(e^{-as} - e^{-bs})}{s}. \quad (\text{A.61})$$

An important analysis is the calculation of steady states that can be done also in the frequency domain:

$$\lim_{t \rightarrow \infty} f(t) = \lim_{s \rightarrow 0} s F(s). \quad (\text{A.62})$$

Figure A.5 shows in which way the individual transformed functions can be combined. Every block represents one or more differential equations that reflects the input/output behavior. If one has calculated the resulting function in the frequency domain one can get back to the time domain using several tables published in the literature for the back transformation.

Example: The right scheme shows a feedback from the output back to the input. The differential equations are as follows:

$$\text{Block } F_1: \quad \dot{y} = -y + e, \quad y(0) = 0 \quad (\text{A.63})$$

$$\text{Block } F_2: \quad g = 2y. \quad (\text{A.64})$$

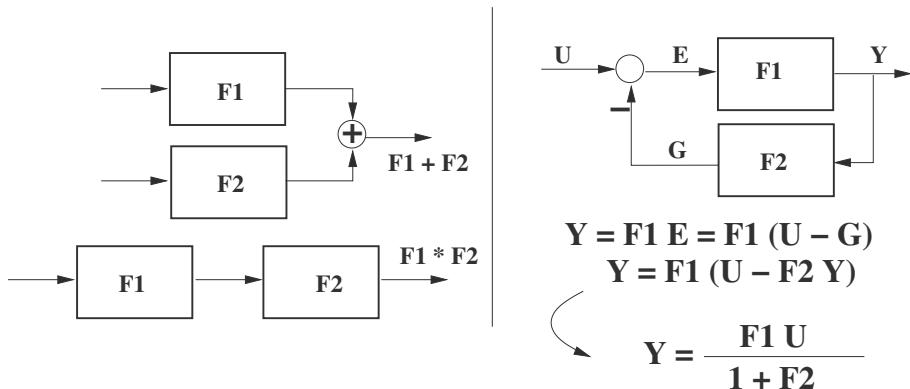


FIGURE A.5: Individual elements can be combined using algebraic manipulations. F_j represents the respective transfer functions, Y the output and U the input signal.

Laplace transformation leads to:

$$\text{Block } F_1: \quad sY = -Y + E \quad \longrightarrow \quad Y = \underbrace{\frac{1}{1+s}}_{F_1} E \quad (\text{A.65})$$

$$\text{Block } F_2: \quad G = \underbrace{\frac{2}{1+2}}_{F_2} Y. \quad (\text{A.66})$$

Using the formula from above, one obtains:

$$Y = \frac{F_1}{1+F_2} U = \frac{1}{1+s} \frac{1}{1+2} U = \frac{1}{3(1+s)} U. \quad (\text{A.67})$$

If one is interested in the response for a step input ($U = 1/s$), the solution in the frequency domain is Y , which can be transformed back in the time domain $y(t)$:

$$Y(U = 1/s) = \frac{1}{3(s+s^2)} \quad \longrightarrow \quad y(t) = \frac{(1-e^{-t})}{3} \quad (\text{A.68})$$

The frequency domain also allows different possibilities for model analysis. In the frequency domains one considers the response of the system if it is stimulated by a sinusoidal function with a given frequency. A linear system is characterized by the fact that the output shows the same frequency; however a phase shift and amplification/damping of the amplitude occur (see next figure). Amplitude and phase give important information on system's dynamics. In the figure amplitude and phase for varying frequencies are shown for the differential equation $\ddot{y} + \dot{y} + y = 1$. Such dynamics are characterized as filters since some frequencies are amplified while others are damped.

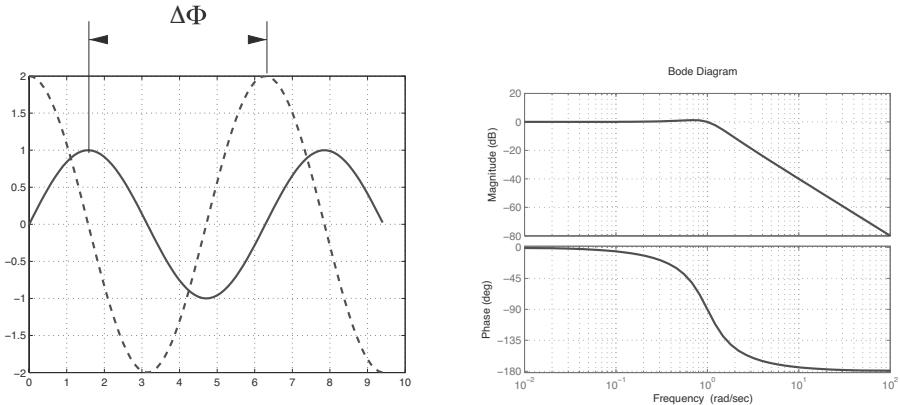


FIGURE A.6: Left: A linear system shows a different amplitude and a phase shift if it is stimulated with a certain frequency ω . Right: Amplitude and phase shift for varying frequencies for the example $\ddot{y} + \dot{y} + y = 1$. For high frequencies, the system is damped.

Linear Optimization

Each constrained polytope M is the convex span of its corners. If the m corners \underline{x}_j are known, the polytope can be presented as follows:

$$M = \{\underline{x} \in \mathbb{R}^n \mid \underline{x} = \sum_{j=1}^m \lambda_j \underline{x}_j\} \quad (\text{A.69})$$

for $\lambda_j \geq 0$ with $\sum_{j=1}^m \lambda_j = 1$

where \underline{x}_j are M 's corners.

Figure 11.7 illustrates an example:

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \text{ and } \underline{b} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \quad (\text{A.70})$$

The first equation presents a plane; together with the second equation, the intersection with $x_2 = 0$ -plane. The polyhedron has two corners that can be described by the coordinates

$$\underline{x}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \text{ and } \underline{x}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \quad (\text{A.71})$$

The polyhedron is spanned by both vectors, if λ varies between zero and one.

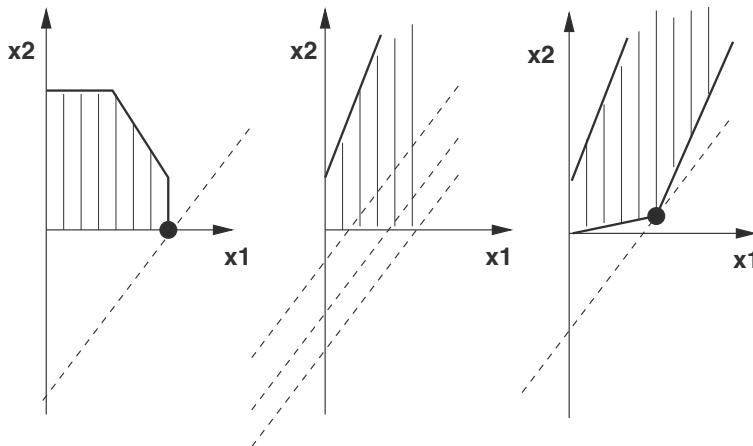


FIGURE A.7: On the left side: Polyhedron with solution to the optimization problem. In the center: Open polyhedron with no solution. On the right side: Open polyhedron with one solution.

The following **proposition** is imperative: $M \neq \emptyset$ is a closed polyhedron and $f = \underline{c}^T \underline{x}$ a linear objective function. Then, the optimization has a solution that is on the corner of M .

Nonrestrained polyhedrons may have a solution in some cases. This is shown in Figure A.7. A 2-dimensional plane is considered, which presents the feasible region (inequality constraints apply here). The objective function is linear and occurs as a straight line with an intercept z that should be minimized. The objective function is:

$$\min \Phi = x_2 - x_1 \rightarrow x_2 = x_1 + z. \quad (\text{A.72})$$

In the first case shown on left in the figure, the solution can be determined graphically by shifting the line down until it intersects with a corner of the feasible region. In the second case a construction of the feasible region is not possible. In the third case it is.

To calculate the optimal values of the objective function it is necessary to determine the corners. This can be done as follows: A is a $n \times q$ matrix and $\text{rank}(A) = n$. Each choice of n linearly independent column of A forms a basis of A ; the corresponding matrix is called a basis matrix. A solution of the equation system is called a basis solution if all non basis variables equal zero. The choice of columns presents degrees of freedom. There are as many linearly independent columns as rows, as the matrix that results from the columns is

invertible. The values in vector \underline{x} must be sorted correspondingly:

$$\text{with } A = [A_B \quad A_N] \rightarrow A \underline{x} = [A_B \quad A_N] \begin{pmatrix} \underline{x}_B \\ \underline{x}_N \end{pmatrix} = \underline{b}. \quad (\text{A.73})$$

Thus, A_B is a $n \times n$ matrix and A_N is a $n \times (q - n)$ matrix. From this equation, the values for \underline{x}_B with n elements can be determined:

$$\underline{x}_B = A_B^{-1} \underline{b} + A_B^{-1} A_N \underline{x}_N. \quad (\text{A.74})$$

The first part of the solution ($\underline{x}_N = 0$) is the basis solution:

$$\boxed{\underline{x}_B = A_B^{-1} \underline{b}.} \quad (\text{A.75})$$

The basis solution can be understood as those conditions that must be at least fulfilled to solve the equation system. **Proposition:** M is polyhedron $M = \{x | A x = b, x \geq 0\}$. Point $\underline{x} \in \mathbb{R}^n$ is then a corner point in M , if \underline{x} is a basis solution of the system $A \underline{x} = \underline{b}$ and $\underline{x} \geq 0$. The complete solution, now including $(q - n)$ degrees of freedom, is as below, if freedom is expressed using a parameter vector $\underline{\lambda}_N$ with $q - n$ elements.

$$\underline{x} = \begin{pmatrix} \underline{x}_B \\ \underline{x}_N \end{pmatrix} = \begin{pmatrix} \underline{x}_B \\ 0_N \end{pmatrix} + \begin{pmatrix} -A_B^{-1} A_N \\ I_N \end{pmatrix} \underline{\lambda}_N \quad (\text{A.76})$$

$$= \underline{x} - W_N \underline{\lambda}_N \quad (\text{A.77})$$

$$\text{with } W_N = \begin{pmatrix} A_B^{-1} A_N \\ -I_N \end{pmatrix}.$$

This solution can be entered directly in the objective function. Together with the equation from above, one obtains:

$$f : \underline{c}^T \underline{x} = \begin{pmatrix} \underline{c}_B^T & \underline{c}_N^T \end{pmatrix} \begin{pmatrix} \underline{x}_B \\ \underline{x}_N \end{pmatrix} = \underbrace{\underline{c}_B^T \underline{x}_B}_{\text{value of obj.}} + \underbrace{(\underline{c}_N^T - \underline{c}_B^T A_B^{-1} A_N)}_{\underline{\gamma}_N^T} \underline{x}_N. \quad (\text{A.78})$$

The relationship is as follows:

$$\underline{\gamma}_N^T = -\underline{c}^T \cdot W_N. \quad (\text{A.79})$$

If now $\underline{\gamma}_N^T \leq 0$ (i.e., all values of the vector are smaller than zero), then the basis solution \underline{x}_B is the optimal solution to the problem. This means that the degrees of freedom of the objective function to be maximized lead to

smaller values. If the values for γ_N are negative, then values for f are smaller than the values for the chosen corner. If the value is greater than zero, then there is another corner that has a greater value for f , as the solution is always represented by one corner only. This is applied in the numerical solution of linear optimization problems, for example, in SIMPLEX-algorithms.

This page intentionally left blank

This page intentionally left blank

"**Systems Biology: Mathematical Modeling and Model Analysis** is a rich resource of mathematical methods and approaches that can be utilized to analyze and understand biological systems. It will be particularly attractive to engineers and mathematicians who want to learn the basics of modern biology in a condensed fashion and then apply the tools of their trades to relevant biological questions. ... **Systems Biology** devotes a good portion of the material to state-of-the-art model diagnostics and engineering techniques, such as linear systems analysis and control theory, which so far are rarely found in systems biology texts and are therefore a welcome addition to the repertoire of textbook literature. ... Laudably, all topics are illustrated with step-by-step examples, and many of them are reinforced with exercises."

—Eberhard Voit, Georgia Institute of Technology

"... a meticulous and far-reaching treatment of this critically important subject. ... comprehensive without sacrificing depth ... an exhaustive survey of modeling approaches. Each chapter is thoughtfully crafted to draw in novices to the field while still engaging to experts. Additionally, a number of well-designed exercises complement each chapter."

—Markus Covert, Stanford University

Drawing on the latest research in the field, this book presents many methods for modeling and analyzing biological systems, in particular cellular systems. It shows how to use predictive mathematical models to acquire and analyze knowledge about cellular systems. The text also explores how the models are systematically applied in biotechnology.

Features

- Presents methods for the analysis of motifs, modules, and large-scale networks
- Explains how deterministic models and graphs can be used in the analysis and verification of networks
- Shows how to model the complete synthesis of macromolecules and compare measured data
- Uses the lactose uptake system to demonstrate various modeling and analysis tools
- Contains many analytical and numerical examples



CRC Press
Taylor & Francis Group
an informa business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

K16367

ISBN: 978-1-4665-6789-4



9 781466 567894