Edda Klipp
Wolfram Liebermeister
Christoph Wierling
Axel Kowald

Systems Biology

Edda Klipp
Wolfram Liebermeister
Christoph Wierling
Axel Kowald

Systems Biology

A Textbook

Second, Completely Revised
and Enlarged Edition

WILEY-VCH
Verlag GmbH & Co. KGaA

Authors

Prof. Dr. h.c. Edda Klipp
Theoretical Biophysics
Humboldt-Universität zu Berlin
Invalidenstr. 42
10115 Berlin
Germany

Dr. Wolfram Liebermeister
Institute of Biochemistry
Charité - Universitätsmedizin Berlin
Charitéplatz 1
10117 Berlin
Germany

Dr. Christoph Wierling
Alacris Theranostics GmbH
Fabeckstr. 60-62
14195 Berlin
Germany

and

Max Planck Institute for Molecular Genetics
Ihnestr. 63-73
14195 Berlin
Germany

Dr. Axel Kowald
Theoretical Biophysics
Humboldt University Berlin
Invalidenstr. 42
10115 Berlin
Germany

Cover

Cover design by Wolfram Liebermeister. The cover picture was provided with kind permission by Jörg Bernhardt.

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

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

Print ISBN: 978-3-527-33636-4
ePDF ISBN: 978-3-527-67566-1
ePub ISBN: 978-3-527-67567-8
Mobi ISBN: 978-3-527-67568-5

Typesetting Thomson Digital, Noida, India

Printed on acid-free paper

Contents

Guide to Different Topics of the Book

Biological systems and processes

Metabolism (2, 3, 4, 11.1, 11.4, 12.1)
Gene regulatory network (7, 8.2, 9)
Gene expression regulation (2, 9)
Signaling systems (8.2, 12.2)
Cell cycle (12.3)
Development (7.3)
Aging (12.4)

Concepts for biological function

Qualitative behavior (3, 7.1)
Parameter sensitivity and robustness (4.2, 10.2)
Modularity and functional subsystems (6.4, 8.3)
Robustness against failure (10.2)
Information (10.3, 15.6)
Population heterogeneity (10.3)
Optimality (3.2, 11.1, 11.2)
Evolution (11.3)
Population dynamics and game theory (11.4)

Model types with different levels of abstraction

Statistical particle models (15.6)
Stochastic biochemical models (5.1.2, 7.2, 9.4, 15.4)
Kinetic models (4, 5.1.1, 11, 12, 16.7)
Constraint-based models (3.2)
Discrete models (7.1)
Spatial models (7.3)

Mathematical frameworks to describe cell states

Topological (network structures) (3.1, 8)
Structural stoichiometric models (3)
Dynamical systems (4, 12, 15.2)
Deterministic linear models (6.3)
Deterministic kinetic models (4, 9.1, 12)
Uncertain parameters (10.1)
Optimization and control theory (4.2, 11.1,11.2,15.5)

Experimental Techniques

Experimental techniques (14)

Modeling skills

Model building (2, 3, 4, 5.1, 7)
Model annotation (5.4)
Parameter estimation (6.1)
Model testing and selection (6.2)
Local sensitivity analysis/control theory (4.2, 10.1, 10.2)
Global sensitivity/uncertainty analysis (10.1)
Model reduction (6.3)
Model combination (5.4, 6.4)
Network theory (8)
Statistics (15.3, 15.7)
Optimization of model outputs and structure (11.1)
Optimal temporal control (11.2, 15.5)

Practical issues in modeling

Use of databases (5.3, 16)
Data formats (5.2, 5.4)
Data sources (5.3, 16)
Modeling software (5.1, 17)
Simulation techniques and tools (5.1)
Model visualization (5.1)
Data visualization (8.3)

About the Authors

Edda Klipp (born 1965) studied biophysics at the Humboldt-Universität zu Berlin and obtained a PhD of theoretical biophysics at the HU Berlin and an honorary doctor at Gothenburg University. She is professor for theoretical biophysics at HU Berlin. A founding member of the International Society for Systems Biology, her research interests include mathematical modeling of cellular systems, application of physical principles in biology, systems biology, and development of modeling tools.

Wolfram Liebermeister (born 1972) studied physics in Tübingen and Hamburg and obtained a PhD of theoretical biophysics at the Humboldt University of Berlin. In his work on complex biological systems, he highlights functional aspects such as variability, information, and optimality.

Christoph Wierling (born 1973) studied biology in Münster and holds a PhD in biochemistry obtained from the Free University Berlin. He was leading a research group for systems biology at the Max Planck Institute for Molecular Genetics, Berlin. Currently he is heading the bioinformatics and modeling unit at Alacris Theranostics, a Berlin-based company applying NGS and systems biology approaches for translational research and personalized medicine. His research interests focus on modeling and simulation of biological systems and the development of systems biology software.

Axel Kowald (born 1963) holds a PhD in mathematical biology from the National Institute for Medical Research, London. His current research interests focus on the mathematical modeling of processes involved in the biology of aging and systems biology.

Part One
Introduction to Systems Biology

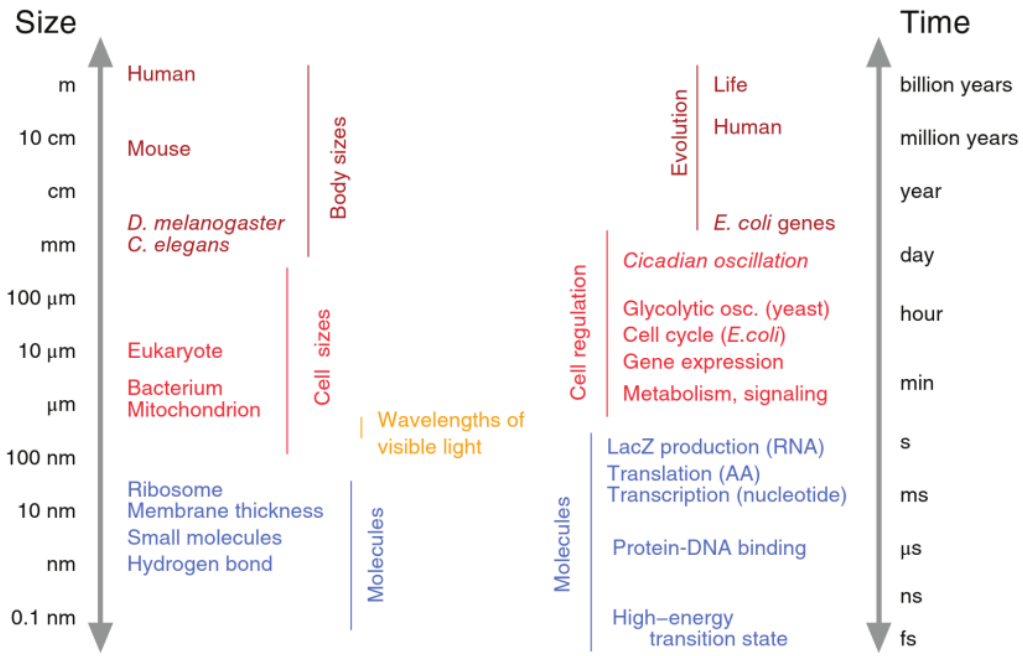


Figure 1.1 Length and time scales in biology. (Data from the BioNumbers database at bionumbers.hms.harvard.edu.)

allows for the use of model organisms and for the critical transfer of insights gained from one cell type to other cell types. Applications include, for example, prediction of protein function from similarity, prediction of network properties from optimality principles, reconstruction of phylogenetic trees, or the identification of regulatory DNA sequences through cross-species comparisons. However, the evolutionary process also leads to genetic variations within species. Therefore, personalized medicine and research is an important new challenge for biomedical research.

1.2 Models and Modeling

If we observe biological phenomena, we are confronted with various complex processes that often cannot be explained from first principles and the outcome of which cannot reliably be foreseen from intuition. Even if general biochemical principles are well established (e.g., the central dogma of transcription and translation or the biochemistry of enzyme-catalyzed reactions), the biochemistry of individual molecules and systems is often unknown and can vary considerably between species. Experiments lead to biological hypotheses about individual processes, but it often remains unclear whether these hypotheses can be combined into a larger coherent picture because it is often difficult to foresee the global behavior of a complex system from knowledge of its parts. Mathematical modeling and computer simulations can help us to understand the internal nature and dynamics of these processes and to arrive at predictions about their future development and the effect of interactions with the environment.

1.2.1 What Is a Model?

The answer to this question will differ among communities of researchers. In a broad sense, a model is an abstract representation of objects or processes that explains features of these objects or processes (Figure 1.2). A biochemical reaction network can be represented by a graphical sketch showing dots for metabolites and arrows for reactions; the same network could also be described by a system of differential equations, which allows simulating and predicting the dynamic behavior of that network. If a model is used for simulations, it needs to be ensured that it faithfully predicts the system's behavior – at least those aspects that are supposed to be covered by the model. Systems biology models are often based on well-established physical laws that justify their general form, for instance, the thermodynamics of chemical reactions. Besides this, a computational model needs to make specific statements about a system of interest – which are partially justified by experiments and biochemical knowledge, and partially by mere extrapolation from other systems. Such a model can

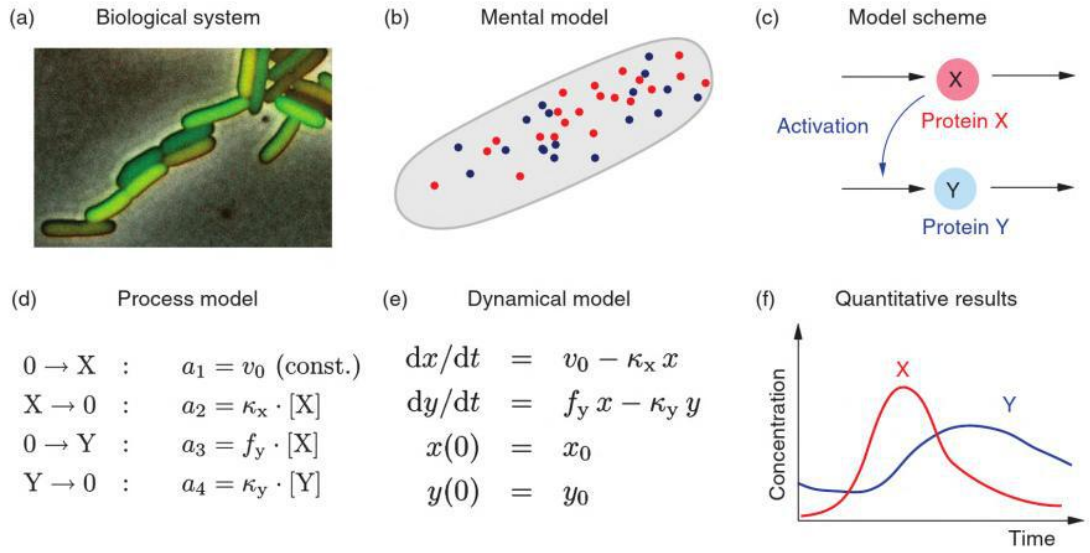


Figure 1.2 Typical abstraction steps in mathematical modeling. (a) *E. coli* bacteria produce thousands of different proteins. If a specific protein type is labeled with a fluorescent marker, cells glow under the microscope according to the concentration of this marker. (Courtesy of M. Elowitz.) (b) In a simplified mental model, we assume that cells contain two enzymes of interest, X (red) and Y (blue), and that the molecules (dots) can freely diffuse within the cell. All other substances are disregarded for the sake of simplicity. (c) The interactions between the two protein types can be drawn in a wiring scheme: each protein can be produced or degraded (black arrows). In addition, we assume that proteins of type X can increase the production of protein Y. (d) All individual processes to be considered are listed together with their rates a (occurrence per time). The mathematical expressions for the rates are based on a simplified picture of the actual chemical processes. (e) The list of processes can be translated into different sorts of dynamic models, in this case, deterministic rate equations for the protein concentrations x and y. (f) By solving the model equations, predictions for the time-dependent concentrations can be obtained. If the predictions do not agree with experimental data, this indicates that the model is wrong or too much simplified. In both cases, the model has to be refined.

summarize established knowledge about a system in a coherent mathematical formulation. In experimental biology, the term "model" is also used to denote a species that is especially suitable for experiments; for example, a genetically modified mouse may serve as a model for human genetic disorders.

1.2.2 Purpose and Adequateness of Models

Modeling is a subjective and selective procedure. A model represents only specific aspects of reality but, if done properly, this is sufficient since the intention of modeling is to answer particular questions. If the only aim is to predict system outputs from given input signals, a model should display the correct input–output relation, while its interior can be regarded as a black box. However, if instead a detailed biological mechanism has to be elucidated, then the system's structure and the relations between its parts must be described realistically. Some models are meant to be generally applicable to many similar objects (e.g., Michaelis–Menten kinetics holds for many enzymes, the promoter–operator concept is applicable to many genes, and gene regulatory motifs are common), while others are specifically tailored to one particular object (e.g., the 3D structure of a protein, the sequence of a gene, or a model of deteriorating mitochondria during aging). The mathematical part can be kept as simple as possible to allow for easy implementation and comprehensible results. Or it can be modeled very realistically and be much more complicated. None of the characteristics mentioned above makes a model wrong or right, but they determine whether a model is appropriate to the problem to be solved. The phrase "essentially, all models are wrong, but some are useful" coined by the statistician George Box is indeed an appropriate guideline for model building.

1.2.3 Advantages of Computational Modeling

Models gain their reference to reality from comparison with experiments, and their benefits therefore depend on the quality of the experiments used. Nevertheless, modeling combined with experimentation has a lot of advantages compared with purely experimental studies:

- Modeling drives conceptual clarification. It requires verbal hypotheses to be made specific and conceptually rigorous.

- Modeling highlights gaps in knowledge or understanding. During the process of model formulation, unspecified components or interactions have to be determined.
- Modeling provides independence of the modeled object.
- Time and space may be stretched or compressed *ad libitum*.
- Solution algorithms and computer programs can be used independently of the concrete system.
- Modeling is cheap compared with experiments.
- Models exert by themselves no harm on animals or plants and help to reduce ethical problems in experiments. They do not pollute the environment.
- Modeling can assist experimentation. With an adequate model, one may test different scenarios that are not accessible by experiment. One may follow time courses of compounds that cannot be measured in an experiment. One may impose perturbations that are not feasible in the real system. One may cause precise perturbations without directly changing other system components, which is usually impossible in real systems. Model simulations can be repeated often and for many different conditions.
- Model results can often be presented in precise mathematical terms that allow for generalization. Graphical representation and visualization make it easier to understand the system.
- Finally, modeling allows for making well-founded and testable predictions.

The attempt to formulate current knowledge and open problems in mathematical terms often uncovers a lack of knowledge and requirements for clarification. Furthermore, computational models can be used to test whether proposed explanations of biological phenomena are feasible. Computational models serve as repositories of current knowledge, both established and hypothetical, about how systems might operate. At the same time, they provide researchers with quantitative descriptions of this knowledge and allow them to simulate the biological process, which serves as a rigorous consistency test.

1.3 Basic Notions for Computational Models

1.3.1 Model Scope

Systems biology models consist of mathematical elements that describe properties of a biological system, for instance, mathematical variables describing the concentrations of metabolites. As a model can only describe certain aspects of the system, all other properties of the system (e.g., concentrations of other substances or the environment of a cell) are neglected or simplified. It is important – and, to some extent, an art – to construct models in such ways that the disregarded properties do not compromise the basic results of the model.

1.3.2 Model Statements

Alongside the model elements, a model can contain various kinds of statements and equations describing facts about the model elements, most notably, their temporal behavior. In kinetic models, the basic modeling paradigm considered in this book, the dynamics is determined by a set of ordinary differential equations describing the substance balances. Statements in other model types may have the form of equality or inequality constraints (e.g., in flux balance analysis), maximality postulates, stochastic processes, or probabilistic statements about quantities that vary in time or between cells.

1.3.3 System State

In dynamical systems theory, a system is characterized by its *state*, a snapshot of the system at a given time. The state of the system is described by the set of variables that must be kept track of in a model: in deterministic models, it needs to contain enough information to predict the behavior of the system for all future times. Each modeling framework defines what is meant by the state of the system. In kinetic rate equation models, for example, the state is a list of substance concentrations. In the corresponding stochastic model, it is a probability distribution or a list of the current number of molecules of a species. In a Boolean model of gene regulation, the state is a string of bits indicating for each gene whether it is expressed ("1") or not expressed ("0"). Also, the temporal behavior can be described in fundamentally different ways. In a *dynamical system*, the future states are determined by the current state, while in a *stochastic process*, the future states are not precisely predetermined. Instead, each possible future history has a certain probability to occur.

1.3.4 Variables, Parameters, and Constants

The quantities in a model can be classified as variables, parameters, and constants. A *constant* is a quantity with a fixed value, such as the natural number e or Avogadro's number (number of molecules per mole). *Parameters* are

quantities that have a given value, such as the K_m value of an enzyme in a reaction. This value depends on the method used and on the experimental conditions and may change. *Variables* are quantities with a changeable value for which the model establishes relations. A subset of variables, the *state variables*, describes the system behavior completely. They can assume independent values and each of them is necessary to define the system state. Their number is equivalent to the dimension of the system. For example, the diameter d and volume V of a sphere obey the relation $V = \pi d^3/6$, where π and 6 are constants, V and d are variables, but only one of them is a state variable since the relation between them uniquely determines the other one.

Whether a quantity is a variable or a parameter depends on the model. In reaction kinetics, the enzyme concentration appears as a parameter. However, the enzyme concentration itself may change due to gene expression or protein degradation, and in an extended model, it may be described by a variable.

1.3.5 Model Behavior

Two fundamental factors that determine the behavior of a system are (i) influences from the environment (input) and (ii) processes within the system. The system structure, that is, the relation among variables, parameters, and constants, determines how endogenous and exogenous forces are processed. However, different system structures may still produce similar system behavior (output); therefore, measurements of the system output often do not suffice to choose between alternative models and to determine the system's internal organization.

1.3.6 Model Classification

For modeling, processes are classified with respect to a set of criteria.

- A structural or *qualitative* model (e.g., a network graph) specifies the interactions among model elements. A *quantitative* model assigns values to the elements and to their interactions, which may or may not change.
- In a *deterministic* model, the system evolution through all following states can be predicted from the knowledge of the current state. *Stochastic* descriptions give instead a probability distribution for the successive states.
- The nature of values that time, state, or space may assume distinguishes a *discrete* model (where values are taken from a discrete set) from a *continuous* model (where values belong to a continuum).
- *Reversible* processes can proceed in a forward and backward direction. Irreversibility means that only one direction is possible.
- *Periodicity* indicates that the system assumes a series of states in the time interval $\{t,\ t+\Delta t\}$ and again in the time interval $\{t+i\Delta t,\ t+(i+1)\Delta t\}$ for $i = 1,2, \ldots$.

1.3.7 Steady States

The concept of stationary states is important for the modeling of dynamical systems. *Stationary states* (other terms are *steady states* or *fixed points*) are determined by the fact that the values of all state variables remain constant in time. The asymptotic behavior of dynamic systems, that is, the behavior after a sufficiently long time, is often stationary. Other types of asymptotic behavior are oscillatory or chaotic regimes.

The consideration of steady states is actually an abstraction that is based on a separation of time scales. In nature, everything flows. Fast and slow processes – ranging from formation and breakage of chemical bonds within nanoseconds to growth of individuals within years – are coupled in the biological world. While fast processes often reach a quasi-steady state after a short transition period, the change of the value of slow variables is often negligible in the time window of consideration. Thus, each steady state can be regarded as a quasi-steady state of a system that is embedded in a larger nonstationary environment. Despite this idealization, the concept of stationary states is important in kinetic modeling because it points to typical behavioral modes of the system under study and it often simplifies the mathematical problems.

Other theoretical concepts in systems biology are only rough representations of their biological counterparts. For example, the representation of gene regulatory networks by Boolean networks, the description of complex enzyme kinetics by simple mass action laws, or the representation of multifarious reaction schemes by black boxes proved to be helpful simplifications. Although being a simplification, these models elucidate possible network properties and help to check the reliability of basic assumptions and to discover possible design principles in nature. Simplified models can be used to test mathematically formulated hypotheses about system dynamics, and such models are easier to understand and to apply to different questions.

1.3.8 Model Assignment Is Not Unique

Biological phenomena can be described in mathematical terms. Models developed during the last few decades range from the description of glycolytic oscillations with

ordinary differential equations to population dynamics models with difference equations, stochastic equations for signaling pathways, and Boolean networks for gene expression. However, it is important to realize that a certain process can be described in more than one way: a biological object can be investigated with different experimental methods and each biological process can be described with different (mathematical) models. Sometimes, a modeling framework represents a simplified limiting case (e.g., kinetic models as limiting case of stochastic models). On the other hand, the same mathematical formalism may be applied to various biological instances: statistical network analysis, for example, can be applied to cellular transcription networks, the circuitry of nerve cells, or food webs.

The choice of a mathematical model or an algorithm to describe a biological object depends on the problem, the purpose, and the intention of the investigator. Modeling has to reflect essential properties of the system and different models may highlight different aspects of the same system. This ambiguity has the advantage that different ways of studying a problem also provide different insights into the system. However, the diversity of modeling approaches makes it also very difficult to merge established models (e.g., for individual metabolic pathways) into larger supermodels (e.g., models of complete cell metabolism).

1.4 Networks

The network is a crucial concept in systems biology. We study protein–protein interaction networks, protein–RNA interaction networks, metabolic networks (see Chapters 3 and 4 and Section 12.1), signaling networks (Section 12.2), guilt-by-association networks, and networks connecting gene defects with diseases or diseases with other diseases via common gene defects [1]. Throughout this book, you will find more examples.

Networks are best represented by graphs that consist of nodes and edges, which connect the nodes, as illustrated in Figure 1.3. In protein–protein interaction networks, for example, nodes are proteins and edges are their interactions as can for instance be determined by yeast two-hybrid experiments (see Chapter 14). If appropriate, one can introduce different types of nodes for different types of components. For example, the metabolites and converting enzymes in metabolic networks can be represented with bipartite networks, which possess two types of nodes – one for metabolites and the other for enzymes – that are never directly connected by an edge, but only via the other type of node. Petri net type of modeling

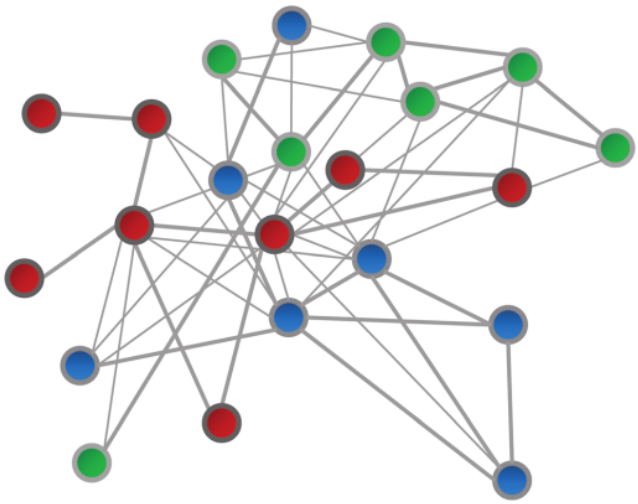


Figure 1.3 Network with nodes (circles) and edges (lines between circles). Different node colors indicate different types of connected components (e.g., proteins, mRNAs, and metabolites).

takes that property into account representing metabolites as places and enzyme-catalyzed reactions as transitions (see Section 7.1). By contrast, classical metabolic modeling considers only one type of node, but different types in different approaches. Systems of ordinary differential equations describing metabolite dynamics take metabolites as nodes and enzymatic reactions as edges (Chapter 4), while flux balance analysis restricts itself to steady states and now focusses on the fluxes through the reactions (now as nodes) that are linked by the stationary metabolites as edges.

1.5 Data Integration

Systems biology has evolved rapidly in the last few years, driven by the new high-throughput technologies. The most important impulse was given by large sequencing projects such as the Human Genome Project, which resulted in the full sequence of the human and other genomes [2,3]. Proteomic technologies have been used to identify the translation status of complete cells (2D gels, mass spectrometry) and to elucidate protein–protein interaction networks involving thousands of components [4]. However, to validate such diverse high-throughput data, one needs to correlate and integrate such information. Thus, an important part of systems biology is data integration.

On the lowest level of complexity, data integration implies common schemes for data storage, data representation, and data transfer. For particular experimental

for developing mathematical models of all parts of a cell of *E. coli* strain K-12.

1.7.2 *Saccharomyces cerevisiae*

The yeast *S. cerevisiae* is a unicellular fungus, belonging to the ascomycetes (Figure 1.4b). It is not only a useful organism needed for the production of wine, beer, and bread, but also the best studied eukaryotic model system. The cells are easy to grow and double under optimal conditions every 90–100 min. Like *E. coli*, also *S. cerevisiae* can live under aerobic as well as anaerobic conditions. If oxygen is present, the majority of energy is generated via oxidative phosphorylation at the inner mitochondrial membrane and without oxygen energy is produced via glycolysis and fermentation. The yeast normally propagates as a diploid organism via mitosis. Under stress, however, the diploid cells can undergo sporulation, producing four haploid cells in the process. These haploid cells belong to one of two mating classes (sexes), called "a" and "α". Haploids can either propagate via normal mitosis or mate with other haploids of the different mating class, resulting again in diploid cells. This life cycle makes *S. cerevisiae* interesting for genetic studies; it has also been extensively used by experimental and modeling studies of the cell cycle, glycolysis, osmotic shock, and mating process [21–28]. Cell division occurs in *S. cerevisiae* in an asymmetric fashion called budding and single-cell studies have shown that yeast cells exhibit replicative senescence with a maximum of 30–40 divisions [29]. Since this process is very reminiscent of the replicative senescence known from human fibroblasts [30], *S. cerevisiae* is also employed as a model system for investigations of the aging process. Furthermore, *S. cerevisiae* was also the first eukaryotic organism to be sequenced and its genome consists of about 12 Mbp containing roughly 6000 genes distributed over 16 chromosomes [31]. Homologous recombination (the exchange of sequences between similar strands of DNA) is very efficient in *S. cerevisiae*, which makes the organism also a convenient model for studies of synthetic biology. Using this mechanism, it was possible to replace the complete chromosome 16 with a new, synthetic one through 11 successive rounds of transformation (see Chapter 14) [32]. The synthetic chromosome was streamlined by removing all introns and superfluous tRNA genes and using only two of the three possible stop codons. This opens the possibility to extend the genetic code by a further amino acid once all chromosomes are modified in this way. A good online resource for further information about this model organism is the *Saccharomyces* Genome Database (www.yeastgenome.org).

1.7.3 *Caenorhabditis elegans*

Of course, model systems for multicellular organisms are also needed and the nematode *C. elegans* (Figure 1.4c) has become such a model since Sidney Brenner introduced it to the research community [33]. Like the other model organisms, it is easy to cultivate (feeding on bacteria or synthetic medium) and thousands of the about 1 mm long animals can live on a large Petri dish. Wild populations of *C. elegans* consist mainly of hermaphrodites together with a few males. Hermaphrodites not only are capable of self-fertilization (leading to natural inbred lines), but can also mate with males. The hermaphrodite then lays eggs that develop into larvae after hatching and after a total of four larval stages (L1–L4) the adult animal emerges. The complete life cycle from egg to egg takes between 2.5 and 5.5 days, depending on the temperature. The total lifespan of *C. elegans* is rather short with 2–3 weeks. This made *C. elegans* another popular model system for the investigation of the aging process [34]. However, the nematode is also an important model for other fields of research such as molecular biology or neurology. RNA interference (RNAi), for instance, is an important experimental technique (Chapter 14) that was developed based on experiments in *C. elegans* [35]. Furthermore, adult nematodes have a fixed number of somatic cells that is identical for all individuals (1031 in the male and 959 in the hermaphrodite), which makes it possible to generate very detailed anatomical models of the worm. The "slidable worm" (www.wormatlas.org/slidableworm.htm), which is a resource available on the webpage of the WormAtlas database, presents the results of such anatomical studies using an easy-to-use interface. *C. elegans* is also the only animal for which the complete wiring diagram (connectome) of the nervous system has been determined (using electron microscopy serial sections) [36,37]. Finally, *C. elegans* has also been the first multicellular organism for which the complete genome sequence has been determined [38,39]. The 97 Mbp contain approximately 19 000 genes dispersed over six chromosomes. Good online starting points for more information are WormBase (www.wormbase.org), WormBook (www.wormbook.org/), or WormAtlas (www.wormatlas.org/).

1.7.4 *Drosophila melanogaster*

The fruit fly *D. melanogaster* (Figure 1.4d) is another, immensely popular, model organism that shares many of the properties of *C. elegans*. The animals are easy to breed in captivity and because of their small size (around 1 mm) it is possible to perform studies involving thousands of

individuals (e.g., for selection or population studies). The generation time (about 7 days at 29 °C) and lifespan (about 30 days at 29 °C) are very short and depend strongly on the ambient temperature. This facilitates, for example, artificial selection studies, which take several generations [40]. *D. melanogaster* has four chromosomes ($2n = 8$), which can even be studied under the light microscope because of a phenomenon called polyteny. As in many insect larvae, the cells of the salivary glands of *D. melanogaster* undergo multiple rounds of replication without cell division, leading to hundreds of sister chromatids aligned to each other. Polytene chromosomes are found in cells that need to express a large amount of a specific gene product and transcriptionally active areas appear under the microscope as swollen regions, so-called puffs. Although this technique is now outdated regarding the analysis of transcriptional activity, polytene chromosomes are still valuable for taxonomic problems. After staining, the puffs form a specific banding pattern that can be used to identify chromosomal deletions and duplications. This can be used in taxonomy to differentiate and classify closely related subspecies. *D. melanogaster* was arguably the most important model species for investigating developmental processes in multicellular organisms [41], which has led to the discovery of Hox genes [42]. These genes code for a set of transcription factors that contain a common 180 bp motif (the homeodomain) and control the development of the anterior–posterior axis of the animal. A unique feature of these genes is that they are arranged on the chromosomes in the same linear order as the body region that they affect (called collinearity). Thus, Hox genes at one end of the cluster control the development of the anterior region (head), while the genes at the other end of the cluster influence the development of the posterior region (tail). Although originally found in *Drosophila*, Hox genes have been found in many metazoans, including vertebrates [43]. The complete genome was sequenced in 2000 [44] and somewhat surprisingly the number of genes is with approximately 14 000 clearly smaller than the number of genes in *C. elegans*. Further information, tools, and resources are available at FlyBase (flybase.org) and Ensembl Genome Browser (www.ensembl.org/Drosophila_melanogaster).

1.7.5 *Mus musculus*

The last model system that we want to introduce here is the house mouse *M. musculus domesticus* (Figure 1.4e). It is clearly the model organism with the largest similarity to humans and is therefore also of great relevance for human research. Humans and mice are both mammals and thus share a common ancestor roughly 80 million years ago, a rather short time span compared with the other model organisms. Consequently, the genome structure and organization is also very similar. The mouse genome, sequenced in 2002 [45], contains 2.5 Gbp and is thus somewhat smaller than the human genome with 2.9 Gbp [2,3], although both genomes contain approximately 20 000–25 000 genes. The similarity at the gene level is quite amazing insofar that for more than 99% of mouse genes a homolog can also be found in the human genome [3], and vice versa. The mouse is also a popular model system because it is very amenable to genetic manipulations. The first mice were cloned in 1998 [46] and today it is common routine to create transgenic mice by introducing DNA constructs into fertilized egg cells and to study the function of existing genes by knocking them out or down (see Chapter 14). The Knockout Mouse Project (KOMP), for instance, aims at generating and providing mouse embryonic stem cells (and eventually whole mice) with single-gene knockout for every gene in the mouse genome (www.komp.org). Because mice have been used for such a long time as model species, many different inbred strains have been developed, which differ in various aspects of their phenotype (e.g., size, lifespan, and disease susceptibility). Of special interest are the various strains of nude mice that have a deletion of the FOXN1 gene, which prevents the formation of a functioning thymus. Without a thymus, these mice cannot produce mature T lymphocytes and therefore lack most forms of immune response (the lack of fur is a side effect of this mutation). As a consequence, they are valuable tools to study tumor development and are also used for transplantation studies, since they do not reject allo- or xenografts. Useful starting points for further information are, for instance, the Mouse Genome Informatics (www.informatics.jax.org/), the Mouse Atlas Project (www.emouseatlas.org), or the Ensembl Genome Browser (www.ensembl.org/Mus_musculus).

References

1 Goh, K.I., Cusick, M.E., Valle, D., Childs, B., Vidal, M., and Barabási, A.L. (2007) The human disease network. *Proc. Natl. Acad. Sci. USA*, 104 (21), 8685–8690.
2 Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, 409, 860–921.
3 Venter, J.C. *et al.* (2001) The sequence of the human genome. *Science*, 291, 1304–1351.
4 Mering, C. *et al.* (2002) Comparative assessment of large-scale data sets of protein–protein interactions. *Nature*, 417, 399–403.
5 Brazma, A. *et al.* (2001) Minimum Information About a Microarray Experiment (MIAME): toward standards for microarray data. *Nat. Genet.*, 29, 365–371.

6 Mack, S.J., Milius, R.P., Gifford, B.D., Sauter, J., Hofmann, J., Osoegawa, K., Robinson, J., Groeneweg, M., Turenchalk, G.S., Adai, A., Holcomb, C., Rozemuller, E.H., Penning, M.T., Heuer, M.L., Wang, C., Salit, M.L., Schmidt, A.H., Parham, P.R., Müller, C., Hague, T., Fischer, G., Fernandez-Viña, M., Hollenbach, J.A., Norman, P.J., and Maiers, M. (2015) Minimum Information for Reporting Next Generation Sequence Genotyping (MIRING): guidelines for reporting HLA and KIR genotyping via next generation sequencing. *Hum. Immunol.* doi: 10.1016/j.humimm.2015.09.011.

7 Taylor, C.F. *et al.* (2003) A systematic approach to modeling, capturing, and disseminating proteomics experimental data. *Nat. Biotechnol.*, 21, 247–254.

8 Hermjakob, H. *et al.* (2004) The HUPO PSI's molecular interaction format: a community standard for the representation of protein interaction data. *Nat. Biotechnol.*, 22, 177–183.

9 Hucka, M. *et al.* (2003) The Systems Biology Markup Language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19, 524–531.

10 Lloyd, C.M. *et al.* (2004) CellML: its future, present and past. *Prog. Biophys. Mol. Biol.*, 85, 433–450.

11 Le Novere, N., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A. *et al.* (2009) The Systems Biology Graphical Notation. *Nat. Biotechnol.*, 27 (8), 735–741.

12 Gitton, Y. *et al.* (2002) A gene expression map of human chromosome 21 orthologues in the mouse. *Nature*, 420, 586–590.

13 Ideker, T. *et al.* (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, 292, 929–934.

14 Kanehisa, M. and Bork, P. (2003) Bioinformatics in the post-sequence era. *Nat. Genet.*, 33, 305–310.

15 Tavazoie, S. *et al.* (1999) Systematic determination of genetic network architecture. *Nat. Genet.*, 22, 281–285.

16 Urbanczyk-Wochniak, E. *et al.* (2003) Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. *EMBO Rep.*, 4, 989–993.

17 Le Novere, N. *et al.* (2005) Minimum Information Requested in the Annotation of Biochemical Models (MIRIAM). *Nat. Biotechnol.*, 23, 1509–1515.

18 Waltemath, D., Adams, R., Beard, D.A., Bergmann, F.T., Bhalla, U.S., Britten, R. *et al.* (2011) Minimum Information About a Simulation Experiment (MIASE). *PLoS Comput. Biol.*, 7 (4), e1001122.

19 Blattner, F.R., Plunkett, G., 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M. *et al.* (1997) The complete genome sequence of *Escherichia coli* K-12. *Science*, 277 (5331), 1453–1462.

20 Lukjancenko, O., Wassenaar, T.M., and Ussery, D.W. (2010) Comparison of 61 sequenced *Escherichia coli* genomes. *Microb. Ecol.*, 60 (4), 708–720.

21 Hynne, F., Dano, S., and Sorensen, P.G. (2001) Full-scale model of glycolysis in *Saccharomyces cerevisiae*. *Biophys. Chem.*, 94 (1–2), 121–163.

22 Klipp, E., Nordlander, B., Kruger, R., Gennemark, P., and Hohmann, S. (2005) Integrative model of the response of yeast to osmotic shock. *Nat. Biotechnol.*, 23 (8), 975–982.

23 Diener, C., Schreiber, G., Giese, W., del Rio, G., Schroder, A., and Klipp, E. (2014) Yeast mating and image-based quantification of spatial pattern formation. *PLoS Comput. Biol.*, 10 (6), e1003690.

24 Kofahl, B. and Klipp, E. (2004) Modelling the dynamics of the yeast pheromone pathway. *Yeast*, 21 (10), 831–850.

25 Adrover, M.À., Zi, Z., Duch, A., Schaber, J., González-Novo, A., Jimenez, J., Nadal-Ribelles, M., Clotet, J., Klipp, E., and Posas, F. (2011) Time-dependent quantitative multicomponent control of the G_1-S network by the stress-activated protein kinase Hog1 upon osmostress. *Sci. Signal.*, 4 (192), ra63. Erratum: *Sci. Signal.*, 4 (197), er5 (2011).

26 Chen, K.C., Calzone, L., Csikasz-Nagy, A., Cross, F.R., Novak, B., and Tyson, J.J. (2004) Integrative analysis of cell cycle control in budding yeast. *Mol. Biol. Cell*, 15 (8), 3841–3862.

27 Rizzi, M., Baltes, M., Theobald, U., and Reuss, M. (1997) *In vivo* analysis of metabolic dynamics in *Saccharomyces cerevisiae*. II. Mathematical model. *Biotechnol. Bioeng.*, 55 (4), 592–608.

28 Kaizu, K., Ghosh, S., Matsuoka, Y., Moriya, H., Shimizu-Yoshida, Y., and Kitano, H. (2010) A comprehensive molecular interaction map of the budding yeast cell cycle. *Mol. Syst. Biol.*, 6, 415.

29 Jazwinski, S.M. (1990) Aging and senescence of the budding yeast *Saccharomyces cerevisiae*. *Mol. Microbiol.*, 4 (3), 337–343.

30 Hayflick, L. (1965) The limited *in vitro* lifetime of human diploid cell strains. *Exp. Cell Res.*, 37, 614–636.

31 Mewes, H.W., Albermann, K., Bahr, M., Frishman, D., Gleissner, A., Hani, J. *et al.* (1997) Overview of the yeast genome. *Nature*, 387 (6632 Suppl.), 7–65.

32 Annaluru, N., Muller, H., Mitchell, L.A., Ramalingam, S., Stracquadanio, G., Richardson, S.M. *et al.* (2014) Total synthesis of a functional designer eukaryotic chromosome. *Science*, 344 (6179), 55–58.

33 Brenner, S. (1973) The genetics of *Caenorhabditis elegans*. *Genetics*, 77, 71–94.

34 Johnson, T.E. (2013) 25 years after age-1: genes, interventions and the revolution in aging research. *Exp. Gerontol.*, 48 (7), 640–643.

35 Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391 (6669), 806–811.

36 White, J.G., Southgate, E., Thomson, J.N., and Brenner, S. (1986) The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. B*, 314 (1165), 1–340.

37 Jarrell, T.A., Wang, Y., Bloniarz, A.E., Brittin, C.A., Xu, M., Thomson, J.N. *et al.* (2012) The connectome of a decision-making neural network. *Science*, 337 (6093), 437–444.

38 C. elegans Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*, 282, 2012–2018.

39 Hillier, L.W., Coulson, A., Murray, J.I., Bao, Z., Sulston, J.E., and Waterston, R.H. (2005) Genomics in *C. elegans*: so many genes, such a little worm. *Genome Res.*, 15 (12), 1651–1660.

40 Rose, M. and Charlesworth, B. (1980) A test of evolutionary theories of senescence. *Nature*, 287 (5778), 141–142.

41 Nusslein-Volhard, C. and Wieschaus, E. (1980) Mutations affecting segment number and polarity in *Drosophila*. *Nature*, 287 (5785), 795–801.

42 Scott, M.P. and Weiner, A.J. (1984) Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of *Drosophila*. *Proc. Natl. Acad. Sci. USA*, 81 (13), 4115–4119.

43 Gehring, W.J. (1992) The homeobox in perspective. *Trends Biochem. Sci.*, 17 (8), 277–280.

44 Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G. *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science*, 287 (5461), 2185–2195.

45 Mouse Genome Sequencing Consortium, Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420 (6915), 520–562.

46 Wakayama, T., Perry, A.C., Zuccotti, M., Johnson, K.R., and Yanagimachi, R. (1998) Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature*, 394 (6691), 369–374.

Further Reading

The early days of systems biology: Kitano, H. (2001) *Foundations of Systems Biology*, MIT Press, Cambridge, MA.

The early days of systems biology: Kitano, H. (2002) Systems biology: a brief overview. *Science*, 295 (5560), 1662–1664.

Numbers in cell biology: Flamholz, A., Philips, R., and Milo, R. (2014) The quantified cell. *Mol. Biol. Cell*, 25 (22), 3497–3500.

Numbers in cell biology: Milo, R. and Phillips, R. (2014) *Cell Biology by the Numbers*, Garland Science.

Systemic thinking in cell biology: Lazebnik, Y. (2002) Can a biologist fix a radio? Or, what I learned while studying apoptosis. *Cancer Cell*, 2, 179–182.

Physical constraints on cell function: Dill, K.A., Ghosh, K., and Schmit, J.D. (2011) Physical limits of cells and proteomes. *Proc. Natl. Acad. Sci. USA*, 108 (44), 17876–17882.

Modeling of Biochemical Systems

2

Over the last two decades, formulation of formal (often mathematical) concepts and computational modeling have become more and more familiar in biology, although they have been applied much earlier. Leonardo of Pisa used in his book from 1202 the Fibonacci numbers to describe the growth of a rabbit population. Lotka and Volterra characterized population dynamics of lynx and hares (predators and preys) with their famous equation in the 1920s. Michaels and Menten developed a model to determine the rate for kinetic reactions in 1913. In the mid of the last century, Ludwig von Bertalanffy became famous for his work on biophysics of open systems, thermodynamics of living systems, and the introduction of the notion of "Fließgleichgewicht," roughly translated with dynamic equilibrium.

Since that time, many different approaches have been introduced and applied for in-depth understanding of biology, to relate independent biological observations to each other (what has the protein content measured in one experiment to do with the mRNA amounts measured independent of another technique?). Concepts from physics and engineering have fruitfully invaded biological research, one example being concepts of control systems for gene regulation. Biology is on its way to a quantitative science like physics, chemistry, or geography. It requires numbers and mathematical predictions to complement biological measurements and to derive useful predictions from them. We can compare it, in some sense, with the weather forecast: Data are combined with models and evaluated by computers to predict tomorrow's weather. The forecast is often not precise, but getting better. However, biology is much more complicated with different species, different cell types making a higher organism, and processes of development and evolution. Nevertheless, many promising results indicate that it is worth trying to provide a mathematical description of biological systems.

Here, we will give a short and cursory introduction to model concepts and how to formulate a first model for the process of interest. More detailed explanations of construction and analysis procedures for models follow in the later chapters.

2.1 Overview of Common Modeling Approaches for Biochemical Systems

Summary

We give an overview of frequently used modeling approaches in systems biology such as network-based models, rule-based models, or statistical models and discuss their fields of application, strengths, and basic underlying principles. Links to the individual chapters with in-depth explanation, examples, and questions are provided.

Understanding a biological system is not a straight, unidirectional process. It requires understanding of its topology or the structure of the system. It involves analyses of its dynamical behavior and the control mechanisms at play. Also, it requires interpreting how its design and function relate one another in the overall context. The

Systems Biology: A Textbook, Second Edition. Edda Klipp, Wolfram Liebermeister, Christoph Wierling, and Axel Kowald.

2.2.1 Basic Components of ODE Models

To formulate an ODE model for a dynamic biochemical reaction network, we need the following information:

1) The basic building blocks are all compounds and all reactions converting these compounds into each other. A list of the reactions and their substrates and products gives us the stoichiometry of the network. This approach is further detailed in Chapter 3.
2) The modeler must set the boundary of the system. Which components should the model follow? They become internal components. Which components are not considered relevant for the model? They are completely left out. Which components determine the model behavior, but are not changed by its dynamics? They are called external components. For a metabolic pathway, for example, all metabolites within the pathway may be internal metabolites, while the concentration of a nutrient provided in the medium at fixed experimental conditions may be an external metabolite. All other cellular components may be ignored for that specific study.
3) For all reactions that are part of the model, assign kinetic laws (see Chapter 4).
4) Determine the values of the kinetic parameters used in the kinetic laws. They can be taken either from databases or literature or they can be fitted to experimental data (Chapter 6).

On the basis of the well-formulated model, we can perform different analysis steps:

5) Find out whether the system has a steady state or not. Is the steady state uniquely defined or do we obtain several steady states depending on the parameter values? Calculate steady state concentrations and reaction rates.
6) Simulate the time course for a given set of parameter values and initial conditions (tools and techniques are introduced in Chapter 5).
7) Analyze the effect of perturbations. The impact of small changes of parameter values is studied by sensitivity analysis and – for biochemical networks – by metabolic control analysis (see Section 4.2). One may also test the effect of complete knockouts of enzymes catalyzing the individual reactions or of knockdowns or overexpression.

2.2.2 Illustrative Examples of ODE Models

To illustrate the steps introduced above, we consider two little models: one for a metabolic network and the other for a small regulatory network.

2.2.2.1 Metabolic Example

Metabolism serves the uptake of nutrients to convert them into energy, mostly in the form of ATP, and into building blocks such as amino acids and lipids. All metabolic reactions are catalyzed by enzymes. The metabolic network in Figure 2.1 resembles the first steps in glycolysis, a major catabolic pathway for the uptake and initial phosphorylation of glucose, which is afterward distributed to other catabolic and anabolic pathways to provide building blocks and energy. Let us call extracellular glucose S_0 and its concentration S_0. Intracellular glucose is S_1, singly phosphorylated glucose (glucose-6-phosphate, G6P) is S_2, and doubly phosphorylated fructose (fructose-1,6-bisphosphate, $F16P_2$) is S_4. G6P is provided for other synthetic pathways producing S_3. Phosphorylation is carried out by consuming ATP (A_3) and converting it into ADP (A_2).

We can describe the dynamics of the network in Figure 2.1a by a set of ordinary differential equations as follows:

$$\begin{aligned} \frac{dS_1}{dt} &= v_1 - v_2, \\ \frac{dS_2}{dt} &= v_2 - v_3 - v_4, \\ \frac{dA_2}{dt} &= -\frac{dA_3}{dt} = v_2 + v_4. \end{aligned} \tag{2.1}$$

When we assign kinetics to the reactions, we can simulate the system. A simple choice of kinetics is mass action kinetics, where the reaction rate is proportional to the concentration of its substrates:

$$v_1 = k_1, \quad v_2 = k_2 \cdot S_1 \cdot A_3, \quad v_3 = k_3 \cdot S_3, \quad v_4 = k_4 \cdot S_2 \cdot A_3. \tag{2.2}$$

You can find more information on kinetic laws in Chapter 4. We will now use this set of equations to simulate the dynamic behavior of the network. Starting with an ATP concentration of 1, an ADP concentration of 0, and zero concentrations of the internal sugars (S_1, S_2), we find that ATP is consumed and ADP is produced. S_1 is produced in an unlimited fashion through the uptake reaction v_1, but S_2 is produced only as long as ATP is available, afterward it declines. Since there is unlimited supply of S_0, the system has no steady state, that is, no state with $dS_i/dt = 0\,(i = 1, \ldots, n)$.

A steady state with $dS_1/dt = dS_2/dt = 0$ can be obtained if we consider that ATP and ADP are kept constant by other cellular processes, that is, $dA_3/dt = dA_2/dt = 0$. Then we can consider them as external variables, as shown in Figure 2.1c with the dynamics represented in Figure 2.1d.

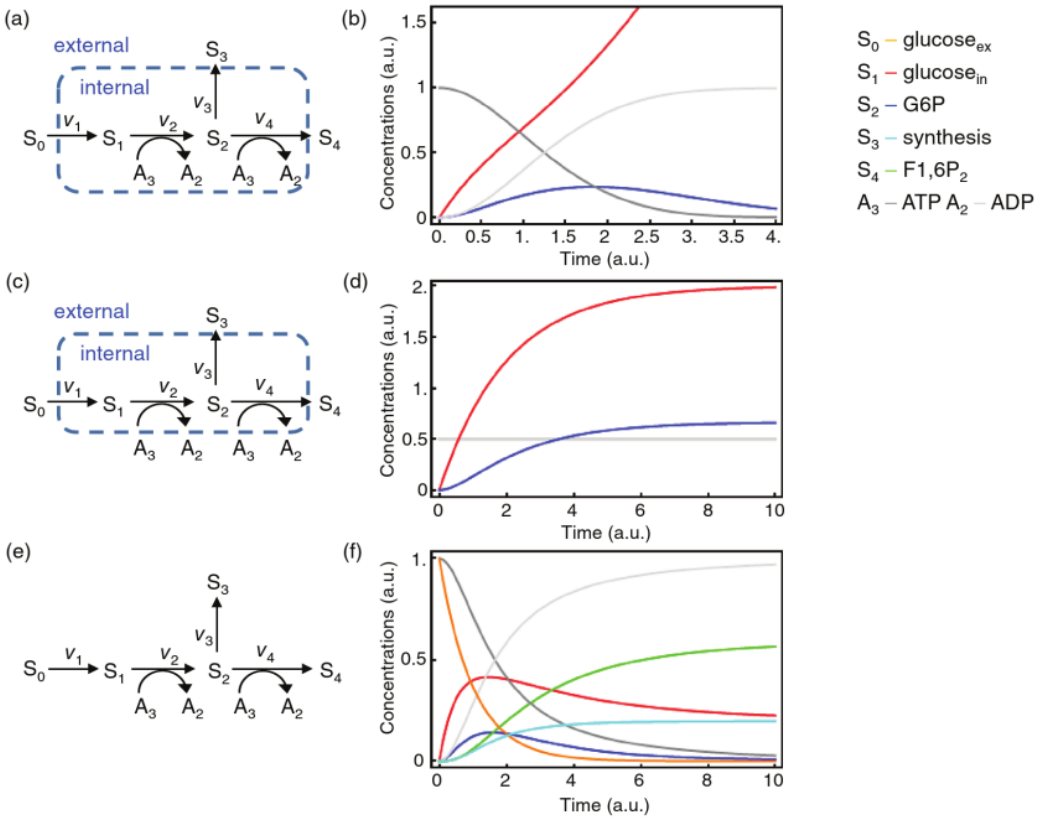


Figure 2.1 Example of a metabolic model. (a) Network representation with S_0 and S_3 considered external and S_1, S_2 as well as A_2 and A_3 treated as internal variables. (b) Time course resulting from dynamic simulation of the network shown in part (a). (c) The same network as in part (a), but here A_2 and A_3 are also treated as external. (d) Time course resulting from dynamic simulation of the network shown in part (c). (e) Network representation with all components considered as internal and, therefore, dynamic. (f) Dynamics of the network shown in part (e). The dynamics and kinetics are listed in Eqs. (2.1) and (2.2). Parameter values: $k_i = 1$ $(i = 1, \ldots, 4)$.

If S_0, S_3, and S_4 were internal variables that can change ($dS_0/dt = -v_1$, $dS_3/dt = v_3$, $dS_4/dt = v_4$, respectively), then the mass provided by S_0 remains within the system and it approaches a state where S_0, A_3, and S_2 decline to 0, while A_2 reaches 1, S_1 and S_4 approach about 0.2, and S_3 about 0.6. However, the model will never reach a true steady state (Figure 2.1e and f).

2.2.2.2 Regulatory Network Example

Stem cell research is of increasing importance in biological research and of great interest in health care. Besides many other promises, it provides the hope that in the future many diseases can be cured by administration of healthy cells of a specific tissue to a diseased person that have been grown out of reprogrammed induced pluripotent stem cells (iPS cells) originating from the same person. The three genes (and gene products) considered as the main regulators of stemness of cells are Oct4, Sox2, and Nanog. They activate each other, but they are controlled by epigenetic marking and by growth factors. Cellular differentiation is accompanied by hypermethylation of their promoters and by downregulation of their gene expression. In order to create iPS cells, many experimental procedures have been tested. The addition of viral plasmid containing four factors – Sox2, Oct4, c-Myc, and the microRNA Klf4 – that was introduced by Takahashi and Yamanaka in 2006 [2] was most successful.

Here, we will use a simple model to study some basic properties of that system. Let us first assume that Oct4, Sox2, and Nanog stabilize each other. Their expression is suppressed by the epigenetic marking (the DNA methylation), but the proteins also prevent DNA methylation (see Figure 2.2). This mutual inhibition can be described on different levels of detail (e.g., including the joint stabilization of the proteins or not). We will use the following differential equation system that focuses on the collective effect of stemness markers on epigenetic marking and vice versa, that is, the double negative feedback (resulting in a positive feedback) that

each component has on itself:

$$\frac{\mathrm{d}OSN}{\mathrm{d}t} = v_1 - v_2 - v_3 = k_1 - k_2 \cdot \frac{EM^{n_1}}{\left(K_1^{n_1} + EM^{n_1}\right)} - k_3 \cdot OSN,$$

$$\frac{\mathrm{d}EM}{\mathrm{d}t} = v_4 - v_5 - v_6 = k_4 - k_5 \cdot \frac{OSN^{n_2}}{\left(K_2^{n_2} + OSN^{n_2}\right)} - k_6 \cdot EM. \tag{2.3}$$

OSN denotes the common activity of the stemness markers Oct4, Sox2, and Nanog. *EM* represents the level of epigenetic marking. The activity of OSN increases linearly, but EM inhibits it in a fashion described with a Hill function (see Chapter 4). Its basal degradation is proportional to its current concentration. The respective rules hold for EM with OSN as inhibitor. The behavior is illustrated in Figure 2.2. It shows that the system has three steady states. One steady state is unstable, the other two steady states feature either low levels of epigenetic marking and high expression of the stemness factors, indicating that the cell is a stem cell, or high levels of epigenetic marking and low levels of Oct4, Sox2, and Nanog, indicating differentiation. In isolation, the system will always reach one of these states, depending on the initial conditions, and then remain there. It can only be

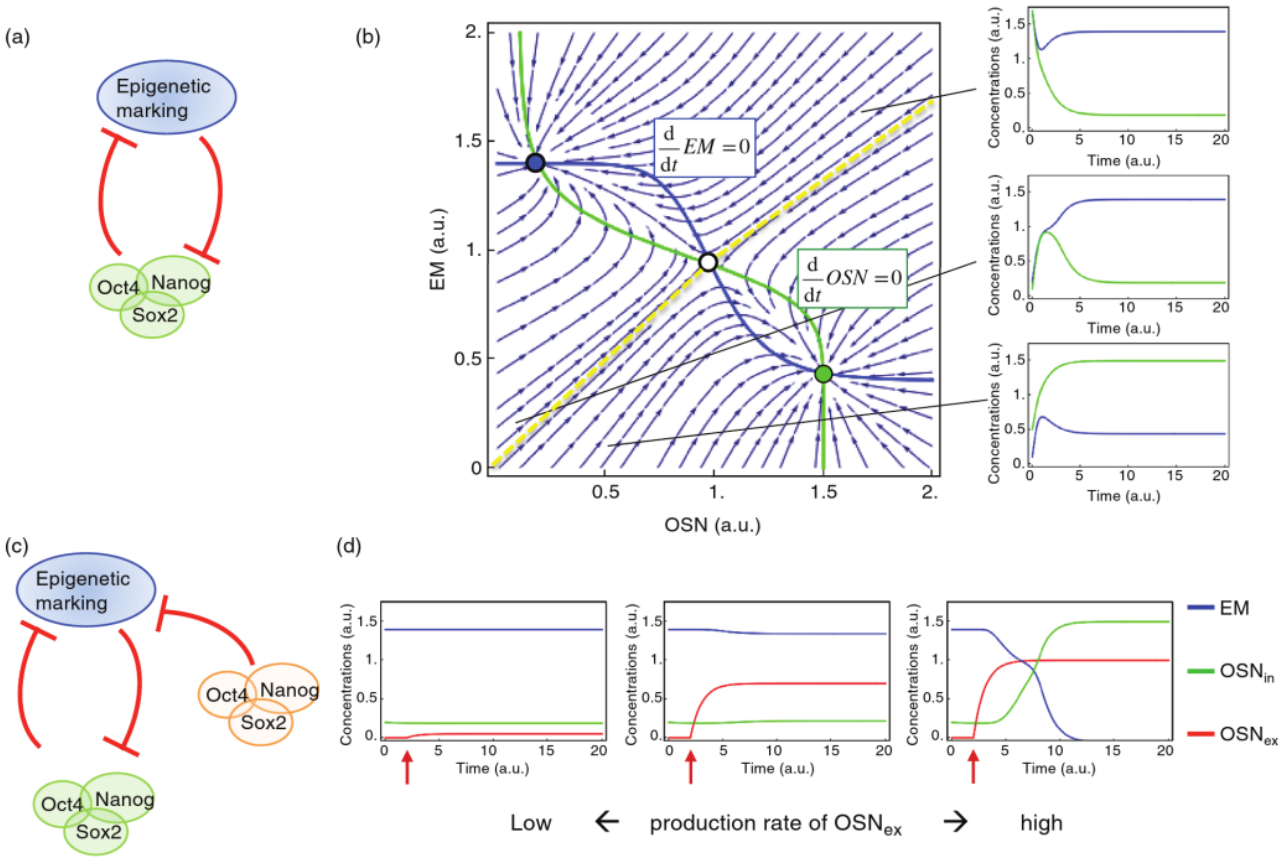


Figure 2.2 Network and dynamics of a model for epigenetic regulation of stemcellness. (a) Epigenetic regulation (EM) and the three factors Oct4, Sox2, and Nanog (OSN) responsible for pluripotency. The dynamics is described with the ordinary differential equations system (2.3). (b) The phase plan showing a plot of *EM* against *OSN* represents the three steady states of this system, one stable state for high values of *EM* (indicated in blue), one stable state for high values of *OSN* (indicated in green), and one in-between. The state in-between is unstable. The small light blue arrows indicate the actual flow of the system at each point in the phase plan. The green line is the line with no change (nullcline) of *OSN*; hence, the flow arrows cross it always vertically. The blue line for steady values of EM is always crossed horizontally. The dashed yellow line is called separatrix since it separates the basin of attraction for the steady state featuring high *EM* and low *OSN* (blue dot) from the basin of attraction for the steady state with high *OSN* and low *EM* (green dot). The small time plots to the right exemplify that any starting condition within these basins of attraction leads to the respective stable state. (c) The system can be pushed out of its current stable state by supply of another component, here by OSN transcribed from an exogenous vector, which inhibits EM but is not regulated by it. (d) The effect depends on the expression strength of external OSN; only if it is expressed strongly enough and for sufficient time, it can reverse the cellular decision from high *EM* to high *OSN* (and low *EM*).

moved out of this state by external cues. Under natural conditions, stem cells are forced into differentiation by external signaling compounds such as Wnt or growth factors. When trying to reprogram cells away from the differentiated state toward induced pluripotency, the strategy introduced by Takahashi and Yamanaka favors OSN through the expression from a viral vector. This has the effect that these four compounds are expressed and active, for example, in regulating epigenetic marking, but they are not under epigenetic regulation themselves. If their expression is strong and long enough, they push the cells back into conditions featuring pluripotency with high expression also of endogenous stem cell factors.

Cellular reprogramming with viral vectors has provided many opportunities to study the reprogramming process in detail and determine the contribution of individual regulatory mechanisms, such as cell cycle progression. It is less suited for long-term application in human patients, which is an interesting medical aim, since it implies using viral material and since the uncontrolled expression of pluripotency factors can also lead to unintended side effects such as cancerogenesis. Hence, the search for alternative ways for reprogramming is ongoing, for example, by using small molecules.

References

1 Klipp, E., Liebermeister, W., Helbig, A., Kowald, A., and Schaber, J. (2007) Systems biology standards: the community speaks. *Nat. Biotechnol.*, 25, 390–391.

2 Takahashi, K. and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126, 663–676.

Further Reading

Agent-Based Modeling: An, G., Mi, Q., Dutta-Moscato, J., and Vodovotz, Y. (2009) Agent-based models in translational systems biology. *Wiley Interdiscip. Rev. Syst. Biol. Med.*, 1 (2), 159–171.

Introduction to Mathematical Concepts in Biology: Edelstein-Keshet, L. (1988) *Mathematical Models in Biology*, SIAM.

Introduction to Mathematical Concepts in Biology: Herbert, S. (2014) *Systems Biology: Introduction to Pathway Modeling*, Ambrosius Publishing.

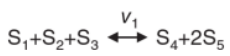
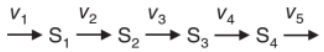
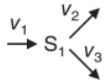
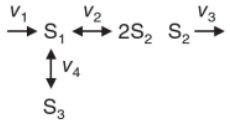
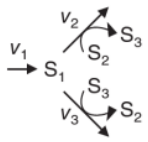
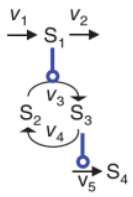
Boolean Modeling: Kauffman, S.A. (1987) Developmental logic and its evolution. *Bioessays*, 6 (2), 82–87.

Boolean Modeling: Kauffman, S.A., Peterson, C., Samuelsson, B., and Troein, C. (2003) Random Boolean network models and the yeast transcriptional network. *Proc. Natl. Acad. Sci. USA*, 100, 14796–14799.

Boolean Modeling: Keener, J. and Sneyd, J. (1998) *Mathematical Physiology*, Springer, New York.

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

Table 3.1 Different reaction networks, their stoichiometric matrices, and the respective system of ODEs.

	Network	*Stoichiometric matrix*	*ODE system*
N1		$\mathbf{N} = \begin{pmatrix} -1 \\ -1 \\ -1 \\ 1 \\ 2 \end{pmatrix}$	$\frac{dS_1}{dt} = \frac{dS_2}{dt} = \frac{dS_3}{dt} = -v_1$ $\frac{dS_4}{dt} = v_1$ $\frac{dS_5}{dt} = 2v_1$
N2		$\mathbf{N} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$	$\frac{dS_1}{dt} = v_1 - v_2$ $\frac{dS_2}{dt} = v_2 - v_3$ $\frac{dS_3}{dt} = v_3 - v_4$ $\frac{dS_4}{dt} = v_4 - v_5$
N3		$\mathbf{N} = \begin{pmatrix} 1 & -1 & -1 \end{pmatrix}$	$\frac{dS_1}{dt} = v_1 - v_2 - v_3$
N4		$\mathbf{N} = \begin{pmatrix} 1 & -1 & 0 & -1 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$	$\frac{dS_1}{dt} = v_1 - v_2 - v_4$ $\frac{dS_2}{dt} = 2v_2 - v_3$ $\frac{dS_3}{dt} = v_4$
N5		$\mathbf{N} = \begin{pmatrix} 1 & -1 & -1 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix}$	$\frac{dS_1}{dt} = v_1 - v_2 - v_4$ $\frac{dS_2}{dt} = -v_2 + v_3$ $\frac{dS_3}{dt} = v_2 - v_3$
N6		$\mathbf{N} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\frac{dS_1}{dt} = v_1 - v_2$ $\frac{dS_2}{dt} = v_4 - v_3$ $\frac{dS_3}{dt} = -v_4 + v_3$ $\frac{dS_4}{dt} = v_5$

Note that external metabolites are neither drawn in the network nor included in the stoichiometric matrix. Thin arrows denote reactions and bold arrows denote activation.

state, we can also consider the vector $\mathbf{J} = (J_1, J_2, \ldots, J_r)^{\mathrm{T}}$ containing the steady-state fluxes. With these notions, the balance equation reads

$$\frac{d\mathbf{S}}{dt} = \mathbf{N}\mathbf{v}, \tag{3.5}$$

a compact form that is suited for various types of analyses.

3.1.2 Information Encoded in the Stoichiometric Matrix N

The stoichiometric matrix contains important information about the structure of the metabolic network. Using the stoichiometric matrix, we may calculate which combinations of individual fluxes are possible in steady state (i.e., calculate the admissible steady-state flux space).

We may easily find out dead ends and unbranched reaction pathways. In addition, we may discover the conservation relations for the included reactants.

In steady state, it holds that

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}t} = \mathbf{N}\mathbf{v} = \mathbf{0}. \tag{3.6}$$

Note that $\mathbf{0}$ is a vector with length n, that is, $\mathbf{0} = (0, 0, \ldots, 0)^{\mathrm{T}}$. The right equality sign in Eq. (3.6) denotes a linear equation system for determination of the rates $\mathbf{v}$. From linear algebra, it is known that this equation has nontrivial solutions only for $\mathrm{Rank}(\mathbf{N}) < r$ (see Section 15.1 for an introduction to linear algebra). A kernel matrix $\mathbf{K}$ fulfilling

$$\mathbf{N}\mathbf{K} = \mathbf{0} \tag{3.7}$$

expresses the respective linear dependencies between the columns of the stoichiometric matrix [3]. $\mathbf{K}$ consists of $r - \mathrm{Rank}(\mathbf{N})$ basis vectors as columns and can be determined using the Gauss algorithm (see mathematical textbooks). The kernel is not uniquely defined. Multiplication of $\mathbf{K}$ with a regular matrix $\mathbf{Q}$ of appropriate size ($\mathbf{K}' = \mathbf{K} \cdot \mathbf{Q}$, equivalently to linear combination of the columns of $\mathbf{K}$) yields another valid kernel $\mathbf{K}'$ of $\mathbf{N}$.

Every possible set $\mathbf{J}$ of steady-state fluxes can be expressed as linear combination of the columns $\mathbf{k}_i$ of $\mathbf{K}$:

$$\mathbf{J} = \sum_{i=1}^{r-\mathrm{Rank}(\mathbf{N})} \alpha_i \cdot \mathbf{k}_i. \tag{3.8}$$

The coefficients must have units corresponding to the units of reaction rates (e.g., $\mathrm{mM\,s^{-1}}$).

If the entries in a certain row are zero in all basis vectors, we have found an equilibrium reaction. In any steady state, the net rate of this reaction must be zero. For the reaction system N4 in Table 3.1, it holds that $r = 4$ and $\mathrm{Rank}(\mathbf{N}) = 3$. Its kernel consists of only one column $\mathbf{K} = \begin{pmatrix} 1 & 1 & 1 & 0 \end{pmatrix}^{\mathrm{T}}$. Hence, $v_4 = \sum_{i=1}^{1} \alpha \cdot 0 = 0$. In any steady state, the rates of production and degradation of S_3 must be equal, thereby leading to zero net change.

If all basis vectors contain the same entries for a set of rows, this indicates an unbranched reaction path. In each steady state, the net rate of all respective reactions is equal.

Up to now, we have not been concerned about (ir) reversibility of reactions in the network. The irreversibility of a reaction does not affect the stoichiometric matrix. However, it has consequences for the choice of basis vectors $\mathbf{k}_i$ for the kernel $\mathbf{K}$. A set of basis vectors must be chosen to satisfy the signs of fluxes when calculated by Eq. (3.8).

Example 3.1

For the network N2 in Table 3.1, we have $r = 5$ reactions and $\mathrm{Rank}(\mathbf{N}) = 4$. The kernel matrix contains just $1 = 5 - 4$ basis vectors, which are multiples of $\mathbf{k}_1 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \end{pmatrix}^{\mathrm{T}}$. This means that in steady state the flux through all reactions must be equal.

Network N3 comprises $r = 3$ reactions and has $\mathrm{Rank}(\mathbf{N}) = 1$. Each representation of the kernel matrix contains $3 - 1 = 2$ basis vectors, for example,

$$\mathbf{K} = \begin{pmatrix} \mathbf{k}_1 & \mathbf{k}_2 \end{pmatrix} \quad \text{with} \quad \mathbf{k}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{k}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \tag{3.9}$$

and for the steady-state flux holds

$$\mathbf{J} = \alpha_1 \cdot \mathbf{k}_1 + \alpha_2 \cdot \mathbf{k}_2. \tag{3.10}$$

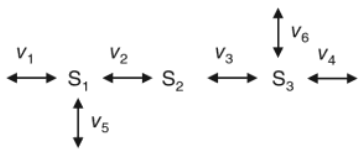
Network N6 can present a small signaling cascade. It has five reactions and $\mathrm{Rank}(\mathbf{N}) = 3$. Two basis vectors of the kernel are

$$\mathbf{k}_1 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \end{pmatrix}^{\mathrm{T}},$$
$$\mathbf{k}_2 = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 \end{pmatrix}^{\mathrm{T}}. \tag{3.11}$$

If we calculate the possible steady-state fluxes according to Eq. (3.10), we can easily see that in every steady state it holds that production and degradation of S_1 are balanced ($J_1 = J_2$) and that the fluxes through the cycle are equal ($J_3 = J_4$). In addition, J_5 must be equal to zero, otherwise S_4 would accumulate. One could prevent the last effect by also including the degradation of S_4 into the network.

Example 3.2

Consider the reaction scheme



(3.12)

The system comprises $r = 6$ reactions. The stoichiometric matrix reads

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

with $\mathrm{Rank}(\mathbf{N}) = 3$. Thus, the kernel matrix is spanned by three basis vectors, for example, $\mathbf{k}_1 = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & -1 \end{pmatrix}^{\mathrm{T}}$, $\mathbf{k}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}^{\mathrm{T}}$, and $\mathbf{k}_3 = \begin{pmatrix} -1 & -1 & -1 & -1 & 0 & 0 \end{pmatrix}^{\mathrm{T}}$. The entries for the second and third reactions are always equal; thus, in any steady state, the fluxes through reactions 2 and 3 must be equal.

3.1.3 The Flux Cone

The stoichiometric analysis of biochemical network analysis can be modified by considering only irreversible reactions (e.g., by splitting reversible reactions into two irreversible ones). Based on such a unidirectional representation, the basis vectors (Eq. (3.8)) form a convex cone in the flux space. This mapping relates stoichiometric analysis to the concepts of convex geometry as follows. The steady-state assumption requires that a flux vector is an element of the null space of the stoichiometric matrix **N** spanned by matrix **K**. A row of **K** can be interpreted as a hyperplane in flux space. The intersection of all these hyperplanes forms the null space. Provided that all reactions are unidirectional or irreversible, the intersection of the null space with the semipositive orthant of the flux space forms a polyhedral cone, the flux cone. The intersection procedure results in a set of rays or edges starting at 0, which fully describe the cone. The edges are represented by vectors and any admissible steady state of the system is a positive combination of these vectors. An illustration is presented in Figure 3.1.

3.1.4 Elementary Flux Modes and Extreme Pathways

A stringent definition of the term "pathway" in a metabolic network is not straightforward. A descriptive definition of a pathway is a set of reactions that are linked by common metabolites. Typical examples include glycolysis or different amino acid synthesis pathways. More detailed inspection of metabolic maps such as the *Boehringer chart* [4] shows that metabolism is highly interconnected and better addressed as a network. Pathways that are known for a long time from biochemical experience are already hard to recognize, and it is even harder to find out new pathways, for example, in metabolic maps that have been reconstructed from sequenced genomes of bacteria.

The problem of clearly identifying functional pathways has been elaborated in the concepts of *elementary flux modes* [3,5–10] and *extreme pathways* [11–14]. In both cases, the stoichiometry of a metabolic network is investigated to discover which direct routes are possible that lead from one external metabolite to another external metabolite. Both approaches use the steady-state assumption and take into account that some reactions are reversible, while others are irreversible. Despite these two constraints, we still obtain too many solutions, especially for larger networks. For elementary flux modes, this problem is solved by the requirement that they cannot be further decomposed, while extreme fluxes are bound to the generating vectors of the flux cone, as explained below.

We start with defining a *flux mode* **M**. It is the set of flux vectors that represent direct routes through the network between external metabolites. In mathematical terms, it is defined as the set

$$\mathbf{M} = \left\{ \mathbf{v} \in \mathbf{R}^r | \mathbf{v} = \lambda \mathbf{v}^*, \lambda > 0 \right\}, \tag{3.13}$$

where $\mathbf{v}^*$ is an r-dimensional vector (unequal to the null vector) fulfilling two conditions:

1) the steady-state condition $\mathbf{Nv} = \mathbf{0}$, that is, Eq. (3.6), and
2) sign restriction, that is, the flux directions in $\mathbf{v}^*$ fulfill the prescribed irreversibility relations. $\mathbf{v}^{\text{irr}}$ denotes the subvector of $\mathbf{v}^*$ that contains only nonnegative fluxes.

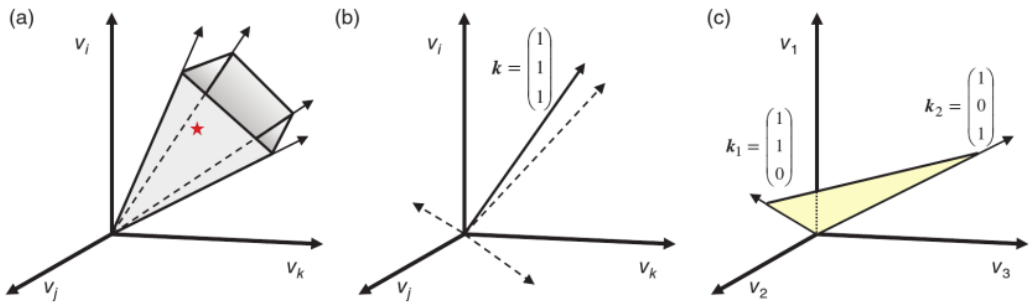


Figure 3.1 Flux cone: schematic representation of the subspace of feasible steady states within the space spanned by all positive-valued vectors for rates of irreversible reactions, v_i, $i = 1, \ldots, r$. Only three dimensions are shown. Feasible solutions are linear combinations of basis vectors of matrix **K** (see text). (a) Illustrative representation of the flux cone for a higher dimensional system (with $r - \text{Rank}(\mathbf{N}) = 4$). The basis vectors of **K** are rays starting at the origin. The line connecting the four rays indicates possible limits for real flux distributions set by constraints. The little asterisk indicates one special feasible solution for the fluxes. (b) The flux cone for an unbranched reaction chain of arbitrary length, such as the network N2 in Table 3.1, is just a ray since **K** is represented by a single basis vector containing only 1s. (c) The flux cone for network N3 in Table 3.1 is the plane spanned by the basis vectors $\mathbf{k}_1 = \begin{pmatrix} 1 & 1 & 0 \end{pmatrix}^T$ and $\mathbf{k}_2 = \begin{pmatrix} 1 & 0 & 1 \end{pmatrix}^T$.

A flux mode $\mathbf{M}$ comprising $\mathbf{v}$ is called reversible if the set $\mathbf{M}'$ comprising $-\mathbf{v}$ is also a flux mode.

A flux mode is an *elementary flux mode* if it uses a minimal set of reactions and cannot be further decomposed, that is, the vector $\mathbf{v}$ cannot be represented as nonnegative linear combination of two vectors that fulfill conditions (1) and (2) but contain more zero entries than $\mathbf{v}$. An elementary flux mode is a minimal set of enzymes that could operate at steady state, with all the irreversible reactions used in the appropriate direction. The number of elementary flux modes is at least as high as the number of basis vectors of the null space. The set of elementary flux modes is uniquely defined. Pfeiffer *et al.* [6] developed a software ("Metatool") to calculate the elementary flux modes for metabolic networks.

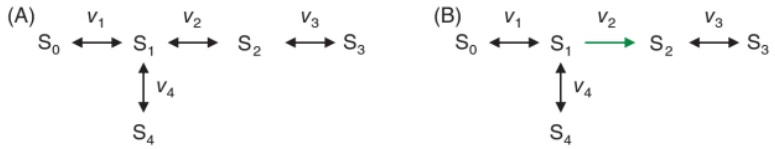
A flux mode is an *extreme pathway* if

1) all reactions are nonnegative, that is, $\mathbf{v} = \mathbf{v}^{\mathrm{irr}}$;
2) it belongs to the edges of the flux cone, which also means that it represents a basis vector of $\mathbf{K}$.

To achieve the first, reversible reactions are broken down into their forward and backward components and exchange fluxes have to be defined in the appropriate direction. This way, the set of extreme pathways is a

Example 3.3

The systems (A) and (B) differ by the fact that reaction 2 is either reversible or irreversible.



(3.14)

The elementary flux modes connect the external metabolites S_0 and S_3, S_0 and S_4, or S_3 and S_4. The stoichiometric matrix and the flux modes for case (A) and case (B) are

$$\mathbf{N} = \begin{pmatrix} 1 & -1 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{pmatrix}, \quad \mathbf{v}^{\mathrm{A}} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ -1 \end{pmatrix}, \quad \text{and}$$

$$\mathbf{v}^{\mathrm{B}} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ -1 \end{pmatrix}. \tag{3.15}$$

The possible routes are illustrated in Figure 3.2.

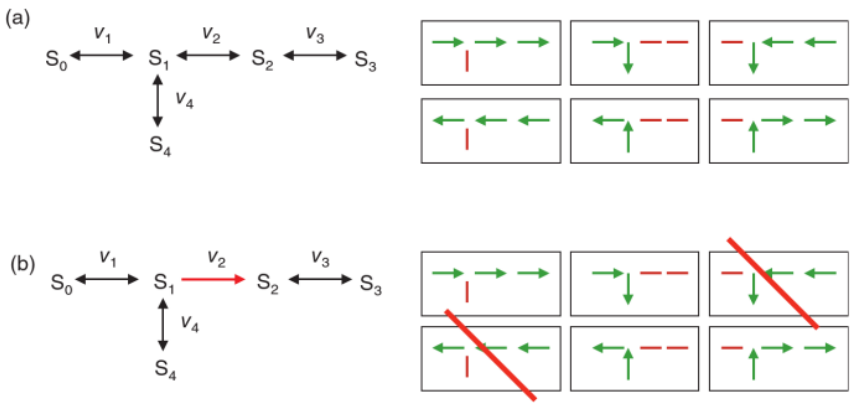


Figure 3.2 Schematic representation of elementary flux modes for the reaction network depicted in Eq. (3.14).

subset of the set of elementary flux modes and the extreme pathways are systemically independent.

Elementary flux modes and extreme pathways can be used to understand the range of metabolic pathways in a network, to test a set of enzymes for production of a desired product and detect nonredundant pathways, to reconstruct metabolism from annotated genome sequences and analyze the effect of enzyme deficiency, to reduce drug effects, and to identify drug targets. A specific application, the flux balance analysis will be explained in Section 3.2.1.

3.1.5 Conservation Relations – Null Space of $\mathbf{N}^T$

If a chemical entity is neither added to nor removed from the reaction system (neither produced nor degraded), its total concentration remains constant. This also holds if the substance interacts with other compounds by forming complexes.

For the mathematical derivation of the conservation relations [3], we consider a matrix $\mathbf{G}$ fulfilling

$$\mathbf{GN} = \mathbf{0}. \tag{3.16}$$

Due to Eq. (3.5), it follows

$$\mathbf{G}\dot{\mathbf{S}} = \mathbf{GNv} = \mathbf{0}. \tag{3.17}$$

Integrating this equation leads directly to the conservation relations

$$\mathbf{GS} = \text{constant}. \tag{3.18}$$

The number of independent rows of $\mathbf{G}$ is equal to $n - \text{Rank}(\mathbf{N})$, where n is the number of metabolites in the system. $\mathbf{G}^T$ is the kernel matrix of $\mathbf{N}^T$; hence, it has similar properties to $\mathbf{K}$. Matrix $\mathbf{G}$ can also be found using the Gauss algorithm. It is not unique, but every linear combination of its rows is again a valid solution (equivalent to a premultiplication of $\mathbf{G}$ with a regular matrix of appropriate size, i.e., $\mathbf{PG} = \mathbf{G}'$). There exists a simplest representation $\mathbf{G} = \begin{pmatrix} \mathbf{G}_0 & \mathbf{I}_{n-\text{Rank}(\mathbf{N})} \end{pmatrix}$. Finding this representation may be helpful for a simple statement of conservation relations, but this may necessitate renumbering and reordering of metabolite concentrations (see below).

Importantly, conservation relations can be used to simplify the system of differential equations $\dot{\mathbf{S}} = \mathbf{Nv}$ describing the dynamics of our reaction system. The idea is to eliminate linear dependent differential equations and to replace them by appropriate algebraic equations. Below the procedure is explained systematically [2].

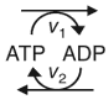
First we have to reorder the rows in the stoichiometric matrix $\mathbf{N}$ as well as in the concentration vector $\mathbf{S}$ such that a set of independent rows is on top and the dependent rows are at the bottom. Then the matrix $\mathbf{N}$ is split into the independent part $\mathbf{N}_{\text{indep}}$ and the dependent part $\mathbf{N}'$ and a *link matrix* $\mathbf{L}$ is introduced in the following way:

$$\mathbf{N} = \begin{pmatrix} \mathbf{N}_{\text{indep}} \\ \mathbf{N}' \end{pmatrix} = \mathbf{L}\mathbf{N}_{\text{indep}} = \begin{pmatrix} \mathbf{I}_{\text{Rank}(\mathbf{N})} \\ \mathbf{L}' \end{pmatrix} \mathbf{N}_{\text{indep}}. \tag{3.22}$$

$\mathbf{I}_{\text{Rank}(\mathbf{N})}$ is the identity matrix of size $\text{Rank}(\mathbf{N})$. The

Example 3.4

Consider a set of two reactions comprising a kinase and a phosphatase reaction

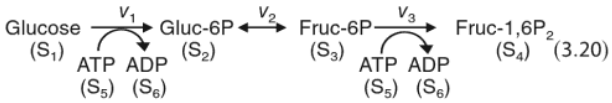


(3.19)

The metabolite concentration vector reads $\mathbf{S} = \begin{pmatrix} ATP & ADP \end{pmatrix}^T$, and the stoichiometric matrix is $\mathbf{N} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$ yielding $\mathbf{G} = \begin{pmatrix} 1 & 1 \end{pmatrix}$. From the condition $\mathbf{GS}$=constant, it follows that $ATP+ADP$=constant. Thus, we have a conservation of adenine nucleotides in this system. The actual values of $ATP+ADP$ must be determined from the initial conditions.

Example 3.5

For the following model of the upper part of glycolysis



(3.20)

the stoichiometric matrix $\mathbf{N}$ (note the transpose!) and a possible representation of the conservation matrix $\mathbf{G}$ are given by

$$\mathbf{N}^T = \begin{pmatrix} -1 & 1 & 0 & 0 & -1 & 1 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & -1 & 1 \end{pmatrix} \quad \text{and}$$

$$\mathbf{G} = \begin{pmatrix} 2 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \mathbf{g}_3 \end{pmatrix}. \tag{3.21}$$

The interpretation of the second and third rows of $\mathbf{G}$ is straightforward, showing the conservation of adenine nucleotides ($\mathbf{g}_2$, $ADP + ATP$ = constant) and the conservation of sugars ($\mathbf{g}_3$), respectively. The interpretation of the first row is less intuitive. If we construct the linear combination $\mathbf{g}_4 = -\mathbf{g}_1 + 3 \cdot \mathbf{g}_2 + 2 \cdot \mathbf{g}_3 = \begin{pmatrix} 0 & 1 & 1 & 2 & 3 & 2 \end{pmatrix}$, we find the conservation of phosphate groups.

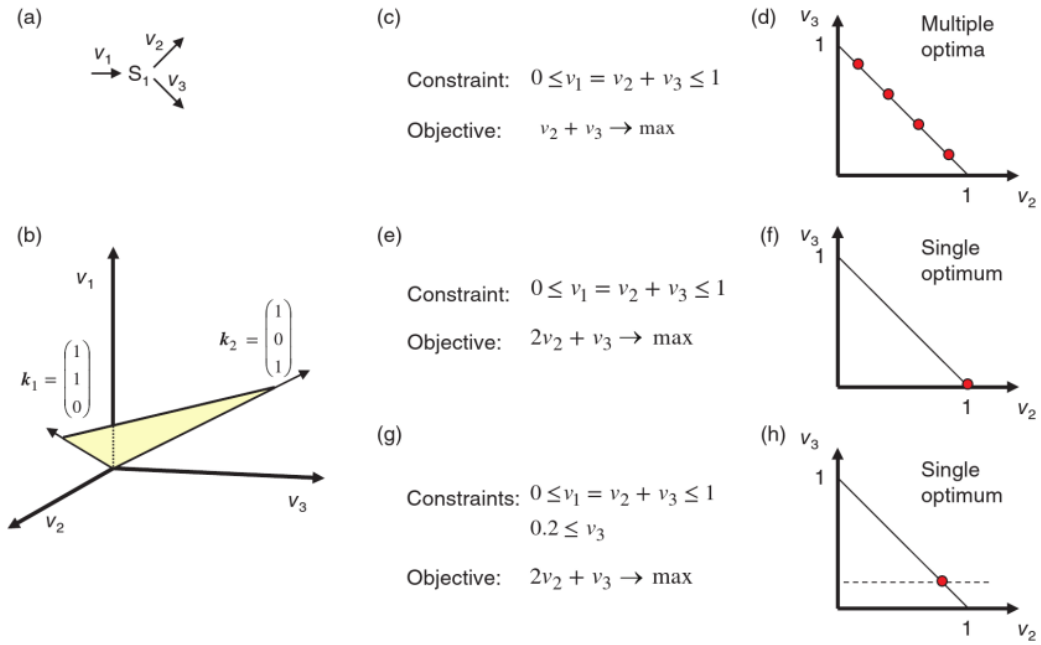


Figure 3.3 Constraint-based flux balance analysis. (a) Simple branched network. (b) The kernel vectors $\mathbf{k}_1$ and $\mathbf{k}_2$ span a plane of admissible steady-state fluxes in the flux space, that is, a two-dimensional flux cone. All solutions must lie on this plane. (c–h) Different examples for constraints, objectives, and resulting optimal fluxes. (c and d) The constraint of an upper and a lower bound for v_1 and the objective of maximizing v_1 (here equivalently to maximizing $v_2 + v_3$) yield infinitely many optimal solutions lying on the line $v_2 + v_3 = 1$. (e and f) If the objective is instead $2v_2 + v_3 \rightarrow$ max, we obtain a single optimal solution at $v_2 = 1,\ v_3 = 0$. (g and h) The stronger constraint $0.2 \leq v_3$ shifts the solution to $v_2 = 0.8,\ v_3 = 0.2$.

(or increase) in free energy between two compounds in a network is independent of the reaction path that is taken to get from one compound to the other one. According to the second law of thermodynamics, the Gibbs free energy must decrease in any occurring reaction. For a forward reaction, the difference of chemical potentials (Equation 3.33) must be negative and the reaction affinity must be positive. In general, for a reaction j, we obtain the condition

$$\sum_i \mu_i n_{ij} v_j \leq 0. \tag{3.35}$$

Therefore, a given flux pattern $\mathbf{v} = (v_1, \ldots, v_r)^{\mathrm{T}}$ is only feasible if condition (3.35) can be satisfied by some vector $(\mu_1, \ldots, \mu_n)^{\mathrm{T}}$ of chemical potentials. This condition can be tested using the stoichiometric matrix [24].

Flux balance analysis does not require that condition (3.35) is fulfilled and can therefore lead to incorrect flux signs. This problem can be avoided by predefining some of the flux directions, which will restrict the solution space in advance. *Energy balance analysis* [25,26], in contrast, ensures thermodynamically feasible fluxes by a joint optimization of the fluxes v_j and the chemical potential differences $\Delta_r G_j$. Besides the conditions (3.31), it imposes the additional requirements (3.34) and (3.35), which leads to an optimization problem with nonlinear constraints.

The chemical potentials are related not only to the flux directions, but also to the substance concentrations: for an ideal mixture (with vanishing mixing enthalpy), the chemical potential of substance i at pressure p and temperature T reads

$$\mu_i(p, T) = \mu_i^0(p, T) + RT \ln S_i, \tag{3.36}$$

where S_i denotes the concentration of metabolite i in mM. If the standard chemical potentials μ_i^0 are known (e.g., calculated by the group contribution method [27,28]), Eq. (3.36) translates to constraints between flux directions and substance concentrations. These constraints can be used to determine ranges of possible substance concentrations or to check whether measured concentrations are in agreement with the assumed fluxes [29,30].

3.2.4 Applications and Tests of the Flux Optimization Paradigm

Constraint-based methods such as flux balance analysis allow us to predict the metabolic fluxes and the biomass production (corresponding to the maximal growth rate) under different external conditions, for example,

availability of nutrients. The predictions can be used to simulate dynamically the growth of cell populations and the consumption of nutrients [19]. By comparing the predictions of FBA (biomass production or metabolic fluxes) with experimental data, one can check the assumed network structure for errors (e.g., missing reactions) and test Boolean models of gene regulation [31]. For instance, a low predicted growth rate would indicate that the organism is not viable. By testing the networks for deletion mutants, essential genes can be predicted [32]. The accuracy of such predictions (92% for a *Escherichia coli* model [16]) can be used as a quality score to check the consistency of the model structure and to point to missing reactions.

This approach implies that flux patterns in wild-type and mutant cells, under different external conditions, are optimized for the same general objective function. However, a study by Schuetz *et al.* [33] indicates that cells may optimize different objectives depending on the experimental conditions: metabolic fluxes in the central metabolism of *E. coli* cells were compared with predictions based on 11 alternative (linear and nonlinear) objective functions. Under glucose limitation in continuous cultures, cells seemed to maximize their yield of ATP or biomass per glucose consumed. Unlimited growth on glucose in respiring batch cultures, on the other hand, was best described by assuming a maximization of ATP production divided by the sum of squared reaction fluxes. This modified objective can be interpreted as a compromise between large ATP production and small enzymatic costs.

Such considerations of minimal effort had been formulated before in the *principle of minimal fluxes* [34]. Large reaction velocities require large amounts of enzymes, which put a burden on the cell. If the cost of enzyme production plays a role, cells will benefit from flux patterns that require less enzyme production, so pathways that do not contribute to biomass production (or whatever quantity is maximized) should be shut off to save energy and material. The principle of minimal fluxes assumes that the flux pattern has to meet some functional requirement – for example, to yield a prescribed rate of biomass production – while the magnitudes of individual reaction fluxes are minimized.

Even if we accept the assumption of optimality in general, constraint-based methods (i) do not explain by which biological mechanisms changes in flux distributions are actually achieved (e.g., inherent dynamics of the metabolic network, transcriptional regulation), (ii) do not cover the trade-off between cost and benefit of enzyme production, (iii) rely, instead, on ad-hoc assumptions, for example, about maximal fluxes, and (iv) assume a steady state and do not account for dynamic objectives such as fast adaptation to changes of supply and demand.

We will further address the issue of optimality of biological systems and the application to detect organization principles in Chapter 11.

3.2.5 Extensions of Flux Balance Analysis

Metabolic networks are embedded in a highly regulated cellular environment. A number of studies and approaches extended flux balance analysis to address specific biological observations, to integrate molecular and cellular information, and to accommodate further general principles. They are briefly summarized here.

3.2.5.1 Minimization of Metabolic Adjustments

Minimization of metabolic adjustments (MoMA) is a flux-based analysis technique similar to FBA [35]. It is based on the same stoichiometric constraints, but the demand of optimal growth flux for mutants is relaxed. Instead, MoMA provides an approximate solution for a suboptimal growth flux state, which is nearest in flux distribution to the unperturbed state. It is based on the assumption that in case of a knockout of an enzyme-coding gene metabolic fluxes undergo a minimal redistribution with respect to the flux configuration of the wild type.

The mathematical formulation of these requirements leads to a quadratic programming problem:

$$\begin{aligned} &\text{Constraint}: && \mathbf{Nv} = \mathbf{0} \\ &\text{Objective}: && \left\|\mathbf{v}_w - \mathbf{v}_d\right\|^2 \rightarrow \min \end{aligned} \tag{3.37}$$

with $\mathbf{v}_w$ presenting the fluxes of the wild type and $\mathbf{v}_d$ the fluxes for the gene deletion mutant. Figure 3.4 shows a comparison of FBA and MoMA for an illustrative example.

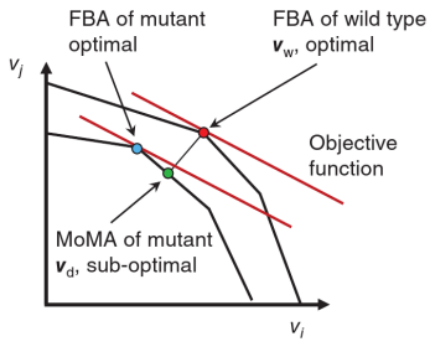


Figure 3.4 The principle of MoMA in comparison with FBA. The coordinates represent two selected fluxes, the outer black line the feasible state space for the wild type and the inner black line the feasible state space for the knockout mutant. The red line shows the objective function. FBA for the wild type yields the solution indicated by the red dot, while FBA for the mutant would result in the solution at the blue dot. MoMA requests that the distance between wild-type FBA solution and the newly attained state is minimal. This is given by the flux distribution marked by the green dot.

3.2.5.2 Flux Variability Analysis

Flux balance analysis often provides not a unique solution but the objective may be achieved by a whole range of alternate optimal solutions. Flux variability analysis (FVA) provides the ranges of possible flux through each individual reaction that is compatible with the steady-state and optimality conditions for the system as a whole [36]. A potential interpretation of the result is that a reaction with a small range of possible fluxes is more important for the functioning of the metabolism than reactions with a wide range of allowed flux solutions. FastFVA is an implementation based on an enhanced algorithm providing solutions significantly faster [37]. Figure 3.5 provides an illustration.

3.2.5.3 Dynamic FBA

Dynamic FBA is an extension of FBA for situations where changes in the metabolic network are relevant and an attempt to adapt the model to these changes over time [38]. This is achieved by a relaxation of the strict assumptions underlying steady-state analysis. The problem can be formulated either as a dynamic optimization problem starting at a given set of initial conditions for the metabolite concentrations or as a static optimization for a set of time intervals. In the second case, the time course is obtained by re-running FBA repeatedly with changing conditions. Since initial metabolite concentrations are given, the update rules also allow to calculate metabolite time profiles in a linearized fashion, that is, $S_i(t + \Delta T) = S_i(t) + \mathbf{N}\mathbf{v}\Delta T$ for time steps of duration ΔT.

3.2.5.4 Regulatory FBA

Metabolic networks are subject to external and internal changes and their dynamics are influenced both by the available amount of nutrients and by the expression of genes coding for the metabolic enzymes. To accommodate the transcriptional regulation of enzymes in the FBA framework, regulatory events may impose temporary, adjustable constraints on the solution space, such as

$$v_k(t) = 0, \quad \text{when} \quad t_1 \le t \le t_2, \tag{3.38}$$

instead of the constant constraints formulated in Eq. (3.29) [39]. The regulatory constraints change the shape of the accessible solution space. Consider, for example, the network N3 given in Table 3.1 and illustrated in Figure 3.3. If we assume a knockout of the gene coding for the enzyme of the second reaction and, hence, set $v_2(t) = 0$ for a certain period, only solution $\mathbf{k}_2 = \begin{pmatrix} 1 & 0 & 1 \end{pmatrix}^{\mathrm{T}}$ would remain. This is equivalent to say that $v_1 = v_3$ or that the entire flux goes through only one of the branches.

For more complex networks with mutual influence of genes on each other, the transcriptional regulatory structure can be described with Boolean logic (Section 7.1), assigning a value of 1 to an expressed gene and a value of 0 to a nonexpressed gene. The relation between genes and their expression dynamics is then described with Boolean rules, that is, combinations of operators such as AND, OR, or NOT. A schematic of the two interconnected networks is represented in Figure 3.6. This approach has been used to study

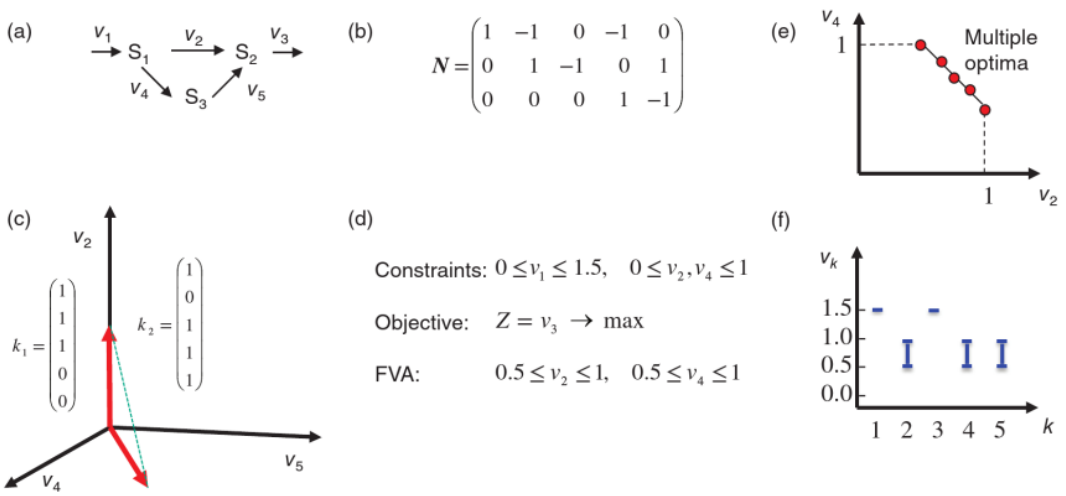


Figure 3.5 Flux variability analysis. For the example network in (a) with the given stoichiometry in (b), we find two representations of the kernel vector **K** represented in (c). The optimization problem formulated in (d) results in multiple solutions where the objective function Z is maximal; fluxes v_2 and v_4 can still vary between 0.5 and 1; however, their sum must equal 1.5, that is, the maximal input flux v_1. (e) shows the multiple optima in a phase plane spanned by the fluxes v_2 and v_4. (f) shows the variability of fluxes in conditions that maximize the objective function under the given constraints.

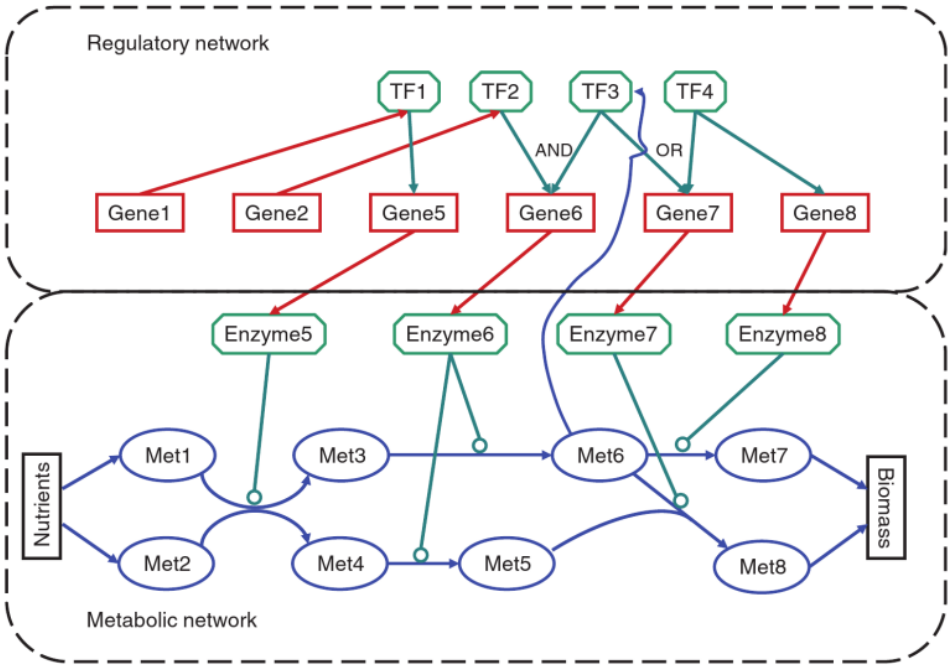


Figure 3.6 Schematic representation of an integrated metabolic and regulatory network. TF: transcription factors; Met: metabolites. Genes code for transcription factors and metabolic enzymes, which in turn catalyze the metabolic reactions. The purpose of the metabolic network is to produce biomass from the nutrients.

whether microorganisms such as *E. coli* can live (i.e., produce sufficient biomass) when living on exhaustible carbon sources and in which order the carbon sources are consumed. Extended analyses integrated large-scale metabolic models with regulatory models for gene expression [40] and for signaling [41] in *E. coli* to study gene expression variability and response to stimuli within the FBA framework.

Exercises

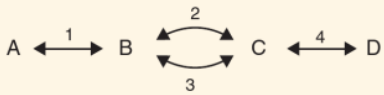
1) A canonical view of the upper part of glycolysis starts with glucose and comprises the following reactions (in brackets: possible abbreviations): The enzyme hexokinase (HK, E_1) phosphorylates glucose (Gluc, S_1) to glucose-6-phosphate (G6P, S_2) under consumption of ATP (S_5) and production of ADP (S_6). The enzyme phosphoglucoisomerase (PGI, E_2) converts glucose-6-phosphate to fructose-6-phosphate (F6P, S_3). The enzyme phosphofructokinase (PFK, E_3) phosphorylates F6P a second time to yield fructose-1,6-bisphosphate (F1,6BP, S_4). The enzyme fructose bisphosphatase catalyzes the reverse reaction (E_4).
   a) Sketch the reaction network and formulate a set of differential equations (without specifying the kinetics of the individual reactions).
   b) Formulate the stoichiometric matrix **N**. What is the rank of **N**?
   c) Calculate steady-state fluxes (matrix **K**) and conservation relations (matrix **G**).
   d) Compare your results with Example 3.5.

2) a) Write down the sets of differential equations for the networks N1–N6 given in Table 3.1 without specifying their kinetics.
   b) Determine the rank of the stoichiometric matrices, independent steady-state fluxes, and conservation relations.
   Do all systems have a (nontrivial) steady state?

3) Inspect networks N3 and N4 in Table 3.1. Can you find elementary flux modes? Use an available tool (e.g., Metatool) to check.

4) Consider the branch point in Figure 3.2a, with reactions $A \rightarrow X$, $B \rightarrow X$, and $X \rightarrow C$. The concentrations of A, B, and C are fixed and the reactions are irreversible. (a) Assume upper bounds $v_1 \leq 1$, $v_2 \leq 2$ and the fitness function $f(\mathbf{v}) = v_3$. Write down the corresponding linear programming problem and compute the resulting flux distribution. (b) Assume, in addition, an upper bound $v_3 \leq 1$. Draw the allowed region in flux space and determine the optimal flux distribution.

5) Show that a circular conversion flux $A \rightarrow B \rightarrow C \rightarrow A$ is thermodynamically unfeasible. Consider the following reaction scheme



with fixed concentrations of A and B and balanced metabolites C and D. Determine all flux distributions that are both stationary and thermodynamically feasible.

References

1 Glansdorff, P. and Prigogine, I. (1971) *Thermodynamic Theory of Structure, Stability and Fluctuations*, Wiley-Interscience, London.
2 Reder, C. (1988) Metabolic control theory: a structural approach. *J. Theor. Biol.*, 135, 175–201.
3 Heinrich, R. and Schuster, S. (1996) *The Regulation of Cellular Systems*, Chapman & Hall, New York.
4 Michal, G. (1999) *Biochemical Pathways*, Spektrum Akademischer Verlag, Heidelberg.
5 Schilling, C.H., Schuster, S., Palsson, B.O., and Heinrich, R. (1999) Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog.*, 15, 296–303.
6 Pfeiffer, T., Sanchez-Valdenebro, I., Nuno, J.C., Montero, F., and Schuster, S. (1999) METATOOL: for studying metabolic networks. *Bioinformatics*, 15, 251–257.
7 Schuster, S., Dandekar, T., and Fell, D.A. (1999) Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol.*, 17, 53–60.
8 Schuster, S., Fell, D.A., and Dandekar, T. (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.*, 18, 326–332.
9 Schuster, S., Hilgetag, C., Woods, J.H., and Fell, D.A. (2002) Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. *J. Math. Biol.*, 45, 153–181.
10 Schuster, S., Hilgetag, C., and Fell, D.A. (1994) Detecting elementary modes of functioning in metabolic networks, in *What Is Controlling Life?* (eds E. Gnaiger, F.N. Gellerich, and M. Wyss), Innsbruck University Press, Innsbruck, pp. 103–105.
11 Schilling, C.H. and Palsson, B.O. (2000) Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. *J. Theor. Biol.*, 203, 249–283.
12 Schilling, C.H., Letscher, D., and Palsson, B.O. (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.*, 203, 229–248.
13 Wiback, S.J. and Palsson, B.O. (2002) Extreme pathway analysis of human red blood cell metabolism. *Biophys. J.*, 83, 808–818.
14 Schilling, C.H. and Palsson, B.O. (1998) The underlying pathway structure of biochemical reaction networks. *Proc. Natl. Acad. Sci. USA*, 95, 4193–4198.
15 Reed, J.L., Famili, I., Thiele, I., and Palsson, B.O. (2006) Towards multidimensional genome annotation. *Nat. Rev. Genet.*, 7, 130–141.
16 Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R. *et al.* (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.*, 3, 121.
17 Thiele, I. and Palsson, B.O. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.*, 5, 93–121.
18 Fell, D.A. and Small, J.R. (1986) Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochem. J.*, 238, 781–786.
19 Varma, A. and Palsson, B.O. (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.*, 60, 3724–3731.
20 Varma, A. and Palsson, B.O. (1994) Predictions for oxygen supply control to enhance population stability of engineered production strains. *Biotechnol. Bioeng.*, 43, 275–285.
21 Edwards, J.S. and Palsson, B.O. (2000) Metabolic flux balance analysis and the *in silico* analysis of *Escherichia coli* K-12 gene deletions. *BMC Bioinform.*, 1, 1.
22 Edwards, J.S. and Palsson, B.O. (2000) The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. USA*, 97, 5528–5533.
23 Ramakrishna, R., Edwards, J.S., McCulloch, A., and Palsson, B.O. (2001) Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. *Am. J. Physiol. Regul. Integr. Comp. Physiol.*, 280, R695–R704.
24 Price, N.D., Schellenberger, J., and Palsson, B.O. (2004) Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies. *Biophys. J.*, 87, 2172–2186.
25 Beard, D.A., Babson, E., Curtis, E., and Qian, H. (2004) Thermodynamic constraints for biochemical networks. *J. Theor. Biol.*, 228, 327–333.
26 Beard, D.A., Liang, S.D., and Qian, H. (2002) Energy balance for analysis of complex metabolic networks. *Biophys. J.*, 83, 79–86.
27 Mavrovouniotis, M.L. (1990) Group contributions for estimating standard Gibbs energies of formation of biochemical compounds in aqueous solution. *Biotechnol. Bioeng.*, 36, 1070–1082.
28 Jankowski, M.D., Henry, C.S., Broadbelt, L.J., and Hatzimanikatis, V. (2008) Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophys. J.*, 95, 1487–1499.
29 Henry, C.S., Jankowski, M.D., Broadbelt, L.J., and Hatzimanikatis, V. (2006) Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. *Biophys. J.*, 90, 1453–1461.
30 Kummel, A., Panke, S., and Heinemann, M. (2006) Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol. Syst. Biol.*, 2, 2006.0034.

Kinetic Models of Biochemical Networks: Introduction

4

4.1 Reaction Kinetics and Thermodynamics

Summary

Kinetic modeling of metabolic reactions has a long tradition and forms the basis of many complex models for metabolic and regulatory networks. In this chapter, we will make you familiar with the basic concept of kinetic models for specific reactions. We introduce the mass action rate law, Michaelis–Menten kinetics, and different extended, applied, or contemporary modeling approaches. The role of the major parameters, K_m and V_{max} is explained. K_m and V_{max} are also related to the parameters of single reaction steps. We will show how you can derive and apply more advanced kinetic expressions. The effect of modifiers – activators and inhibitors – is shown for different kinetic mechanisms. Thermodynamic laws determine and limit the dynamic behavior and the steady state of kinetic systems; therefore, thermodynamic foundations and constraints are briefly introduced.

4.1.1 Kinetic Modeling of Enzymatic Reactions

Deterministic kinetic modeling of individual biochemical reactions has a long history. The Michaelis–Menten model for the rate of an irreversible one-substrate reaction is an integral part of biochemistry and has recently celebrated its centenary. The K_m value is a major characteristic of the interaction between enzyme and substrate. Biochemical reactions are catalyzed by enzymes, that is, specific proteins or ribonucleic acids, which often function in complex with cofactors. They have a catalytic center, are usually highly specific, and remain unchanged by the reaction. One enzyme molecule can catalyze thousands of reactions per second (this so-called turnover number ranges from 10^2 to 10^7 s^{-1}). Enzyme catalysis leads to a rate acceleration of about 10^6 up to 10^{12}-fold compared to the noncatalyzed, spontaneous reaction.

The basic quantities are the concentration S of a substance S, that is, the number n of molecules (or, alternatively, moles) of this substance per volume V, and the rate v of a reaction, that is, the change of concentration S per time t. This type of modeling is macroscopic and phenomenological, compared to the microscopic approach, where single molecules and their interactions are considered. Chemical and biochemical kinetics rely on the assumption that the reaction rate v at a certain point in time and space can be expressed as a unique function of

Systems Biology: A Textbook, Second Edition. Edda Klipp, Wolfram Liebermeister, Christoph Wierling, and Axel Kowald.

the concentrations of all substances at this point in time and space. Classical enzyme kinetics assumes for sake of simplicity a spatial homogeneity (the "well-stirred" test tube) and no direct dependency of the rate on time:

$$v(t) = v(S(t)). \tag{4.1}$$

In more advanced modeling approaches paving the way for whole cell modeling, spatial inhomogeneities are taken into account. Spatial modeling pays tribute to the fact that many components are membrane bound and that cellular structures hinder the free movement of molecules. But, in most cases one can assume that diffusion is rapid enough to allow for an even distribution of all substances in space.

4.1.2 The Law of Mass Action

Biochemical kinetics is based on the mass action law, introduced by Guldberg and Waage in the nineteenth century [1–3]. It states that the reaction rate is proportional to the probability of a collision of the reactants. This probability is in turn proportional to the concentration of reactants to the power of the molecularity, which is the number in which the molecule species enter the reaction. For a simple reaction such as

$$S_1 + S_2 \rightleftharpoons 2P, \tag{4.2}$$

the reaction rate reads

$$v = v_+ - v_- = k_+ S_1 \cdot S_2 - k_- P^2, \tag{4.3}$$

where v is the net rate, v_+ and v_- are the rates of the forward and backward reactions, respectively, and k_+ and k_- are the *kinetic* or *rate constants*, that is, the respective proportionality factors.

The molecularity is 1 for S_1 and S_2 and 2 for P. If we measure the concentration in moles per liter ($\mathrm{mol} \cdot \mathrm{l}^{-1}$ or M) and the time in seconds (s), then the rate has the unit $\mathrm{M} \cdot \mathrm{s}^{-1}$. Accordingly, the rate constants for bimolecular reactions have the unit $\mathrm{M}^{-1}\mathrm{s}^{-1}$. Rate constants for monomolecular reactions have the dimension s^{-1}.

The general mass action rate law for a reaction transforming m_i substrates with concentrations S_i into m_j products with concentrations P_j reads

$$v = v_+ - v_- = k_+ \prod_{i=1}^{m_i} S_i^{n_i} - k_- \prod_{j=1}^{m_j} P_j^{n_j}, \tag{4.4}$$

where n_i and n_j denote the respective molecularities of S_i and P_j in this reaction.

The equilibrium constant K_{eq} (we will also use the simpler symbol q) characterizes the ratio of substrate and product concentrations in equilibrium (S_{eq} and P_{eq}), that is, the state where the thermodynamic affinity vanishes and where the forward and backward rates become equal. The rate constants are related to K_{eq} in the following way:

$$K_{\mathrm{eq}} = \frac{k_+}{k_-} = \frac{\prod_{j=1}^{n_j} P_{j,\mathrm{eq}}}{\prod_{i=1}^{n_i} S_{i,\mathrm{eq}}}. \tag{4.5}$$

The relation between the thermodynamic and the kinetic description of biochemical reactions will be outlined in Section 4.1.3.

The equilibrium constant for the reaction given in Eq. (4.2) is $K_{\mathrm{eq}} = P_{\mathrm{eq}}^2 / \left(S_{1,\mathrm{eq}} \cdot S_{2,\mathrm{eq}}\right)$. The dynamics of the concentrations far from equilibrium is described by the ODEs

$$\frac{\mathrm{d}}{\mathrm{d}t} S_1 = \frac{\mathrm{d}}{\mathrm{d}t} S_2 = -v \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}t} P = 2v. \tag{4.6}$$

The time course of S_1, S_2, and P is obtained by integration of these ODEs (see Section 15.2).

4.1.3 Reaction Thermodynamics

Biochemical reactions in isolation or as part of a larger reaction network are governed by the laws of

Example 4.1

The kinetics of a simple decay like

$$S \rightarrow \tag{4.7}$$

is described by $v = kS$ and $\mathrm{d}S/\mathrm{d}t = -kS$. Integration of this ODE from time $t = 0$ with the initial concentration S_0 to an arbitrary time t with concentration $S(t)$, $\int_{S_0}^{S} \mathrm{d}S/S = -\int_{t=0}^{t} k\mathrm{d}t$, yields the temporal expression $S(t) = S_0\, \mathrm{e}^{-kt}$ (Figure 4.1).

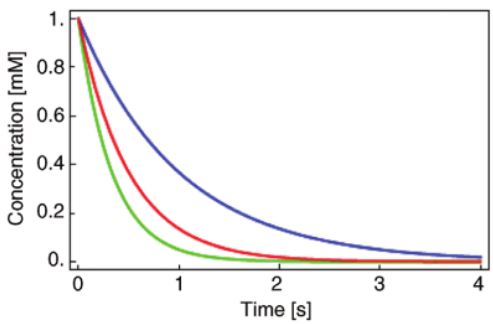


Figure 4.1 Exponential decay of a compound as described in Eq. (4.7). The initial concentration is $S_0 = 1$ mM. The dynamics are shown for three different values of k: $k = 1 \cdot \mathrm{s}^{-1}$ (blue), $k = 2 \cdot \mathrm{s}^{-1}$ (red), and $k = 3 \cdot \mathrm{s}^{-1}$ (green).

thermodynamics. This means that they cannot create or destroy energy, they can only convert it or store it in chemical bonds or release it from there. An important purpose of metabolism is to extract energy from nutrients, which is necessary for the synthesis of molecules, growth, and proliferation. We distinguish between energy-supplying reactions, energy-demanding reactions, and energetically neutral reactions. The principles of reversible and irreversible thermodynamics and their application to chemical reactions allow understanding of energy circulation in the cell.

A biochemical process is characterized by the direction of the reaction, by whether it occurs spontaneously or not, and by the position of the equilibrium. The first law of thermodynamics, that is, the law of energy conservation, tells us that the total energy of a closed system remains constant during any process. The second law of thermodynamics states that a process occurs spontaneously only if it increases the total entropy of the system. Unfortunately, entropy is usually not directly measurable. A more suitable measure is the Gibbs free energy G, which is the energy capable of carrying out work under isotherm–isobar conditions, that is, at constant temperature and constant pressure. The change of the Gibbs free energy is given as

$$\Delta G = \Delta H - T\Delta S, \tag{4.8}$$

where ΔH is the change in enthalpy, ΔS is the change in entropy, and T is the absolute temperature in Kelvin. ΔG is a measure for the driving force, the spontaneity of a chemical reaction. The reaction proceeds spontaneous under release of energy, if $\Delta G < 0$ (exergonic process). If $\Delta G > 0$, then the reaction is energetically not favorable and will not occur spontaneously (endergonic process). $\Delta G = 0$ implies that the system has reached its equilibrium. Endergonic reactions may proceed if they obtain energy from a strictly exergonic reaction by energetic coupling. In tables, Gibbs free energy is usually given for standard conditions (ΔG^0), that is, for a concentration of the reaction partners of 1 M, a temperature of $T = 298$ K, and, for gaseous reactions, a pressure of $p = 98.1$ kPa $= 1$ atm. The unit is kJ mol^{-1}. Gibbs free energy differences satisfy a set of relations as follows. The Gibbs free energy difference for a reaction can be calculated from the balance of free energies of formation of its products and substrates:

$$\Delta G = \sum G_P - \sum G_S. \tag{4.9}$$

The enzyme cannot change the Gibbs free energies of the substrates and products of a reaction, neither their difference, but it changes the way the reaction proceeds microscopically, the so-called reaction path, thereby lowering the activation energy for the reaction. The *transition state theory* explains this as follows. During the course of a reaction, the metabolites must pass one or more transition states of maximal free energy, in which bonds are solved or newly formed. The transition state is unstable; the respective molecule configuration is called an activated complex. It has a lifetime of around one molecule vibration, 10^{-14}–10^{-13} s, and it can hardly be experimentally verified. The difference $\Delta G^{\neq}$ of Gibbs free energy between the reactants and the activated complex determines the dynamics of a reaction: the higher this difference, the lower the probability that the molecules may pass this barrier and the lower the rate of the reaction. The value of $\Delta G^{\neq}$ depends on the type of altered bonds, on steric, electronic, or hydrophobic demands, and on temperature.

Figure 4.2 presents a simplified view of the reaction course of the noncatalyzed reaction and with an enzyme. The substrate and the product are situated in local minima of the free energy; the active complex is assigned to the local maximum. The Gibbs free energy difference ΔG is proportional to the logarithm of the equilibrium constant K_{eq} of the respective reaction:

$$\Delta G = -RT \ln K_{eq}, \tag{4.10}$$

where R is the gas constant, 8.314 J mol^{-1} K^{-1}. The value of $\Delta G^{\neq}$ corresponds to the kinetic constant k_+ of the forward reaction (Eqs. (4.3)–(4.5)) by $\Delta G^{\neq} = -RT \ln k_+$, while $\Delta G^{\neq} + \Delta G$ is related to the rate constant k_- of the backward reaction.

The interaction of the reactants with an enzyme may alter the reaction path and, thereby, lead to lower values of $\Delta G^{\neq}$ as well as higher values of the kinetic constants. However, the enzyme will not change the equilibrium

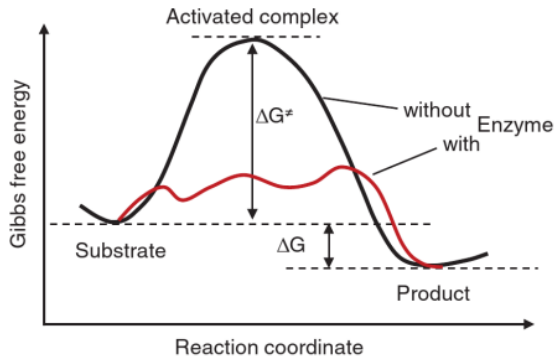


Figure 4.2 Change of Gibbs free energy along the course of a reaction. The substrate and the product are situated in local minima of the free energy; the active complex is assigned to the local maximum. The enzyme may change the reaction path and thereby lower the barrier of Gibbs free energy.

Table 4.1 Values of $\Delta G^{0\prime}$ and K_{eq} for some important reactions[a].

Reaction	$\Delta G^{0\prime}/(kJ\ mol^{-1})$
$2\,H_2 + O_2 \rightarrow 2\,H_2O$	−474
$2\,H_2O_2 \rightarrow 2\,H_2O + O_2$	−99
$PP_i + H_20 \rightarrow 2\,P_i$	−33.49
$ATP + H_20 \rightarrow ADP + P_i$	−30.56
Glucose-6-phosphate $+ H_20 \rightarrow$ Glucose $+ P_i$	−13.82
Glucose $+ P_i \rightarrow$ Glucose-6-phosphate $+ H_20$	+13.82
Glucose-1-phosphate → Glucose-6-phosphate	−7.12
Glucose-6-phosphate → Fructose-6-phosphate	+1.67
Glucose $+ 6\,O_2 \rightarrow 6\,CO_2 + 6\,H_20$	−2890

[a] *Source:* ZITAT: Lehninger, A.L. Biochemistry, 2nd edition, New York, Worth, 1975, p. 397.

constant of the reaction. The Gibbs free energy may assume several local minima and maxima along the path of reaction. They are related to unstable intermediary complexes. Values for the difference of free energy for some biologically important reactions are given in Table 4.1. Note that the free energy differences always refer to specific standard concentrations.

A biochemical reaction is reversible if it may proceed in both directions, leading to a positive or negative sign of the rate v. The actual direction depends on the current reactant concentrations. In theory, every reaction should be reversible. In practice, we can consider many reactions as irreversible, since (i) reactants in cellular environment cannot assume any concentration, (ii) coupling of a chemical conversion to ATP consumption leads to a severe drop in free energy and therefore makes a reaction reversal energetically unfavorable, and (iii) for compound destruction, such as protein degradation, reversal by chance is extremely unlikely.

The detailed consideration of enzyme mechanisms by applying the mass action law for the single events has led to a number of standard kinetic descriptions, which will be explained in the following. For further information on equilibrium thermodynamics in reaction systems also see Section 15.6.

4.1.4 Michaelis–Menten Kinetics

Brown [4] proposed an enzymatic mechanism for invertase, catalyzing the cleavage of saccharose to glucose and fructose. This mechanism holds in general for all one-substrate reactions without backward reaction and without effectors, such as

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_2} \mathrm{E} + \mathrm{P}. \tag{4.11}$$

It comprises a reversible formation of an enzyme–substrate complex ES from the free enzyme E and the substrate S and an irreversible release of the product P. The ODE system for the dynamics of this reaction reads

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -k_1 E \cdot S + k_{-1} ES, \tag{4.12}$$

$$\frac{\mathrm{d}ES}{\mathrm{d}t} = k_1 E \cdot S - (k_{-1} + k_2) ES, \tag{4.13}$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} = -k_1 E \cdot S + (k_{-1} + k_2) ES, \tag{4.14}$$

$$\frac{\mathrm{d}P}{\mathrm{d}t} = k_2 ES. \tag{4.15}$$

The reaction rate is equal to the negative decay rate of the substrate as well as to the rate of product formation:

$$v = -\frac{\mathrm{d}S}{\mathrm{d}t} = \frac{\mathrm{d}P}{\mathrm{d}t}. \tag{4.16}$$

This ODE system (Eqs. (4.12)–(4.16)) cannot be solved analytically. Figure 4.3 shows numerical solutions for different parameter sets.

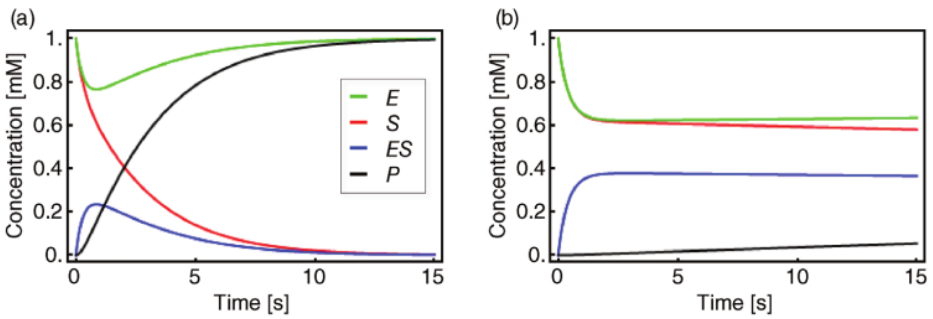


Figure 4.3 Temporal evolution of the equation system (4.12)–(4.15). Shown are S (red), E (green), ES (blue), and P (black). The initial concentrations are in both panels $S(0) = E(0) = 1$ mM and $ES(0) = P(0) = 0$ mM. Parameter values: $k_1 = 1 \cdot \mathrm{M}^{-1}\mathrm{s}^{-1}$, $k_2 = 1 \cdot \mathrm{s}^{-1}$, and either $k_3 = 1 \cdot \mathrm{s}^{-1}$ (a) or $k_3 = 0.01 \cdot \mathrm{s}^{-1}$ (b).

Different assumptions have been used to simplify this system in a satisfactory way. Michaelis and Menten [5] considered a *quasi-equilibrium* between the free enzyme and the enzyme–substrate complex, meaning that the reversible conversion of E and S to ES is much faster than the decomposition of ES into E and P, or in terms of the kinetic constants, that is,

$$k_1, k_{-1} \gg k_2. \tag{4.17}$$

This is the situation as shown in Figure 4.3b.

Briggs and Haldane [6] assumed that during the course of reaction a state is reached where the concentration of the ES complex remains constant, the so-called quasi-steady state. This assumption is justified only if the initial substrate concentration is much larger than the enzyme concentration, $S(t=0) \gg E$, otherwise such a state will never be reached. In mathematical terms, we obtain

$$\frac{\mathrm{d}ES}{\mathrm{d}t} = 0. \tag{4.18}$$

In the following, we derive an expression for the reaction rate from the ODE system (4.12)–(4.15) and the quasi-steady-state assumption for ES. First, adding Eqs. (4.13) and (4.14) results in

$$\frac{\mathrm{d}ES}{\mathrm{d}t} + \frac{\mathrm{d}E}{\mathrm{d}t} = 0 \quad \text{or} \quad E_{\text{total}} = E + ES = \text{constant}. \tag{4.19}$$

This expression shows that enzyme is neither produced nor consumed in this reaction; it may be free or part of the complex, but its total concentration remains constant. Introducing (4.19) into (4.13) under the steady-state assumption (4.18) yields

$$ES = \frac{k_1 E_{\text{total}} S}{k_1 S + k_{-1} + k_2} = \frac{E_{\text{total}} S}{S + ((k_{-1} + k_2)/k_1)}. \tag{4.20}$$

For the reaction rate, this gives

$$v = \frac{k_2 E_{\text{total}} S}{S + ((k_{-1} + k_2)/k_1)}. \tag{4.21}$$

In enzyme kinetics, it is convention to present Eq. (4.21) in a simpler form, which is important in theory and practice

$$v = \frac{V_{\max} S}{S + K_{\mathrm{m}}}. \tag{4.22}$$

Equation (4.22) is the expression for Michaelis–Menten kinetics. The parameters have the following meaning: the *maximal velocity*,

$$V_{\max} = k_2 E_{\text{total}}, \tag{4.23}$$

is the maximal rate that can be attained, when the enzyme is completely saturated with substrate. The *Michaelis constant*,

$$K_{\mathrm{m}} = \frac{k_{-1} + k_2}{k_1}, \tag{4.24}$$

is equal to the substrate concentration that yields the half-maximal reaction rate. For the quasi-equilibrium assumption (Eq. (4.17)), it holds that $K_{\mathrm{m}} \cong k_{-1}/k_1$. The maximum velocity divided by the enzyme concentration (here $k_2 = V_{\max}/E_{\text{total}}$) is often called the turnover number, k_{cat}. The meaning of the parameters is illustrated in the plot of rate versus substrate concentration (Figure 4.4).

4.1.4.1 How to Derive a Rate Equation

Below, we will present some enzyme kinetic standard examples. Individual mechanisms for your specific enzyme of interest may be more complicated or merely differ from these standards. Therefore, we summarize here the general way of deriving a rate equation.

1) Draw a wiring diagram of all steps to consider (e.g., Eq. (4.11)). It contains all substrates and products (S and P) and n free or bound enzyme species (E and ES).
2) The right sides of the ODEs for the concentrations changes sum up the rates of all steps leading to or away from a certain substance (e.g., Eqs. (4.12)–(4.15)). The rates follow mass action kinetics (Eq. (4.3)).
3) The sum of all enzyme-containing species is equal to the total enzyme concentration E_{total} (the right side of all differential equations for enzyme species sums up to zero). This constitutes one equation.
4) The assumption of quasi-steady state for $n-1$ enzyme species (i.e., setting the right sides of the respective ODEs equal to zero) together with (3) result in n

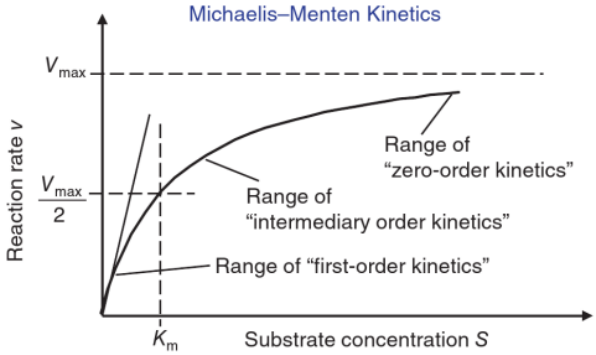


Figure 4.4 Dependence of reaction rate v on substrate concentration S in Michaelis–Menten kinetics. $V_{\max}$ denotes the maximal reaction rate that can be reached for large substrate concentration. K_{m} is the substrate concentration that results in half-maximal reaction rate. For low substrate concentration, v increases almost linearly with S, while for high substrate concentrations v is almost independent of S.

Table 4.3 Types of inhibition for irreversible and reversible Michaelis–Menten kinetics[b].

Name	*Implementation*	*Equation – irreversible*	*Equation – reversible case*	*Characteristics*
Competitive inhibition	I binds only to free E; P-release only from ES-complex $k_{\pm4} = k_{\pm5} = k_6 = 0$	$v = \frac{V_{max}S}{K_m \cdot i_3 + S}$	$v = \frac{V^f_{max}(S/K_{mS}) - V^r_{max}(P/K_{mP})}{(S/K_{mS}) + (P/K_{mP}) + i_3}$	K_m changes, V_{max} remains same. S and I compete for the binding place; high S may out compete I.
Uncompetitive Inhibition	I binds only to the ES-complex; P-release only from ES-complex $k_{\pm3} = k_{\pm5} = k_6 = 0$	$v = \frac{V_{max}S}{K_m + S \cdot i_4}$	$v = \frac{V^f_{max}(S/K_{mS}) - V^r_{max}(P/K_{mP})}{1 + ((S/K_{mS}) + (P/K_{mP})) \cdot i_4}$	K_m and V_{max} change, but their ratio remains same. S may not out compete I
Noncompetitive inhibition	I binds to E and ES; P-release only from ES $K_{I,3} = K_{I,4}$, $k_6 = 0$	$v = \frac{V_{max}S}{(K_m + S) \cdot i_3}$	$v = \frac{V^f_{max}(S/K_{mS}) - V^r_{max}(P/K_{mP})}{(1 + (S/K_{mS}) + (P/K_{mP})) \cdot i_4}$	K_m remains, V_{max} changes. S may not out compete I
Mixed inhibition	I binds to E and ES; P-release only from ES $K_{I,3} \neq K_{I,4}$, $k_6 = 0$	$v = \frac{V_{max}S}{K_m \cdot i_4 + S \cdot i_3}$		K_m and V_{max} change. $K_{I,3} > K_{I,4}$: competitive-noncompetitive inhibition $K_{I,3} < K_{I,4}$: noncompetitive-uncompetitive inhibition
Partial Inhibition	I may bind to E and ES; P-release from ES and ESI $K_{I,3} \neq K_{I,4}$, $k_6 \neq 0$	$v = \frac{V_{max}S\left(1 + \left((k_6 I)/k_2 K_{I,3}\right)\right)}{K_m \cdot i_4 + S \cdot i_3}$		K_m and V_{max} change if $k_6 > k_2$: activation instead of inhibition.

[b] The following abbreviations are used: $K_{I,3} = \frac{k_{-3}}{k_3}$, $K_{I,4} = \frac{k_{-4}}{k_4}$, $i_3 = 1 + \frac{I}{K_{I,3}}$, $i_4 = 1 + \frac{I}{K_{I,4}}$.

4.1.5.1 Substrate Inhibition

A common characteristic of enzymatic reaction is the increase of the reaction rate with increasing substrate concentration S up to the maximal velocity V_{max}. But in some cases, a decrease of the rate above a certain value of S is recorded. A possible reason is the binding of a further substrate molecule to the enzyme–substrate complex yielding the complex ESS that cannot form a product. This kind of inhibition is reversible if the second substrate can be released. The rate equation can be derived using the scheme of uncompetitive inhibition by replacing the inhibitor by another substrate. It reads

$$v = k_2 ES = \frac{V_{max}S}{K_m + S(1 + (S/K_I))}. \tag{4.33}$$

This expression has an optimum, that is, a maximal value of v, at

$$S_{opt} = \sqrt{K_m K_I} \quad \text{with} \quad v_{opt} = \frac{V_{max}}{1 + 2\sqrt{K_m/K_I}}. \tag{4.34}$$

The dependence of v on S is shown in Figure 4.6. A typical example for substrate inhibition is the binding of two succinate molecules to malonate dehydrogenase, which possesses two binding pockets for the carboxyl group. This is schematically represented in Figure 4.6.

4.1.5.2 Binding of Ligands to Proteins

Every molecule that binds to a protein is a ligand, irrespective of whether it is subject of a reaction or not. Below we consider binding to monomer and oligomer proteins. In oligomers, there may be interactions between the binding sites on the subunits.

Consider binding of one ligand (S) to a protein (E) with only one binding site:

$$\mathrm{E} + \mathrm{S} \rightleftharpoons \mathrm{ES}. \tag{4.35}$$

The binding constant K_B is given by

$$K_B = \left(\frac{ES}{E \cdot S}\right)_{eq}. \tag{4.36}$$

The reciprocal of K_B is the dissociation constant K_D. The fractional saturation Y of the protein is determined by the number of subunits that have bound ligands, divided by the total number of subunits. The fractional saturation for one subunit is

$$Y = \frac{ES}{E_{total}} = \frac{ES}{ES + E} = \frac{K_B \cdot S}{K_B \cdot S + 1}. \tag{4.37}$$

The plot of Y versus S at constant total enzyme concentration is a hyperbola, like the plot of v versus S in the Michaelis–Menten kinetics (Eq. (4.22)). At a process where the binding of S to E is the first step followed by

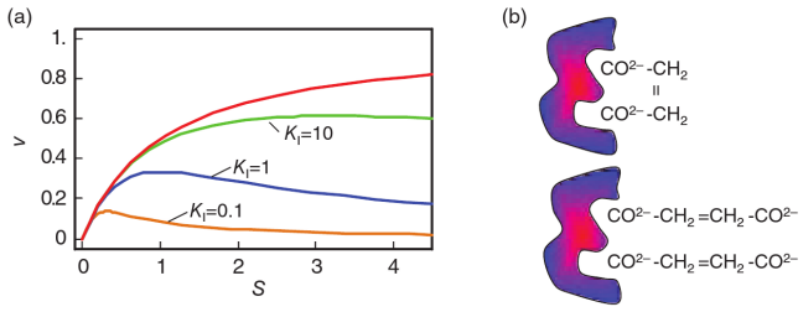


Figure 4.6 Substrate inhibition. (a) Plot of reaction rate v against substrate concentration S for an enzyme with substrate inhibition. The upper curve shows Michaelis–Menten kinetics without inhibition, the lower curves show kinetics for the indicated values of binding constant K_I. Parameter values: $V_{max} = 1$, $K_m = 1$. (b) Visualization of a possible mechanism for substrate inhibition: The enzyme (gray item) has two binding pockets to bind different parts of a substrate molecule (upper scheme). In case of high substrate concentration, two different molecules may enter the binding pockets, thereby preventing the specific reaction (lower scheme).

product release and where the initial concentration of S is much higher as the initial concentration of E, the rate is proportional to the concentration of ES and it holds

$$\frac{v}{V_{max}} = \frac{ES}{E_{total}} = Y. \tag{4.38}$$

If the protein has several binding sites, then interactions may occur between these sites, that is, the affinity to further ligands may change after binding of one or more ligands. This phenomenon is called *cooperativity*. Positive or negative cooperativity denote increase or decrease in the affinity of the protein to a further ligand, respectively. Homotropic or heterotropic cooperativity denotes that the binding to a certain ligand influences the affinity of the protein to a further ligand of the same or another type, respectively.

4.1.5.3 Positive Homotropic Cooperativity and the Hill Equation

Consider a dimeric protein with two identical binding sites. The binding to the first ligand facilitates the binding to the second ligand.

$$E_2 + S \xrightarrow{\text{slow}} E_2S, \quad E_2S + S \xrightarrow{\text{fast}} E_2S_2, \tag{4.39}$$

where E is the monomer and E_2 is the dimer. The fractional saturation is given by

$$Y = \frac{E_2S + 2 \cdot E_2S_2}{2 \cdot E_{2,total}} = \frac{E_2S + 2 \cdot E_2S_2}{2 \cdot E_2 + 2 \cdot E_2S + 2 \cdot E_2S_2}. \tag{4.40}$$

If the affinity to the second ligand is strongly increased by binding to the first ligand, then E_2S will react with S as soon as it is formed and the concentration of E_2S can be neglected. In the case of complete *cooperativity*, that is, every protein is either empty or fully bound, Eq. (4.39) reduces to

$$E_2 + 2S \rightarrow E_2S_2. \tag{4.41}$$

The binding constant reads

$$K_B = \frac{E_2S_2}{E_2 \cdot S^2}, \tag{4.42}$$

and the fractional saturation is

$$Y = \frac{2 \cdot E_2S_2}{2 \cdot E_{2,total}} = \frac{E_2S_2}{E_2 + E_2S_2} = \frac{K_B \cdot S^2}{1 + K_B \cdot S^2}. \tag{4.43}$$

Generally, for a protein with n subunits it holds:

$$v = V_{max} \cdot Y = \frac{V_{max} \cdot K_B \cdot S^n}{1 + K_B \cdot S^n}. \tag{4.44}$$

This is the general form of the *Hill equation*. To derive it, we assumed complete homotropic cooperativity. The plot of the fractional saturation Y versus substrate concentration S is a sigmoid curve with the inflection point at $1/K_B$. The quantity n (often "h" is used instead) is termed the *Hill coefficient*.

The derivation of this expression was based on experimental findings concerning the binding of oxygen to hemoglobin (Hb) [13,14]. In 1904, Bohr *et al.* found that the plot of the fractional saturation of Hb with oxygen against the oxygen partial pressure had a sigmoid shape. Hill (1909) explained this with interactions between the binding sites located at the hem subunits. At this time, it was already known that every subunit hem binds one molecule of oxygen. Hill assumed complete cooperativity and predicted an experimental Hill coefficient of 2.8. Today it is known that hemoglobin has four binding sites, but that the cooperativity is not complete. The sigmoid binding characteristic has the advantage that Hb binds

strongly to oxygen in the lung with a high oxygen partial pressure while it can release O_2 easily in the body with low oxygen partial pressure.

4.1.5.4 The Monod–Wyman–Changeux Model for Sigmoid Kinetics

The Monod model [15] explains sigmoid enzyme kinetics by taking into account the interaction of subunits of an enzyme. We will show here the main characteristics and assumptions of this kinetics. The full derivation is given in the web material. It uses the following assumptions: (i) the enzyme consists of n identical subunits, (ii) each subunit can assume an active (R) or an inactive (T) conformation, (iii) all subunits change their conformations at the same time (concerted change), and (iv) the equilibrium between the R and the T conformation is given by an allosteric constant

$$L = \frac{T_0}{R_0}. \tag{4.45}$$

The binding constants for the active and inactive conformations are given by K_R and K_T, respectively. If substrate molecules can only bind to the active form, that is, if $K_T = 0$, the rate can be expressed as

$$V = \frac{V_{max} K_R S}{(1 + K_R S)} \frac{1}{\left[1 + \{L/((1 + K_R S)^n)\}\right]}, \tag{4.46}$$

where the first factor $(V_{max} K_R S)/(1 + K_R S)$ corresponds to the Michaelis–Menten rate expression, while the second factor $\left[1 + (L/(1 + K_R S)^n)\right]^{-1}$ is a regulatory factor.

For $L = 0$, the plot v versus S is hyperbola as in Michaelis–Menten kinetics. For $L > 0$, we obtain a sigmoid curve shifted to the right. A typical value for the allosteric constant is $L \cong 10^4$ (Figure 4.7).

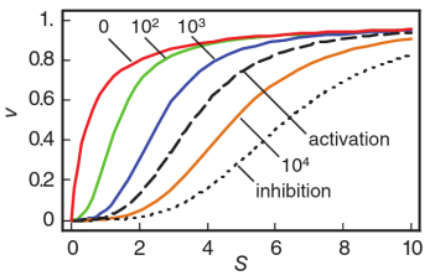


Figure 4.7 Model of Monod, Wyman, and Changeux: Dependence of the reaction rate on substrate concentration for different values of the allosteric constant L, according to Eq. (4.46). Parameters: $V_{max} = 1$, $n = 4$, $K_R = 2, K_T = 0$. The value of L is indicated at the curves. Obviously, increasing value of L causes stronger sigmoidity. The influence of activators or inhibitors (compare Eq. (4.47)) is illustrated with the dotted line for $K_I I = 2$ and with the dashed line for $K_A A = 2$ ($L = 10^4$ in both cases).

Up to now we considered in the model of Monod, Wyman, and Changeux only homotropic and positive effects. But this model is also well suited to explain the dependence of the reaction rate on activators and inhibitors. Activators A bind only to the active conformation and inhibitors I bind only to the inactive conformation. This shifts the equilibrium to the respective conformation. Effectively, the binding to effectors changes L:

$$L' = L \frac{(1 + K_I I)^n}{(1 + K_A A)^n}, \tag{4.47}$$

where K_I and K_A denote binding constants. The interaction with effectors is a heterotropic effect. An activator weakens the sigmoidity, while an inhibitor strengthens it.

A typical example for an enzyme with sigmoid kinetics that can be described with the Monod model is the enzyme phosphofructokinase, which catalyzes the transformation of fructose-6-phosphate and ATP to fructose-1,6-bisphosphate. AMP, NH_4, and K^+ are activators, ATP is an inhibitor.

4.1.6 Generalized Mass Action Kinetics

Mass action kinetics (see Section 4.1.1) has experienced refinements in different ways. The fact that experimental results frequently do not show the linear dependence of rate on concentrations as assumed in mass action laws is acknowledged in power law kinetics used in the S-systems approach. Here, the rate reads

$$\frac{v_j}{v_j^0} = k_j \prod_{i=1}^{n} \left(\frac{S_i}{S_i^0}\right)^{g_{j,i}}, \tag{4.48}$$

where the concentrations S_i and rates v_j are normalized to some standard value denoted by superscript 0, and $g_{i,j}$ is a real number instead of an integer as in Eq. (4.4). The normalization yields dimensionless quantities. The power law kinetics can be considered as a generalization of the mass action rate law. The exponent $g_{i,j}$ is equal to the concentration elasticities, that is, the scaled derivatives of rates with respect to substrate concentrations (see Section 4.3, Eq. (4.107)). Substrates and effectors (their concentrations both denoted by S_i) enter expression (4.48) in the same formal way, but the respective exponents $g_{i,j}$ will be different. The exponents $g_{i,j}$ will be positive for substrates and activators, but should assume a negative value for inhibitors.

4.1.7 Approximate Kinetic Formats

In metabolic modeling studies, approximate kinetic formats are used (for a recent review see Ref. [16]). They

preassume that each reaction rate v_j is proportional to the enzyme concentration E_j. The rates, enzyme concentrations, and substrate concentrations are normalized with respect to a references state, which is usually a steady state. This leads to the general expression

$$\frac{v_j}{v_j^0} = \frac{E_j}{E_j^0} \cdot f\left(\frac{\mathbf{S}}{\mathbf{S}^0}, \boldsymbol{\varepsilon}_c^0\right), \tag{4.49}$$

where $\boldsymbol{\varepsilon}_c$ is the matrix of concentration elasticities as explained in Section 4.3. One example is the so-called lin-log kinetics

$$\frac{\mathbf{v}}{\mathbf{v}^0} = \frac{\mathbf{E}}{\mathbf{E}^0}\left(\mathbf{I} + \boldsymbol{\varepsilon}_c^0 \cdot \ln\frac{\mathbf{S}}{\mathbf{S}^0}\right), \tag{4.50}$$

where $\mathbf{I}$ is the $r \times r$ identity matrix. Another example is an approximation of the power-law kinetics

$$\ln\frac{\mathbf{v}}{\mathbf{v}^0} = \ln\frac{\mathbf{E}}{\mathbf{E}^0} + \boldsymbol{\varepsilon}_c^0 \ln\frac{\mathbf{S}}{\mathbf{S}^0}. \tag{4.51}$$

Approximative kinetics simplify the determination of model parameters and, especially, of concentration elasticities, since Eq. (4.51) as set of linear equations in the elasticity coefficients.

4.1.8 Convenience Kinetics and Modular Rate Laws

The convenience kinetics [17] has been introduced to ease parameter estimation and to have a kinetic mechanism, where all parameters are independent on each other and not related via the Haldane relation (Eq. (4.28)). It is a generalized form of Michaelis–Menten kinetics that covers all possible stoichiometries, and describes enzyme regulation by activators and inhibitors. For a reaction with stoichiometry

$$n_{-1}S_1 + n_{-2}S_2 + \cdots \leftrightarrow n_{+1}P_1 + n_{+2}P_2 + \cdots, \tag{4.52}$$

it reads

$$v = E_{\text{total}} \cdot f_{\text{reg}} \cdot \frac{k_{\text{cat}}^{\text{for}} \prod_i \left(S_i/K_{\text{m},S_i}\right)^{n_{-i}} - k_{\text{cat}}^{\text{back}} \prod_j \left(P_j/K_{\text{m},P_j}\right)^{n_{+j}}}{\prod_i \left(1 + \left(S_i/K_{\text{m},S_i}\right) + \cdots + \left(S_i/K_{\text{m},S_i}\right)^{n_{-i}}\right) + \prod_j \left(1 + \left(P_j/K_{\text{m},P_j}\right) + \cdots + \left(P_j/K_{\text{m},P_j}\right)^{n_{+j}}\right) - 1}, \tag{4.53}$$

with enzyme concentration E_{total} and turnover rates $k_{\text{cat}}^{\text{for}}$ and $k_{\text{cat}}^{\text{back}}$. The regulatory prefactor f_{reg} is either 1 (in case of no regulation) or a product of terms $M/(K_A + M)$ or $1 + M/K_A$ for activators and $K_I/(K_I + M)$ for inhibitors. Activation constants K_A and inhibition constants K_I are measured in concentration units. M is the concentration of the modifier.

In analogy to Michaelis–Menten kinetics, K_m values denote substrate concentrations, at which the reaction rate is half-maximal if the reaction products are absent; K_I and K_A values denote concentrations, at which the inhibitor or activator has its half-maximal effect. In this respect, many parameters in convenience kinetics are comparable to the kinetic constants measured in enzyme assays. This is important for parameter estimation (see Section 4.2).

To facilitate thermodynamic independence of the parameters, we introduce new system parameters that can be varied independently, without violating any thermodynamic constraints (see Section 4.1.1). For each reaction, we define the velocity constant $K_V = \left(k_{\text{cat}}^{\text{for}} \cdot k_{\text{cat}}^{\text{back}}\right)^{1/2}$ (geometric mean of the turnover rates in both directions). Given the equilibrium and velocity constants, the turnover rates can be written as $k_{\text{cat}}^{\text{for}} = K_V\left(K_{\text{eq}}\right)^{-1/2}$, $k_{\text{cat}}^{\text{back}} = K_V\left(K_{\text{eq}}\right)^{1/2}$. The equilibrium constants K_{eq} can be expressed by independent parameters such as the Gibbs free energies of formation: for each substance i, we define the dimensionless energy constant $K_i^G = \exp(G_i(0)/(RT))$ with Boltzmann's gas constant $R = 8.314\,\text{J}\,\text{mol}^{-1}\,\text{K}^{-1}$ and absolute temperature T. The equilibrium constants then satisfy $\ln K_{\text{eq}} = -N^T \ln K^G$.

In more general terms, modular rate laws are a family of reversible rate laws for reactions with arbitrary stoichiometries and various types of regulation, including mass-action, Michaelis–Menten, and uni–uni reversible Hill kinetics as special cases 20 385 728. There general form reads

$$v = E_{\text{total}} \cdot f_{\text{reg}} \cdot \frac{T}{D + D_{\text{reg}}}, \tag{4.54}$$

where f_{reg} describes complete or partial regulation (e.g., by an inhibitor), T is the numerator (equivalently to the one as used in equation (4.53)), while the components of the denominator, D and D_{reg}, depend on reaction stoichiometry, selected rate law, allosteric regulation, and on the preferred model parameterization. Five versions of denominator have been introduced:

1) Power-law modular rate law: $D = 1$ (such as mass action kinetics)
2) Common modular rate law: as in Eq. (4.53)
3) Simultaneous binding modular rate law:

$$D = \prod_i \left(1 + \frac{S_i}{K_{\text{m},S_i}}\right)^{n_{-i}} \prod_j \left(1 + \frac{P_j}{K_{\text{m},P_j}}\right)^{n_{+j}}$$

4) Direct binding modular rate law:

$$D = 1 + \prod_i \left(\frac{S_i}{K_{m,S_i}}\right)^{n_{-i}} + \prod_j \left(\frac{P_j}{K_{m,P_j}}\right)^{n_{+j}}$$

5) Force-dependent modular rate law:

$$D = \sqrt{\prod_i \left(\frac{S_i}{K_{m,S_i}}\right)^{n_{-i}} \prod_j \left(\frac{P_j}{K_{m,P_j}}\right)^{n_{+j}}}$$

With a thermodynamically safe parameterization of these rate laws, parameter sets obtained by model fitting, sampling, or optimization are guaranteed to lead to consistent chemical equilibrium states, as demonstrated above for convenience kinetics.

4.2 Metabolic Control Analysis

Summary

Metabolic control analysis (MCA) is a powerful quantitative and qualitative framework for studying the relationship between steady-state properties of a network of biochemical reaction and the properties of the individual reactions. It investigates the sensitivity of steady-state properties of the network to small parameter changes. MCA is a useful tool for theoretical and experimental analysis of control and regulation in cellular systems.

MCA was independently founded by two different groups in the 1970s [18,19] and was further developed by many different groups upon the application to different metabolic systems. A milestone in its formalization was provided in Ref. [20]. Originally intended for metabolic networks, MCA has nowadays found applications also for signaling pathways, gene expression models, and hierarchical networks [21–25].

Metabolic networks are very complex systems that are highly regulated and exhibit a lot of interactions like feedback inhibition or common substrates such as ATP for different reactions. Many mechanisms and regulatory properties of isolated enzymatic reactions are known. The development of MCA was motivated by a series of questions like the following: Can one predict properties or behavior of metabolic networks from the knowledge about their parts, the isolated reactions? Which individual steps control a flux or a steady-state concentration? Is there a rate-limiting step? Which effectors or modifications have the most prominent effect on the reaction rate? In biotechnological production processes, it is of interest which enzyme(s) should be activated in order to increase the rate of synthesis of a desired metabolite. There are also related problems in health care. Concerning metabolic disorders involving overproduction of a metabolite, which reactions should be modified in order to downregulate this metabolite while perturbing the rest of the metabolism as weakly as possible?

In metabolic networks, the steady-state variables, that is, the fluxes and the metabolite concentrations, depend on the value of parameters such as enzyme concentrations, kinetic constants (like Michaelis constants and maximal activities), and other model specific parameters. The effect of perturbations, moreover, depends on the place of the perturbation. As an illustration, in Example 4.2, we discuss a linear metabolic pathway whose enzymes are successively inhibited. We see in Figure 4.8 that an inhibition of the first enzyme has a different temporal effect than inhibition of the later enzymes. Also the steady states (here the values reached at time point 15) are different if different enzymes are hit.

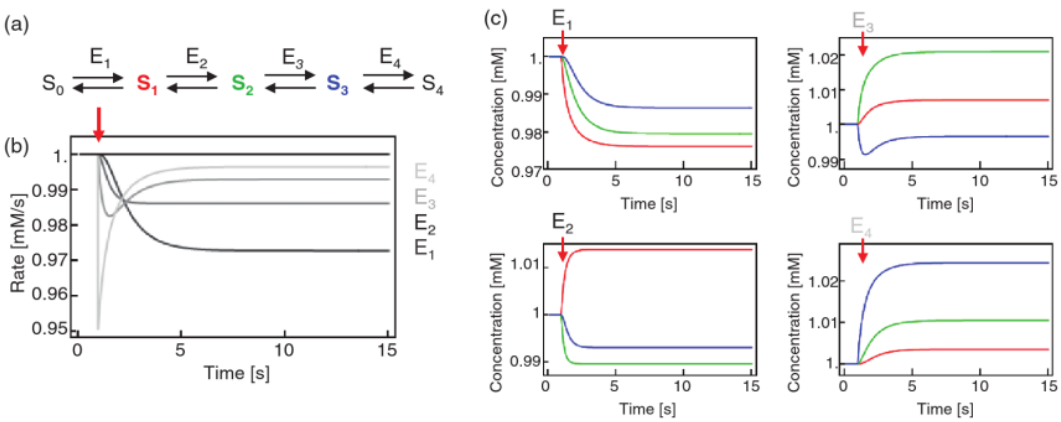
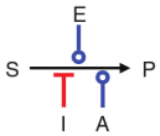


Figure 4.8 The effect of inhibiting an enzyme in an unbranched metabolic pathway depends on the position of that enzyme in the pathway. (a) Scheme of a linear metabolic pathway. Parameter values: see text. (b) Before perturbation, the system is at steady state. At time $t = 1$, one of the enzymes as indicated by gray scale is mildly inhibited by reducing its value by 5% (from 1 to 0.95). The rate of reaction 4 is presented. (c) Dynamics of metabolite concentrations upon different perturbations. Colors of time courses correspond to the colors of metabolites in (a).

Example 4.4

Typical values of elasticity coefficients will be explained for an isolated reaction transforming substrate S into product P. The reaction is catalyzed by enzyme E with the inhibitor I, and the activator A as depicted below



(4.59)

Usually, the elasticity coefficients for metabolite concentrations are in the following range:

$$\varepsilon_S^v = \frac{S}{v}\frac{\partial v}{\partial S} > 0 \quad \text{and} \quad \varepsilon_P^v = \frac{P}{v}\frac{\partial v}{\partial P} \leq 0. \tag{4.60}$$

In most cases, the rate increases with the concentration of the substrate (compare, e.g., Eq. (4.58)) and decreases with the concentration of the product. An exception from $\varepsilon_S^v > 0$ occurs in the case of substrate inhibition (Eq. (4.33)), where the elasticity will become negative for $S > S_{\text{opt}}$. The relation $\varepsilon_P^v = 0$ holds, if the reaction is irreversible or if the product concentration is kept zero by external mechanisms. The elasticity coefficients with respect to effectors I or A should obey

$$\varepsilon_A^v = \frac{A}{v}\frac{\partial v}{\partial A} > 0 \quad \text{and} \quad \varepsilon_I^v = \frac{I}{v}\frac{\partial v}{\partial I} < 0, \tag{4.61}$$

since this is essentially what the notions activator and inhibitor mean.

For the most kinetic laws, the reaction rate v is proportional to the enzyme concentration E. For example, E is a multiplicative factor in the mass action rate law as well as in the maximal rate of the Michelis–Menten rate law. Therefore, it holds that

$$\varepsilon_E^v = \frac{\partial \ln v}{\partial \ln E} = 1. \tag{4.62}$$

More complicated interactions between enzymes and substrates like metabolic channeling (direct transfer of the metabolite from one enzyme to the next without release to the medium) may lead to exceptions from this rule.

Such a parameter might be the enzyme concentration, a kinetic constant, or the concentration of a specific inhibitor or effector.

In a more compact form the flux control coefficient reads

$$C_k^j = \frac{v_k}{J_j}\frac{\partial J_j}{\partial v_k}. \tag{4.65}$$

The respective nonnormalized flux control coefficient is $\tilde{C}_k^j = \partial J_j/\partial v_k$. Equivalently, the *concentration control coefficient* of concentrations S_i^{ss} with respect to v_k reads

$$C_k^i = \frac{v_k}{S_i^{ss}}\frac{\partial S_i^{ss}}{\partial v_k}. \tag{4.66}$$

4.2.1.3 Response Coefficients

The steady state is determined by the values of the parameters. A third type of coefficients expresses the direct dependence of steady-state variables on parameters. The response coefficients are defined as

$$R_m^j = \frac{p_m}{J_j}\frac{\partial J_j}{\partial p_m} \quad \text{and} \quad R_m^i = \frac{p_m}{S_i^{ss}}\frac{\partial S_i^{ss}}{\partial p_m}, \tag{4.67}$$

where the first coefficient expresses the response of the flux to a parameter perturbation, while the latter describes the response of a steady-state concentration.

4.2.1.4 Matrix Representation of the Coefficients

Control, response, and elasticity coefficients are defined with respect to all rates, steady-state concentrations, fluxes, or parameters in the metabolic system and in the respective model. They can be arranged in matrices:

$$\mathbf{C}^J = \left\{C_k^j\right\}, \mathbf{C}^S = \left\{C_k^i\right\}, \mathbf{R}^J = \left\{R_m^j\right\}, \mathbf{R}^S = \left\{R_m^i\right\},$$
$$\boldsymbol{\varepsilon} = \left\{\varepsilon_i^k\right\}, \boldsymbol{\pi} = \left\{\pi_m^k\right\}. \tag{4.68}$$

Matrix representation can also be chosen for all types of nonnormalized coefficients. The arrangement in matrices allows to applying matrix algebra in control analysis. In particular, the matrices of normalized control coefficients can be calculated from the matrices of nonnormalized control coefficient as follows:

$$\begin{aligned} \mathbf{C}^J &= (\mathrm{dg}\boldsymbol{J})^{-1}\cdot\tilde{\mathbf{C}}^J\cdot\mathrm{dg}\mathbf{J} & \mathbf{C}^S &= (\mathrm{dg}\mathbf{S}^{\mathrm{ss}})^{-1}\cdot\tilde{\mathbf{C}}^J\cdot\mathrm{dg}\mathbf{J} \\ \mathbf{R}^J &= (\mathrm{dg}\mathbf{J})^{-1}\cdot\tilde{\mathbf{R}}^J\cdot\mathrm{dg}\mathbf{p} & \mathbf{R}^S &= (\mathrm{dg}\mathbf{S}^{\mathrm{ss}})^{-1}\cdot\tilde{\mathbf{R}}^S\cdot\mathrm{dg}\mathbf{p} \\ \boldsymbol{\varepsilon} &= (\mathrm{dg}\mathbf{v})^{-1}\cdot\tilde{\boldsymbol{\varepsilon}}\cdot\mathrm{dg}\mathbf{S}^{\mathrm{ss}} & \boldsymbol{\pi} &= (\mathrm{dg}\mathbf{v})^{-1}\cdot\tilde{\boldsymbol{\pi}}\cdot\mathrm{dg}\mathbf{p} \end{aligned} \tag{4.69}$$

The symbol "dg" stands for the diagonal matrix, that is, for a system with three reactions it holds $\mathrm{dg}\mathbf{J} = \begin{pmatrix} J_1 & 0 & 0 \\ 0 & J_2 & 0 \\ 0 & 0 & J_3 \end{pmatrix}$.

4.2.2 The Theorems of Metabolic Control Theory

Let us assume that we are interested in calculating the control coefficients for a system under investigation. Usually, the steady-state fluxes or concentrations cannot be expressed explicitly as function of the reaction rates. Therefore, flux and concentration control coefficients

cannot simply be determined by taking the respective derivatives, as we did for the elasticity coefficients in Example 4.3.

Fortunately, the work with control coefficients is eased by of a set of theorems. The first type of theorems, the *summation theorems*, makes a statement about the total control over a flux or a steady-state concentration. The second type of theorems, the *connectivity theorems*, relates the control coefficients to the elasticity coefficients. Both types of theorems together with network information encoded in the stoichiometric matrix contain enough information to calculate all control coefficients.

Here, we will first introduce the theorems. Then, we will present a hypothetical perturbation experiment (as introduced by Kacser & Burns) to illustrate the summation theorem. Finally, the theorems will be derived mathematically.

4.2.2.1 The Summation Theorems

The summation theorems make a statement about the total control over a certain steady-state flux or concentration. The flux control coefficients and concentration control coefficients fulfill, respectively,

$$\sum_{k=1}^{r} C_{v_k}^{J_j} = 1 \quad \text{and} \quad \sum_{k=1}^{r} C_{v_k}^{S_i} = 0, \tag{4.70}$$

for any flux J_i and any steady-state concentration S_i^{st}. The quantity r is the number of reactions. The flux control coefficients of a metabolic network for one steady-state flux sum up to one. This means that all enzymatic reactions can share the control over this flux. The control coefficients of a metabolic network for one steady-state concentration are balanced. This means again that the enzymatic reactions can share the control over this concentration, but some of them exert a negative control while others exert a positive control. Both relations can also be expressed in matrix formulation. We get

$$\mathbf{C}^J \cdot \mathbf{1} = \mathbf{1} \quad \text{and} \quad \mathbf{C}^S \cdot \mathbf{1} = \mathbf{0}. \tag{4.71}$$

The symbols $\mathbf{1}$ and $\mathbf{0}$ denote column vectors with r rows containing as entries only ones or zeros, respectively. The summation theorems for the nonnormalized control coefficients read

$$\tilde{\mathbf{C}}^J \cdot \mathbf{K} = \mathbf{K} \quad \text{and} \quad \tilde{\mathbf{C}}^S \cdot \mathbf{K} = \mathbf{0}, \tag{4.72}$$

where $\mathbf{K}$ is the matrix satisfying $\mathbf{N} \cdot \mathbf{K} = \mathbf{0}$ (see Section 4.2). A more intuitive derivation of the summation theorems is given in the following example according to Kacser and Burns [18].

Example 4.5

The summation theorem for flux control coefficients can be derived using a thought experiment.

Consider the following unbranched pathway with fixed concentrations of the external metabolites, S_0 and S_3:

$$S_0 \overset{v_1}{\longleftrightarrow} S_1 \overset{v_2}{\longleftrightarrow} S_2 \overset{v_3}{\longleftrightarrow} S_3. \tag{4.73}$$

What happens to steady-state fluxes and metabolite concentrations, if we perform an experimental manipulation of all three reactions leading to the same fractional change α of all three rates?

$$\frac{\delta v_1}{v_1} = \frac{\delta v_2}{v_2} = \frac{\delta v_3}{v_3} = \alpha. \tag{4.74}$$

The flux must increase to the same extent, $\delta J/J = \alpha$, but, since rates of producing and degrading reactions increase to the same amount, the concentrations of the metabolites remain constant $\delta S_1/S_1 = \delta S_2/S_2 = 0$.

The combined effect of all changes in local rates on the system variables S_1^{ss}, S_2^{ss}, and J can be written as the sum of all individual effects caused by the local rate changes. For the flux holds

$$\frac{\delta J}{J} = C_1^J \frac{\delta v_1}{v_1} + C_2^J \frac{\delta v_2}{v_2} + C_3^J \frac{\delta v_3}{v_3}. \tag{4.75}$$

It follows

$$\alpha = \alpha\left(C_1^J + C_2^J + C_3^J\right) \quad \text{or} \quad 1 = C_1^J + C_2^J + C_3^J. \tag{4.76}$$

This is just a special case of Eq. (4.70). In the same way, for the change of concentration S_1^{ss}, we obtain

$$\frac{\delta S_1^{\text{ss}}}{S_1^{\text{ss}}} = C_1^{S_1} \frac{\delta v_1}{v_1} + C_2^{S_1} \frac{\delta v_2}{v_2} + C_3^{S_1} \frac{\delta v_3}{v_3}. \tag{4.77}$$

Finally, we get

$$\begin{aligned} 0 &= C_1^{S_1} + C_2^{S_1} + C_3^{S_1} \quad \text{as well as} \\ 0 &= C_1^{S_2} + C_2^{S_2} + C_3^{S_2}. \end{aligned} \tag{4.78}$$

Although shown here only for a special case, these properties hold in general for systems without conservation relations. The general derivation is given in Section 4.2.3.

4.2.2.2 The Connectivity Theorems

Flux control coefficients and elasticity coefficients are related by the expression

$$\sum_{k=1}^{r} C_{v_k}^{J_j} \varepsilon_{S_i}^{v_k} = 0. \tag{4.79}$$

Note that the sum runs over all rates v_k for any flux J_j. Considering the concentration S_i of a specific metabolite

and a certain flux J_j, each term contains the elasticity $\varepsilon_{S_i}^{v_k}$ describing the direct influence of a change of S_i on the rates v_k and the control coefficient expressing the control of v_k over J_j.

The connectivity theorem between concentration control coefficients and elasticity coefficients reads

$$\sum_{k=1}^{r} C_{v_k}^{S_h} \varepsilon_{S_i}^{v_k} = -\delta_{hi}. \tag{4.80}$$

Again, the sum runs over all rates v_k, while S_h and S_i are the concentrations of two fixed metabolites. The symbol $\delta_{hi} = \begin{cases} 0, & \text{if } h \neq i \\ 1, & \text{if } h = i \end{cases}$ is the so-called Kronecker symbol.

In matrix formulation, the connectivity theorems read

$$\mathbf{C}^J \cdot \boldsymbol{\varepsilon} = \mathbf{0} \quad \text{and} \quad \mathbf{C}^S \cdot \boldsymbol{\varepsilon} = -\mathbf{I}, \tag{4.81}$$

where $\boldsymbol{I}$ denotes the identity matrix of size $n \times n$. For nonnormalized coefficients, it holds

$$\tilde{\mathbf{C}}^J \cdot \tilde{\boldsymbol{\varepsilon}} \cdot \mathbf{L} = \mathbf{0} \quad \text{and} \quad \tilde{\mathbf{C}}^S \cdot \tilde{\boldsymbol{\varepsilon}} \cdot \mathbf{L} = -\mathbf{L}, \tag{4.82}$$

Example 4.6

To calculate the control coefficients, we study the following reaction system:

$$P_0 \overset{V_1}{\longleftrightarrow} S \overset{V_2}{\longleftrightarrow} P_2 \tag{4.84}$$

The flux control coefficients obey the theorems

$$C_1^J + C_2^J = 1 \quad \text{and} \quad C_1^J \varepsilon_S^1 + C_2^J \varepsilon_S^2 = 0, \tag{4.85}$$

which can be solved for the control coefficients to yield

$$C_1^J = \frac{\varepsilon_S^2}{\varepsilon_S^2 - \varepsilon_S^1} \quad \text{and} \quad C_2^J = \frac{-\varepsilon_S^1}{\varepsilon_S^2 - \varepsilon_S^1}. \tag{4.86}$$

Since usually $\varepsilon_S^1 < 0$ and $\varepsilon_S^2 > 0$ (see Example 4.4), both control coefficients assume positive values $C_1^J > 0$ and $C_2^J > 0$. This means, that both reactions exert a positive control over the steady-state flux, and acceleration of any of them leads to increase of J, which is in accordance with common intuition.

The concentration control coefficients fulfill

$$C_1^S + C_2^S = 0 \quad \text{and} \quad C_1^S \varepsilon_S^1 + C_2^S \varepsilon_S^2 = -1, \tag{4.87}$$

which yields

$$C_1^S = \frac{1}{\varepsilon_S^2 - \varepsilon_S^1} \quad \text{and} \quad C_2^S = \frac{-1}{\varepsilon_S^2 - \varepsilon_S^1}. \tag{4.88}$$

With $\varepsilon_S^1 < 0$ and $\varepsilon_S^2 > 0$, we get $C_1^S > 0$ and $C_2^S < 0$, that is, increase of the first reaction causes a raise in the steady-state concentration of S while acceleration of the second reaction leads to the opposite effect.

where $\mathbf{L}$ is the link matrix that expresses the relation between independent and dependent rows in the stoichiometric matrix (Section 3.15, Eq. (3.22)). A comprehensive representation of both summation and connectivity theorems for nonnormalized coefficients is given by the following equation:

$$\begin{pmatrix} \tilde{\mathbf{C}}^{\mathbf{J}} \\ \tilde{\mathbf{C}}^{\mathbf{S}} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{K} & \tilde{\boldsymbol{\varepsilon}}\,\mathbf{L} \end{pmatrix} = \begin{pmatrix} \mathbf{K} & \mathbf{0} \\ \mathbf{0} & -\mathbf{L} \end{pmatrix}. \tag{4.83}$$

The summation and connectivity theorem together with the structural information of the stoichiometric matrix are sufficient to calculate the control coefficients for a metabolic network. This shall be illustrated for a small network in the next example.

4.2.3 Matrix Expressions for Control Coefficients

After having introduced the theorems of MCA, we will derive expressions for the control coefficients in matrix form. These expressions are suited for calculating the coefficients even for large-scale models. We start from the steady-state condition

$$\mathbf{N}\mathbf{v}(\boldsymbol{S}^{\mathrm{ss}}(\mathbf{p}), \mathbf{p}) = \mathbf{0}. \tag{4.89}$$

Implicit differentiation with respect to the parameter vector $\mathbf{p}$ yields

$$\mathbf{N}\frac{\partial \mathbf{v}}{\partial \mathbf{S}}\frac{\partial \mathbf{S}^{\mathrm{ss}}}{\partial \mathbf{p}} + \mathbf{N}\frac{\partial \mathbf{v}}{\partial \mathbf{p}} = \mathbf{0}. \tag{4.90}$$

If we chose reaction specific parameters for perturbation, the matrix of nonnormalized parameter elasticities contains nonzero entries in the main diagonal and zeros elsewhere (compare Eq. (4.64)).

$$\frac{\partial \mathbf{v}}{\partial \mathbf{p}} = \begin{pmatrix} \frac{\partial v_1}{\partial p_1} & 0 & & 0 \\ 0 & \frac{\partial v_2}{\partial p_2} & \cdots & 0 \\ & & & \\ 0 & 0 & & \frac{\partial v_r}{\partial p_r} \end{pmatrix}. \tag{4.91}$$

Therefore, this matrix is regular and has an inverse. Furthermore, we consider the Jacobian matrix

$$\mathbf{M} = \mathbf{N}\frac{\partial \mathbf{v}}{\partial \mathbf{S}} = \mathbf{N}\tilde{\boldsymbol{\varepsilon}}. \tag{4.92}$$

The Jacobian $\mathbf{M}$ is a regular matrix if the system is asymptotically stable and contains no conservation

relations. The case with conservation relations is considered below. Here, we may premultiply Eq. (4.90) by the inverse of **M** and rearrange to get

$$\frac{\partial \mathbf{S}^{ss}}{\partial \mathbf{p}} = -\left(\mathbf{N}\frac{\partial \mathbf{v}}{\partial \mathbf{S}}\right)^{-1}\mathbf{N}\frac{\partial \mathbf{v}}{\partial \mathbf{p}} = -\mathbf{M}^{-1}\mathbf{N}\frac{\partial \mathbf{v}}{\partial \mathbf{p}} \equiv \tilde{\mathbf{R}}^{S}. \quad (4.93)$$

As indicated, $\partial \mathbf{S}^{ss}/\partial \mathbf{p}$ is the matrix of nonnormalized response coefficients for concentrations. Postmultiplication by the inverse of the nonnormalized parameter elasticity matrix gives

$$\frac{\partial \mathbf{S}^{ss}}{\partial \mathbf{p}}\left(\frac{\partial \mathbf{v}}{\partial \mathbf{p}}\right)^{-1} = -\left(\mathbf{N}\frac{\partial \mathbf{v}}{\partial \mathbf{S}}\right)^{-1}\mathbf{N} = \tilde{\mathbf{C}}^{S}. \quad (4.94)$$

This is the matrix of nonnormalized concentration control coefficients. The right (middle) site contains no parameters. This means, that the control coefficients do not depend on the particular choice of parameters to exert the perturbation as long as Eq. (4.64) is fulfilled. The control coefficients are only dependent on the structure of the network represented by the stoichiometric matrix **N**, and on the kinetics of the individual reactions, represented by the nonnormalized elasticity matrix $\tilde{\boldsymbol{\varepsilon}} = \partial \mathbf{v}/\partial \mathbf{S}$.

The implicit differentiation of

$$\mathbf{J} = \mathbf{v}(\mathbf{S}^{ss}(\mathbf{p}), \mathbf{p}), \quad (4.95)$$

with respect to the parameter vector **p** leads to

$$\frac{\partial \mathbf{J}}{\partial \mathbf{p}} = \frac{\partial \mathbf{v}}{\partial \mathbf{p}} + \frac{\partial \mathbf{v}}{\partial \mathbf{S}}\frac{\partial \mathbf{S}^{ss}}{\partial \mathbf{p}} = \left(\mathbf{I} - \frac{\partial \mathbf{v}}{\partial \mathbf{S}}\left(\mathbf{N}\frac{\partial \mathbf{v}}{\partial \mathbf{S}}\right)^{-1}\mathbf{N}\right)\frac{\partial \mathbf{v}}{\partial \mathbf{p}} \equiv \tilde{R}^{J}. \quad (4.96)$$

This yields, after some rearrangement, an expression for the nonnormalized flux control coefficients:

$$\frac{\partial \mathbf{J}}{\partial \mathbf{p}}\left(\frac{\partial \mathbf{v}}{\partial \mathbf{p}}\right)^{-1} = \mathbf{I} - \frac{\partial \mathbf{v}}{\partial \mathbf{S}}\left(\mathbf{N}\frac{\partial \mathbf{v}}{\partial \mathbf{S}}\right)^{-1}\mathbf{N} = \tilde{\mathbf{C}}^{J}. \quad (4.97)$$

The normalized control coefficients are (by use of Eq. (4.69))

$$\mathbf{C}^{J} = \mathbf{I} - (dg\mathbf{J})^{-1}\left(\frac{\partial \mathbf{v}}{\partial \mathbf{S}}\left(\mathbf{N}\frac{\partial \mathbf{v}}{\partial \mathbf{S}}\right)^{-1}\mathbf{N}\right)(dg\mathbf{J}) \quad \text{and}$$

$$\mathbf{C}^{S} = -(dg\mathbf{S}^{ss})^{-1}\left(\left(\mathbf{N}\frac{\partial \mathbf{v}}{\partial \mathbf{S}}\right)^{-1}\mathbf{N}\right)(dg\mathbf{J}). \quad (4.98)$$

These equations can easily be implemented for numerical calculation of control coefficients or used for analytical computation.

They are also suited for derivation of the theorems of MCA. The summation theorems for the control coefficients follow from Eq. (4.98) by postmultiplication with the vector **1** (the row vector containing only 1 s), and consideration of the relations $(dg\mathbf{J}) \cdot \mathbf{1} = \mathbf{J}$ and $\mathbf{NJ} = \mathbf{0}$, as shown below:

$$\begin{aligned}\mathbf{C}^{J}\mathbf{1} &= \mathbf{I}\cdot\mathbf{1} - (dg\mathbf{J})^{-1}\left(\frac{\partial \mathbf{v}}{\partial \mathbf{S}}\left(\mathbf{N}\frac{\partial \mathbf{v}}{\partial \mathbf{S}}\right)^{-1}\mathbf{N}\right)(dg\mathbf{J})\mathbf{1} \\ \mathbf{C}^{J}\mathbf{1} &= \mathbf{1} - (dg\mathbf{J})^{-1}\left(\frac{\partial \mathbf{v}}{\partial \mathbf{S}}\left(\mathbf{N}\frac{\partial \mathbf{v}}{\partial \mathbf{S}}\right)^{-1}\mathbf{N}\right)\mathbf{J} = \mathbf{1} - \mathbf{0}\end{aligned}. \quad (4.99)$$

The connectivity theorems result from postmultiplication of Eq. (4.98) with the elasticity matrix $\boldsymbol{\varepsilon} = (dg\mathbf{J})^{-1} \cdot (\partial \mathbf{v}/\partial \mathbf{S}) \cdot dg\mathbf{S}^{ss}$, and using that multiplication of a matrix with its inverse yields the identity matrix **I** of respective type.

If the reaction system involves conservation relations, we eliminate dependent variables as explained in Section 1.2.4. In this case, the nonnormalized coefficients read

$$\tilde{\mathbf{C}}^{J} = \mathbf{I} - \frac{\partial \mathbf{v}}{\partial \mathbf{S}}\mathbf{L}\left(\mathbf{N}_{\mathbf{R}}\frac{\partial \mathbf{v}}{\partial \mathbf{S}}\right)^{-1}\mathbf{N}_{\text{indep}} \quad \text{and}$$

$$\tilde{\mathbf{C}}^{S} = -\mathbf{L}\left(\mathbf{N}_{\mathbf{R}}\frac{\partial \mathbf{v}}{\partial \mathbf{S}}\right)^{-1}\mathbf{N}_{\text{indep}}, \quad (4.99)$$

and the normalized control coefficients are obtained by applying Eq. (4.69).

An example for calculation of flux control coefficients can be found in the web material.

To investigate to implications of control distribution, we will now analyze the control pattern in an unbranched pathway:

$$S_0 \overset{V_1}{\longleftrightarrow} S_1 \overset{V_2}{\longleftrightarrow} S_2 \ldots S_{r-1} \overset{V_r}{\longleftrightarrow} S_r \quad (4.100)$$

with linear kinetics $v_i = k_i S_{i-1} - k_{-i} S_i$, the equilibrium constants $q_i = k_i/k_{-i}$, and fixed concentrations of the external metabolites, S_0 and S_r. In this case, one can calculate an analytical expression for the steady-state flux,

$$J = \frac{S_0 \prod_{j=1}^{r} q_j - S_r}{\sum_{l=1}^{r} \frac{1}{k_l} \prod_{m=l}^{r} q_m}, \quad (4.101)$$

as well as an analytical expression for the flux control coefficients

$$C_i^{J} = \left(\frac{1}{k_i}\prod_{j=i}^{r} q_j\right) \cdot \left(\sum_{l=1}^{r} \frac{1}{k_l} \prod_{m=l}^{r} q_m\right)^{-1}. \quad (4.102)$$

Let us consider two very general cases. First assume that all reactions have the same individual kinetics, $k_i = k_+, k_{-i} = k_-$ for $i = 1, \ldots, r$ and that the equilibrium constants, which are also equal, satisfy $q = k_+/k_- > 1$. In this case, the ratio of two subsequent flux control coefficients

is

$$\frac{C_i^J}{C_{i+1}^J} = \frac{k_{i+1}}{k_i} q_i = q > 1. \tag{4.103}$$

Hence, the control coefficients of the preceding reactions are larger than the control coefficients of the succeeding reactions and flux control coefficients are higher in the beginning of a chain than in the end. This is in agreement with the frequent observation that flux control is strongest in the upper part of an unbranched reaction pathway.

Now assume that the individual rate constants might be different, but that all equilibrium constants are equal to one, $q_i = 1$ for $i = 1, \ldots, r$. This implies $k_i = k_{-i}$. Equation (4.102) simplifies to

$$C_i^J = \frac{1}{k_i} \cdot \left(\sum_{l=1}^{r} \frac{1}{k_l} \right)^{-1}. \tag{4.104}$$

Consider now the relaxation time $\tau_i = 1/(k_i + k_{-i})$ (see Section 4.3) as a measure for the rate of an enzyme. The flux control coefficient reads

$$C_i^J = \frac{\tau_i}{\tau_1 + \tau_2 + \cdots + \tau_r}. \tag{4.105}$$

This expression helps to elucidate two aspects of metabolic control. First, all enzymes participate in the control since all enzymes have a positive relaxation time. There is no enzyme that has all control; that is, determines the flux through the pathway alone. Second, slow enzymes with a higher relaxation time exert in general more control than fast enzymes with a short relaxation time.

The predictive power of flux control coefficients for directed changes of flux is illustrated in the following example.

Example 4.7

Assume that we can manipulate the pathway shown in Figure 4.12 by changing the enzyme concentration in a predefined way. We would like to explore the effect of the perturbation of the individual enzymes. For a linear pathway (see Eqs. (4.100)–(4.102)) consisting of four consecutive reactions, we calculate the flux control coefficients. For $i = 1, \ldots, 4$, it shall hold that (i) all enzyme concentrations are $E_i = 1$, (ii) the rate constants are $k_i = 2, k_{-i} = 1$, and (iii) the concentrations of the external reactants are $S_0 = S_4 = 1$. The resulting flux is $J = 1$ and the flux control coefficients are $\mathbf{C}^J = \begin{pmatrix} 0.533 & 0.267 & 0.133 & 0.067 \end{pmatrix}^T$ according to Eq. (4.98).

If we now perturb slightly the first enzyme, lets say perform a percentage change of its concentration, that is, $E_1 \to E_1 + 1\%$, then Eq. (4.54) implies that the flux increases as $J \to J + C_1^J \cdot 1\%$. In fact, the flux in the new steady state is $J^{E_1 \to 1.01 \cdot E_1} = 1.00531$. Increasing E_2, E_3, or E_4 by 1% leads to flux values of 1.00265, 1.00132, and 1.00066, respectively. A strong perturbation would not yield similar effects. This is illustrated in Figure 4.10.

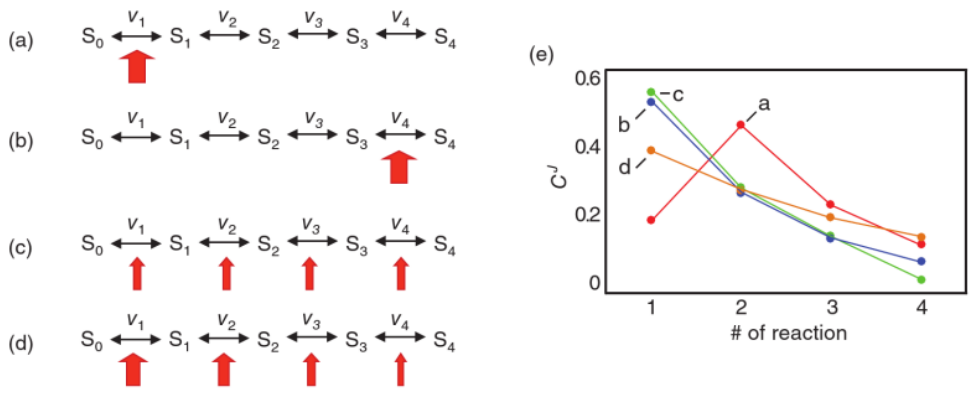


Figure 4.10 Effect of enzyme concentration change on steady-state flux and on flux control coefficients in an unbranched pathway consisting of four reactions. In the reference state, all enzymes have the concentration 1 (in arbitrary units), the control distribution is the same as in case (C), and the steady-state flux is $J = 1$. (a) Change of $E_1 \to 5E_1$ while keeping the other enzyme concentrations constant results in a remarkable drop of control of the first enzyme. The resulting flux is $J^{E_1 \to 5 \cdot E_1} = 1.7741$. (b) The change $E_4 \to 5E_4$ corresponds to $J^{E_4 \to 5 \cdot E_4} = 1.0563$. There is only slight change of control distribution. (c) Equal enzyme concentrations with $E_i \to 2E_i, i = 1, \ldots, 4$ results in $J^{E_i \to 2 \cdot E_i} = 2$. (d) Optimal distribution of enzyme concentration $E_1 = 3.124, E_2 = 2.209, E_3 = 1.562, E_4 = 1.105$ resulting in the maximal steady state flux $J^{\max} = 2.2871$.

Example 4.8

Control analysis can be applied to both metabolic and signaling networks. The small network shown in Figure 4.12a could represent both cases. It shows the production (and degradation) of compound S_1 that modifies the conversion of S_4 into S_2. Such a reaction cycle between two components where both reactions are catalyzed by different enzymes occurs both in metabolic networks (e.g., the conversion of fructose-6-phosphate to fructose-1,6-bisphosphate by phosphofructokinase, PFK, and the reverse reaction catalyzed by fructose bisphosphatase, an important part of glycolysis, see Section 12.1) and in signaling pathways (e.g., the activation of a small G-protein by exchange of GDP with GTP catalyzed by a guanine-nucleotide exchange factor, GEF, and reversely the hydrolysis of GTP to GDP catalyzed by GTPase activating proteins, GAPs, see Section 12.2). In our case, S_2 catalyzes the formation of the next compound S_3, which is in turn degraded. The stoichiometric matrix for this network reads

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

Application of stoichiometric analysis (Chapter 3) reveals a conservation relation, that is, $S_2 + S_4 = \text{constant}$, for all times (Section 3.1.5) and three independent fluxes in steady state, that is, $\mathbf{k}_1 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}^T$, $\mathbf{k}_2 = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 \end{pmatrix}^T$, and $\mathbf{k}_3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}^T$. These independent fluxes are illustrated in Figure 4.12b by different colors. The time-dependent response depends on the initial conditions for the system, that is, whether the system is in steady state when a parameter is perturbed (Figure 4.12c) at a state far away from equilibrium (Figure 4.12d). Looking at the case where we start at steady state, middle panel, we see that a perturbing $S_3(0)$ has initially the largest effect on $S_3(t)$ as indicated by a response coefficient of 1. But this effect declines over time since systems dynamics would lead the system back to its original steady state. Increasing $S_4(0)$ would, however, increase in the long run due to an increase of the conserved moiety of S_2 and S_4. The impact of parameter values is zero at time point $t = 0$, but then increases for producing reactions such as k_1 and k_5 and decreases for degrading reactions such as k_2 and k_6. When looking at the response of S4 (lower panel), we find that increasing $S_3(0)$ has only temporarily a diminishing effect on $S_4(t)$, but not with respect to the new steady state. Parameters k_5 and k_6 have no effect on $S_4(t)$.

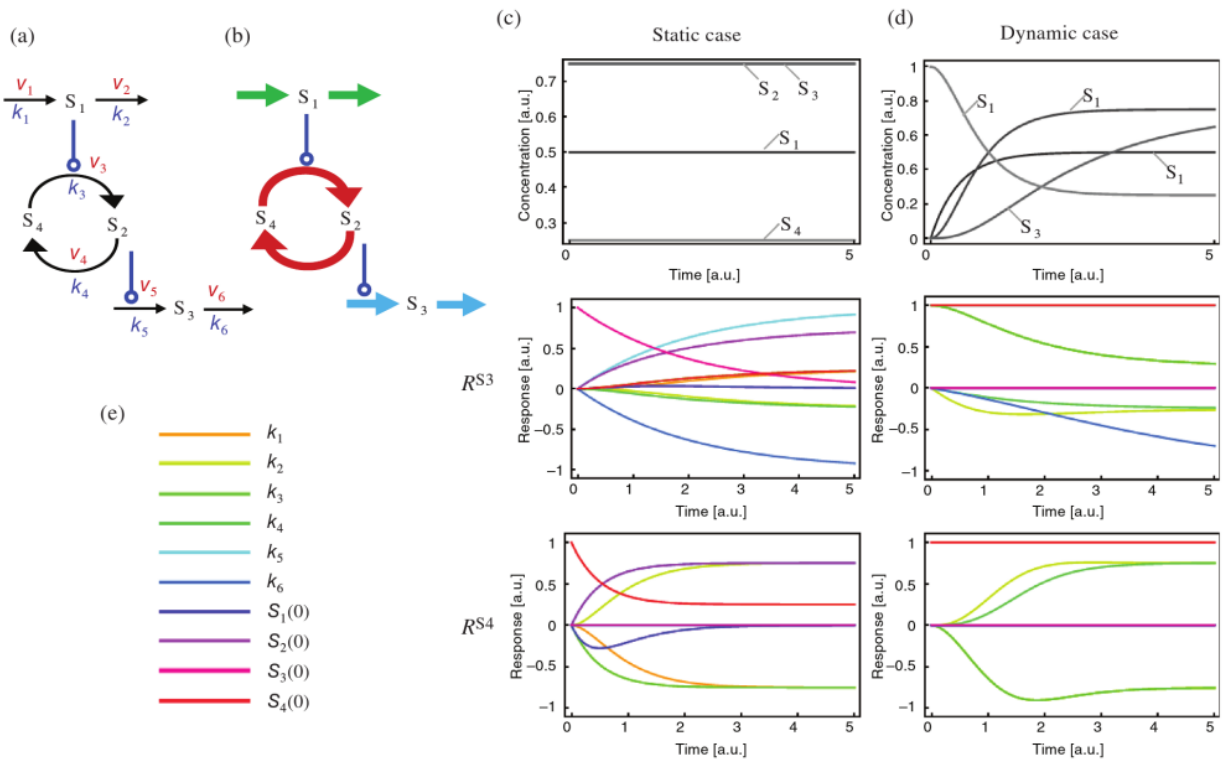


Figure 4.12 Illustration of time-dependent flux response. (a) Example network, (b) Independent steady-state fluxes for the production and degradation of S_1, for the interconversion of S_2 and S_4, and for the production and degradation of S3 are shown by green, red, and blue arrows, respectively. (c) Response analysis for the network at steady state: Top panel: constant concentrations of the substrates, Middle panel: temporal behavior of all response coefficients for S_3, and Lower panel: temporal behavior of all response coefficients for S_4. (d) Response analysis for the system starting at the initial conditions $S_1(0) = S_2(0) = S_3(0) = 0; S_1(0) = 1$. Panels as in C. (e) Color code for the parameters of the system as used in the middle and lower panels.

Parameter values: $k_1 = 1; k_2 = 2; k_3 = 3; k_4 = k_5 = k_6 = 0.5$.