SYSTEMS BIOLOGY

Dynamic Pathway Modelling

Olaf Wolkenhauer

www.sbi.uni-rostock.de

8th October 2012

This is a draft manuscript

While in the past it was more common to make preprints, algorithms and software freely available, authors have become more protective of their work. I do not know whether this is a reflection of the increased competition for recognition or the pressure to commercialise research results. By making these notes available I hope to receive information about errors, misconceptions, missing references or acknowledgment, and suggestions to improve the text. In the modern world of publishing one also feels a need to clarify that the author makes no warranty of any kind, expressed or implied, with regard to programs, algorithms, theoretical concepts, or the documentation contained in this manuscript. The author shall not be liable in any event for incidental or consequential damages in connection with, or arising out of, the use of the material provided here.

Contents

Preface

A motivation not only for this text but for many if not most research projects in molecular biology, cell biology and biomedicine is an improved understanding of disease mechanisms. We like to understand the causal basis of a disease that effects an entire organism. Taking a human body, it consists of organs, which in turn are build from cells. A biological cell is thus the basic building block for living systems and the focus of attention for an understanding of not only the cell but higher levels of structural and functional organisation. Whether we are interested in cancer, neurodegenerative diseases like Parkinson's or Alzheimer's, metabolic diseases like diabetes or the malfunction of the immune system, modern life sciences tries to understand phenomena at the physiological level of an organism at the level of the cell. For example, a cancerous tumor is an uncontrolled proliferation of cells and it seems therefore logical to try and understand how a cell *functions*. The functions of a cell are its growth, differentiation (specialisation), proliferation (division) and apoptosis (programmed cell death). In all of this the cell is interacting or communicating with its environment and other cells. Many drugs are designed to influence this communication, to suppress or stimulate the cell's behavior in a defined way. The generation of signals and transfer of information is achieved through biochemical reactions among molecules. The concept of a *pathway* is used to identify groups of molecules that interact in a specific way so as to realise the functions of cell. Intra- and intercellular interactions and cell functions are nonlinear dynamic processes. To understand how the cell functions requires therefore a theory of dynamic systems. We are going to present formal mathematical concepts that are the basis for an understanding of nonlinear, dynamic molecular interactions in cells.

The focus of this book is on systems biology, an emerging area of research that is a natural conclusion from the advances made in related areas, including genomics, molecular biology, cell biology, biomedicine and bioinformatics. The areas of genomics and bioinformatics have identified and characterised many of the components that make up a living cell and maintain its function. In genomics the genetic information that is encoded in the genome is studied with respect to genes and the proteins they code for. A primary aim of bioinformatics has been to link genome sequences or genes with RNA products and proteins, i.e., to determine whether in a particular experimental context there exist a relationship between genes and proteins, amongst genes and proteins, and across genomes. The principal objective of modern life sciences is to describe the role of these components in developmental and disease mechanisms. While the developments in genomics and bioinformatics have brought tremendous advances in our understanding of molecular and cell biology, it is increasingly recognised that it is the temporal interaction amongst large numbers of molecules that determine phenomena observed at higher (metabolic, cellular, or physiological) levels. This dynamic or systems perspective

and integrative approach (combining data from the genome, transcriptome, proteome, metabolome, ...) is considered in the area of research referred to as *Systems Biology*:

> Systems biology investigates the functioning and function of inter- and intra-cellular dynamic networks, using signal- and systems-oriented approaches.

To understand the functioning and function of cells, systems biology addresses the following central questions:

> *How do the components within a cell interact to bring about its structure and function? (intra-cellular dynamics)*
>
> *How do cells interact to bring about higher levels of structural and functional organisation? (inter-cellular dynamics)*

The functions of a cell do not reside in the molecules themselves but in their interactions, just as life is an emergent, rather than an immanent or inherent, property of matter. Although life, or the function of the cell arise from the material world, they cannot be reduced to it. Systems biology therefore signals a shift, away from molecular characterisation and cataloguing of the components in the cell, towards an understanding of functional activity.

A key feature of the present text is its systems-theoretic perspective. In this setting, the word 'systems' in systems biology is taken to mean a merger of (dynamic) systems theory with (cell) biology. We focus on dynamics and transient changes occurring within cells. These changes, which in most cases will be molecule concentrations, carry information and are at the root of cellular functions that sustain and develop an organism. The concept by which scientists organise these processes are *pathways*, i.e., *networks* of biochemical reactions. A pathway is an abstraction, a model, of an observed reality. The aim for us is to take the concept of pathways, from simple maps or graphs that name the components and indicate graphically and only roughly their relationship, towards a dynamic simulation of the interactions of proteins in a pathway. It will not be possible to address all areas of application and to provide examples from these. However, it is important to emphasise that the methodologies used for modelling and simulation are generic, i.e., they are applicable to a wide range of processes related to intra- and inter-cellular dynamics. In fact, the mathematical concepts and techniques introduced here are widely used in various other areas, including engineering, physics, chemistry. Learning them as generic tools, has a number of advantages for the student who is interested in broad, interdisciplinary training. Mihajlo Mesarović played an important role in defining the discipline systems biology. Already 1968 he wrote [Mes68]:

> "In spite of the considerable interest and efforts, the application of systems theory in biology has not quite lived up to expectations. [...] one of the main reasons for the existing lag is that systems theory has not been directly concerned with some of the problems of vital importance in biology."
> "The real advance in the application of systems theory to biology will come

about only when the biologists start asking questions which are based on the system-theoretic concepts rather than using these concepts to represent in still another way the phenomena which are already explained in terms of biophysical or biochemical principles. [...] then we will not have the 'application of engineering principles to biological problems ' but rather a field of *systems biology* with its own identity and in its own right."

Since then there have been dramatic advances in technologies including, gene and protein expression assays, confocal microscopy, calcium imaging, and fluorescent tagging of proteins, which allow us to observe reactions in time and space. We should not ignore, the fact that as yet we have some way to go with regard to quantitative stimulus-response experiments that generate time series data suitable for conventional system identification techniques. But even if the technologies are available possibly the greatest hurdle and certainly the reason why it is so attractive, is the human factor: advances in the life sciences will rely on experimentalists and theoreticians working closely together; they need each other.

One might argue that mathematical modelling in biology is anything but new. This is true although the emphasis may have been different, which could be summarised as follows:

1. Mathematical biology is most closely associated with:
   a) population modelling – modelling dynamics of infectious disease
   b) morphogenesis and spatial pattern formation
   c) evolutionary dynamics

   ... a main reference is [Mur02].

2. Mathematical physiology is most closely associated with:
   a) cardiac rhythmicity
   b) modelling the circulatory system, respiration, blood, muscle
   c) hormone physiology
   d) renal physiology
   e) hearing and vision

   ... a main reference is [KS01].

3. Systems biology is most closely associated with
   a) molecular and cell biology
   b) genetics and genomics (omics data)
   c) modelling metabolic and signal transduction pathways

The literature on systems biology is reviewed below. The differences to closely relate areas of genomics and bioinformatics are further discussed in Chapter 1.

The outline of the text is as follows. Chapter 1 provides an introduction to the subject area, including a discussion of the scientific approach and the role of modelling. The 'novelty' of systems biology is that it considers signal- and systems-oriented approaches to modelling and simulation of cell-biological and molecular systems. We are going to introduce the concept of a 'system' from a very general perspective which is then refined and adapted to fit the application under consideration. Systems biology considers dynamics, including transient changes of molecular concentrations and differential equations are therefore unavoidable. Chapter 1 provides a gentle introduction to the key ideas.

For the theoretician it is essential to not only have a basic grasp of molecular and cell biology but also to appreciate the generation of data from experiments. Chapter 2 introduces the two basic modelling concepts for biochemical reaction networks: mass action models and the chemical master equation approach. We are going to provide a thorough discussion of the differences and similarities and on the way learn a number of important or useful mathematical techniques.

Differential equation models do not account for random variations and in Chapter 3 we are therefore going to focus on stochastic models of intracellular processes. The role of stochasticity in cells, in modelling and the comparison to conventional kinetic models is an interesting topic for discussion.

Chapter 4 focusses again on nonlinear differential equation models, the different flavors that are available and applications to cell communication networks (signal transduction pathways). The mixture of biology and mathematics, of basic and advanced material is deliberate. In interdisciplinary research it is important to be able to read a broad spectrum of literature and it is important to develop confidence for the experience that not everything can be understood after the first reading.

The cell cycle is an inherently dynamic intracellular process with a long history of mathematical modelling. In Chapter 5 we introduce the biological aspects and models of different degrees of complexity.

The Appendix with its summary of mathematical notation used in the different chapters and a glossary of technical terms is an idea adopted from biological textbooks to help the reader in finding her/his way through the material. Throughout the text, the most important concepts and terms are indicated in the page margin at the place where they are introduced.

Rostock, 8th October 2012

Literature Review

Systems biology is an emerging area of research and which is truly an interdisciplinary area, combining various disciplines and areas of research. A consequence of this is that although there are already many relevant research journal publications, there are currently few suitable textbooks available. In trying to fill a gap with the present text, we should not suggest that it is possible to cover all aspects of systems biology in one

book. Considering the large number of theoretical methodologies, experimental techniques and biological questions, it will be necessary to consult complementary literature. The targeted audience for the present text are graduate and postgraduate students and researchers from a range of disciplines. The aim is to make the text accessible to students and researchers who may be at different levels of their training/experience. Towards this end we are going to illustrate concepts with plots and line drawings wherever possible. Each Section will give numerous references to research publications and books. In addition, we here give a brief list of textbooks that could help the novice to complement the material presented here.

Although the edited volume, [BB01] was written as a textbook and provides a range of examples for models. It covers many methodologies and application areas, but is necessarily limited to brief introductions which do not allow a more comprehensive treatment of the mathematical basis of the models. Since it was written by practitioners it remains a valuable source book with motivating examples. The monograph by Davidson [Dav01] describes how embryonic development is the outcome of a vast spatial and temporal series of differential gene expressions, and how the control of these depends on a hard-wired regulatory program built into the DNA sequence. Apart from few logical wiring diagrams, mathematical modelling and simulation does not play a role in this book. It does however provide a good example for theoreticians to understand the biological challenge related to regulatory systems that are involved in the development of an organism. The edited volume by Fall et al. [FMWT02] comes closer to the present text, is well written with an interdisciplinary audience in mind and is broader in scope. Somewhat more advanced is the standard text in mathematical biology by Murray [Mur02]. It covers a vast range of mathematical techniques and biological examples. In fact, several older texts in the area of mathematical biology are ideal for studies in systems biology but unfortunately some of these texts are out of print. More recently books specifically directed towards systems biology have been published, including [Pal06] whose focus is primarily on metabolic networks and with a bias towards the author's algebraic methods. The textbook by Klipp et al. [KHK+05] has a very similar outlook and structure as the present text but providing more information on the practical aspects of data handling. Uri Alon [Alo06] has written an excellent text with a focus on transcriptional networks. The present text puts more emphasis on the mathematical aspects, providing a systems-theoretic perspective. In particular, we emphasise the assumptions involved in modelling and provide a comparison between alternative modelling approaches. Our text is therefore slightly more "technical" than "practical" with view to applications of the maths presented therein. While there is a tendency to ignore mathematical details and derivations, they are very important in systems biology. To fit any model to a set of data is easy but to construct a model in which each term is in a meaningful correspondence with biological reality, is difficult. The assumptions made in deriving the model and the comparison between alternative formalisms, although tedious and apparently abstract, are the basis for a useful model; one that reflects our understanding of biology and which we trust to make meaningful predictions (suggesting new experiments for which we invest money and time). To say we build a stochastic model by employing the Gillespie algorithm is not enough if one does not understand the assumptions that were

involved in the derivation.

The mathematical concepts employed by all of the authors are mostly standard techniques that are described in a range of books on applied mathematics. [Kre93] is a standard reference in the engineering sciences and covers a large spectrum of basic mathematical techniques. There are excellent introductory treatments of differential equations available, including [BD01] to name only one. Those texts, written for an engineering undergraduate audience, have gone through various editions, are well illustrated and accessible to biologists. A more advanced but still introductory textbook is [HSD04], introductory texts focussing on nonlinear differential equations are [JS07] and [Str00a].

Mathematical modelling and simulation has been applied to metabolic pathways and a number of excellent texts are available, including, [CB04], [Voi00] for introductory material, whereas [Fel97] and [HS96] are more advanced texts, focussing on metabolic control analysis (MCA). The main difference between signalling and metabolic pathways is that for the latter we can concentrate on steady-states, which means that many problems are of algebraic nature and do not require the solution of differential equations. There are a large number of basic maths books aimed at the bio- or life science student. A good, short introduction to the mathematics that are required for any experimentalist are [Pho97] and [CB99], although they avoid differential equations and probability theory. For statistical techniques that are relevant for generating data, we refer to [QK02]. The books by Eason et al. [ECG80] and Batschelet [Bat79], although written for the life scientists, also introduce differential equations and other more advanced material. [MS99] is an introduction to modelling of dynamic systems and is a good complementary text to the present one.

With regard to software tools, an important development for systems biology is the Systems Biology Markup Language (SBML). This standard provides a computer-readable format for representing models of biochemical reaction networks. SBML is applicable to metabolic networks, cell-signaling pathways, genomic regulatory networks, and many other areas in systems biology. It is an international effort to provide a free and open modelling language, supported by a large group of developers. The web-site `www.sbml.org` provides links to a number of software tools for modelling and simulation but also has a repository for SBML code of models published in the literature. These models are an excellent source for hands-on exercises.

For the theoretician or modeler, there are various excellent introductory textbooks for molecular and cell biology. The comparison between mathematical and biological textbooks is striking. Biology textbooks are often heavy, large in size, rich in colorful illustration and images. A good mathematics textbook will have a couple of black & white line drawings but otherwise must appear rather dull and thin to the reader from the life science community. The complexity of systems dealt with and the level of abstraction used to describe such systems is in both areas very similar and yet there are very different means of representing information and generating knowledge.

A broad general introduction to modern life sciences is available, for example, through [P$^+$01] and [Har01]. Focussing on the cell, the book by Alberts et al. [AJL$^+$02] has become almost a standard text. For microorganisms, [MMP00] provides an excellent

introduction and survey of microbiology. The book by Brown [Bro99] is an accessible introduction to the are of genomics. The biochemistry that underlies the reactions in pathways is covered by various books, including [SBT02] or [MVHA99]. The area of signal transduction is developing very rapidly, and there are few textbooks at introductory level available; on example is [Gom03]. For engineers and computer scientists the introductory text [TB04] provides a concise summary of the most important concepts and principles underlying modern life sciences research.

For the biologist who is interested in interdisciplinary research but whose school days instilled a dislike for mathematics, may find parts of the material presented here challenging. Throughout the text we are going to derive virtually all results in detail, rather than just presenting an equation. If the introductory maths texts, which we have described above, are not sufficient, we provide a very basic introduction to mathematical and statistical modelling as a complement to the present text, available from `http://www.sbi.uni-rostock.de/data_handling.htm`. Furthermore, we are going to encourage computational studies and simulations to 'play' with the ideas presented here. A collection of small programmes is available from `www.sbi.uni-rostock.de`.

Acknowledgements

From my research group, Mukhtar Ullah and Thomas Millat contributed material to several sections and took part in teaching of the associated courses at the University of Rostock. Ulf Schmitz helped with LaTeX and generated many of the PSTricks drawings.

This manuscript was produced using LaTeX, specifically MiKTEX and WinEdt. Our thanks go to those who have contributed to these projects!

1 Modelling and Understanding Natural Systems

One of the most spectacular examples for the wonderful complexity and beauty in nature is the life cycle and total metamorphosis of the butterfly (Figure 1.1). During its morphological development the organism undergoes a dramatic transformation, where one genome is associated with two proteoms. The appearance and life of the two animals that emerge from one genome is very different. Nowadays it is argued that the information to enable this fascinating process is encoded in the *genome* of the organism; whereby the genome is understood as the entirety of the genetic information, encoded in a physical structure known as the DNA molecule. How is this process of the development of an organism realised?

genome

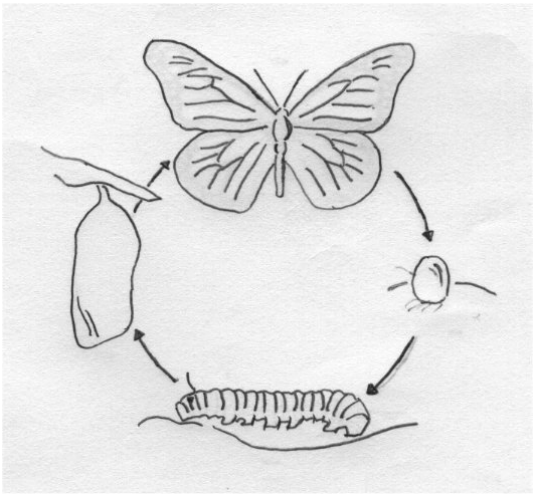


Figure 1.1: The total metamorphosis of a butterfly is an example for one genomes realising two proteomes. It is also a spectacular example for regulation, control and coordination in cell differentiation and the development of an organism.

What has been referred to as the *post-genomics era* of biology, is associated with areas of research that exploit the fact that we have now available the genome sequences for various organisms. The hope has been that using this information we should be able to understand observations at the cell-, phenotypic-, or physiological level. Zooming in from the entirety of an organism to the cell-level, we are covering an enormous scale of magnitude and quantity. While a human can reach heights of say 2.11m, a single cell has a diameter of only about ten micrometers. A human body may consist of some 10^{13} cells, where our largest organ, the liver consists of approximately 300 billion cells alone.

The earth has only about six billion inhabitants, and without loosing the enthusiasm for our research it is sometimes healthy to remind ourselves of the intellectual cosmos in which we are trying to travel.

Not surprisingly then, in many cases where there was an initial hope to discover a simple 'gene/disease' relationship, it was realised that what we are dealing here is a complex web of hierarchical, multi-leveled, regulated dynamic processes. These processes occur within cells and between cells and answering the question as to how a cell takes its place in higher levels of organisation like tissues and organs, means we ought to ask questions about the communication and decision making in cells. The two central aspects of inter- and intra-cellular communication are therefore *signalling* and *gene expression*.

signalling

gene expression

Studying intra- and inter-cellular communication requires sophisticated *technologies* to generate data. The complexity of the processes investigated and thus of the data, motivates the use of mathematics as an extension of common sense reasoning. In the words of Mike Mesarovic "Like you need a set of tools to do experiments, you need a set of concepts to understand".

The scientific approach, by which we wish to investigate questions like those mentioned above, is characterised by the attempt to *model natural systems*[1]. An interesting aspect of interdisciplinary research is the diversity of perspectives and approaches individuals can contribute. The following story explains humorously differences among the modelers in systems biology. A University sends a philosopher, a biologist, a mathematician, a physicist, an engineer and a computer scientist to a hill walking trip in an attempt to stimulate interdisciplinary research. During a break, they rest on a bench, watching a cow in a field nearby. The philosopher asks "I wonder how one could decide on the size of a cow?". Since the object under consideration is a biological species, the biologist responds first: "I have seen many cows in this area and know it is a *big* cow". The mathematician is nervous about the lack of rigor and argues "The true volume is determined by integrating the mathematical function that describes the outer surface of the cow's body." The physicist realises that this function is difficult to get and suggests an experiment: "You lift the cow into a completely filled swimming pool, and then measure the overflowing water, which corresponds directly to the volume of the cow, simple as that!" By now, the engineer had some time to think about the problem and suggests "Let's assume the cow is a sphere...". The computer scientist remained quite all along and is increasingly nervous: "Sorry mates, I thought my laptop wouldn't work up here!".

The underlying philosophy for the present text is to understand cell function through emphasising transformations, processes over the objects. The technological developments in recent years have given us means to characterise the molecular components that make up a cell. For many researchers the function or biological role of a protein is largely defined by its three-dimensional structure. This obsession with molecular characterisation has also led to the misconception of a gene as the causal agent for observations at the phenotype- or physiological level of an organism. The thrust of the present work is that it is systems dynamics that gives rise to biological function. A consequence of

[1] A *natural system* is a system considered in the natural sciences, i.e., physics, chemistry, biology.

this is that the bioinformatics approach, i.e., mining large databases with information about the molecular characterisation of components that make up the cell is necessarily limited.

Rather than focussing on the characterisation of molecular components, we here consider relations between objects and relations between changes of states as more important. This relational outlook is reflected in the following quotations. The physicist Erwin Schrödinger concluded that

> "Life is an emergent, rather than an immanent or inherent, property of matter. Although it arises from the material world, it cannot be reduced to it."

The relational character of our approach is also reflected in the following quotes attributed to the biologist Linus Pauling and universal genius Henri Poincaré, respectively:

> "Life is a relationship among molecules and not a property of any molecule."

> "Science is built up of facts, as a house is with stones. But a collection of facts is no more a science than a heap of stones is a house. "

In fact Poincaré apparently went as fair as saying

> "The aim of science is not things in themselves but the relations between things; outside these relations there is no reality knowable."

1.1 The Systems Biology Approach

Systems biology integrates experimental and modelling approaches to explain the structural and functional organisation of complex (molecular) biological systems as networks of dynamic interactions. It aims at quantitative experimental results and building predictive models. The main focus of the present text are the system-theoretic aspects of systems biology. Systems biology is not genomics, bioinformatics or the integration of data from various Omics technologies. These field are important and complementary but they do not mark the necessary paradigm shift to understand cell function as a well organised system of dynamic processes. Systems biology should be about a shift of focus, from the identification and molecular characterisation of components of the cell, towards an understanding of functional activity. Since systems theory is the study of organisation per se, it seems natural to define systems biology as a merger of (dynamic) systems theory and (cell) biology. This is indeed how it was originally envisaged in 1968. The systems-theoretic perspective on the functional organisation of the cell (as opposed to its structural organisation) is motivated by the conviction that the way the components of the cell are put together, whether we call it structure, order, or organisation, is as material in its causal effects as matter itself. Objects and relations between objects have identical ontological status: Life is a relation among molecules and not a property of any molecule; Causation as the principle of explanation of change, is a relation, not between things, but between changes of states of things.

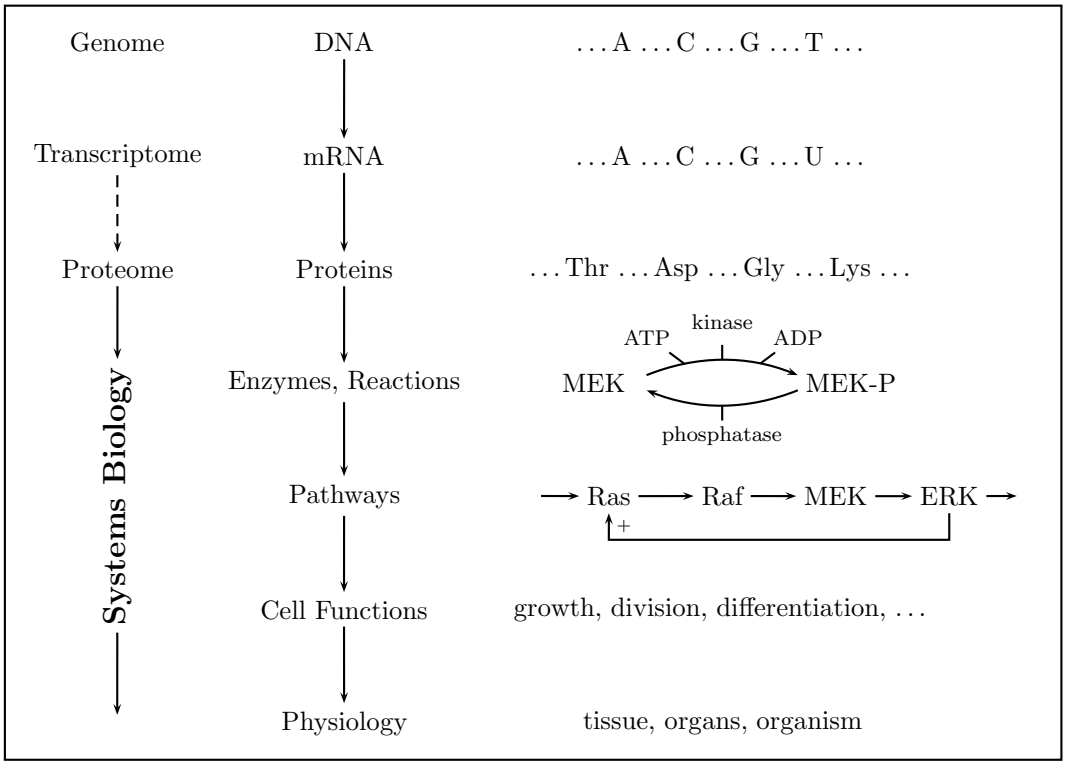


Figure 1.2: Illustration of the scope and remit of systems biology: taking genomics and bioinformatics towards an understanding of functional activity.

Figure 1.2 illustrates the scope of systems biology. The most unfortunate misinterpretation of systems biology is the association with high-throughput, whole-genome data generation. Bioinformatics and Omics approaches are clearly important and complementary but they are also very different (Figure 1.3). In keywords one might summarise the differences between mining and systems-approaches as follows:
The *Mining Approach:*

- ... *pattern recognition*
- ... *association, correlation*
- ... *clustering, classification*

The *Systems Approach:*

- ... *principles (mechanisms) cell function a (spatio-temporal) dynamic process*
- ... *design of stimulus-response experiments*
- ... *modelling and simulation*

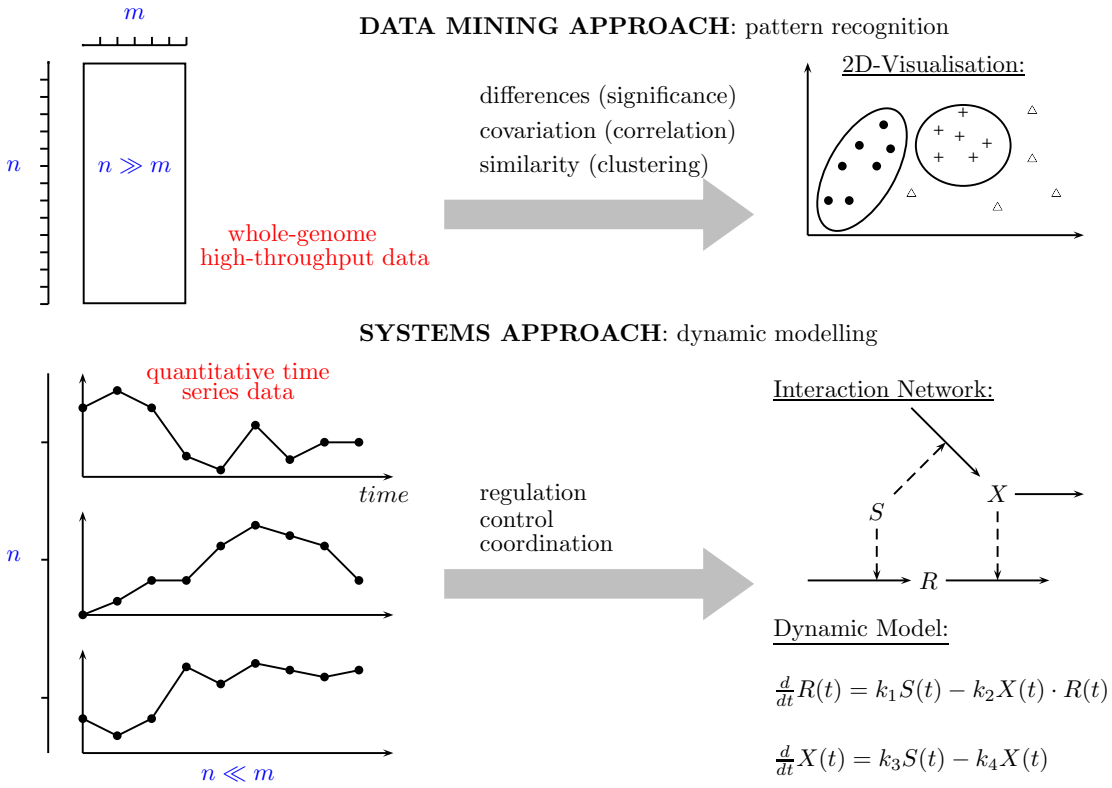


Figure 1.3: Comparison of the mining and systems approach.

1.1.1 Kinds of models

An important question biologists can ask is what mathematical modelling gives them; why should they need mathematical models in studying pathways? The answer we give is that biological function is realised through the dynamic principles of control, regulation, and coordination; all of which are based on feedback mechanisms. Our strategy to demonstrate the necessity of mathematical modelling and systems theory for cell biology is as follows: We first look at the kind of data that are predominantly generated in the life sciences and what type of analysis they require. This will lead us to statistical modelling, testing for differences, correlation and regression. While statistical modelling may not be very popular with experimentalists, it is widely accepted as useful in managing uncertainty. Using a simple three-component biochemical reaction network we show that an understanding of cell function requires a different – system theoretic – conceptual framework to encode causal entailment. We show that the behavior of relative simple systems cannot be understood with conventional experiments. Just like statistical modelling helps us to manage uncertainty, mathematical modelling of nonlinear dynamics helps us to manage organised complexity.

Figure 1.4 summarises the material and arguments in the present section. We are

going to argue for a distinction between a data mining and a systems approach (indicated by the horizontal split). For data mining, statistical modelling provides a basis to manage uncertainty. The systems approach is in our conceptual framework related to dynamic (kinetic) modelling of pathways. The systems approach is to manage complexity arising from dynamic, nonlinear, feedback spatio-temporal interactions. In the mining approach static, qualitative analyzes are prevalent, while the systems approach is inherently dynamic and quantitative. A notable difference between the two approaches is that in bioinformatics, mining biological databases, the analysis of 'data rich', while for kinetic modelling of changes in protein concentrations there is, at present, for most problems, a lack of sufficiently rich quantitative time series data.

Beginning with the column on the left in Figure 1.4, what is referred to as the 'Model Type' describes the information that is sought from data. There are then in the columns to the right formal (statistical/mathematical) representations, as well as quantitative and qualitative visualisations of the models considered. Subsequent sections of the text will guide us through this proposed hierarchy of models. The simplest model is trying to identify the 'presence' of a gene, protein or metabolite. In the data this is reflected in a 'difference' (between two samples, between a sample and an expected value). In practice, this type of analysis is linked to some 'before/after', 'with or without' treatment, considering gene knock-outs, wild-type versus mutant analysis or treating a culture with some perturbation or stimulus of some kind. The next level of analysis considers the 'level' and direction change of more than one variable. Observing the 'covariation' of two variables we may find that one increases in level, while the other decreases. Informally, this leads often to hypotheses about the inhibition or activation of components. Correlation analysis provides the tools for a quantitative analysis of such directional analysis but falls short of describing 'functional relationships'. This will lead us to regression models. We are going to distinguish three interpretations of these models, the first two fitting a curve/plane through the data and thereby providing a summary of the data. We then argue that for cell function, understood as dynamic processes another level of causal, state-space modelling is required. Kinetic or differential equation models will be most powerful with regard to predictions about the dynamics behind cell function but the predictive power comes at a price. To construct such models from data, one requires quantitative, sufficiently rich time series data – something that is very difficult to obtain in systems biology. Despite its value for systems biology, we are going to discuss the limitations of kinetic models, may they be based on differential equations or stochastic representations. Since there is no such thing as a model free of assumptions, we are going to pay attention to this issue.

1.1.1.1 Statistical modelling: Testing for differences

statistical models

The most frequent task for data analysis in cell biology is to look for *differences*. For example, we may look for the expression levels of a gene or protein, comparing wild-type with mutant data or investigating the consequences of a gene knock-out. The data may come from 2D gel, immunoblotting or microarray experiments. Let us associate the outcomes of such comparative experiments with *variables*, that is quantities that can

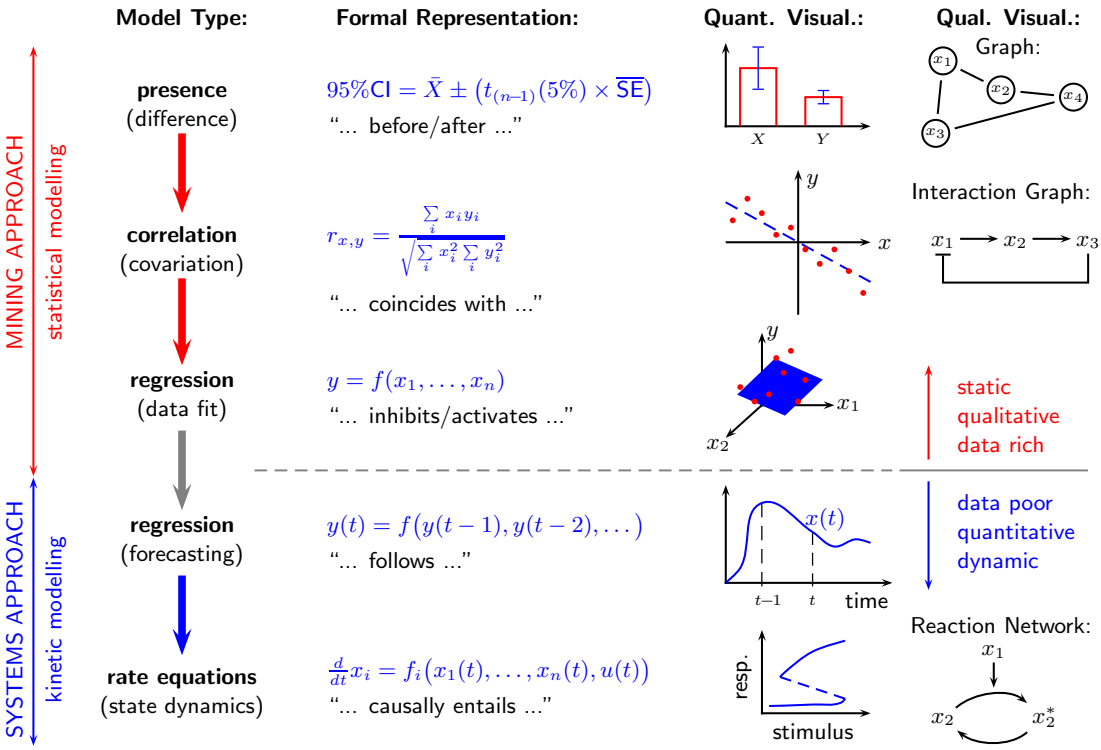


Figure 1.4: Hierarchy and classification of models.

vary. What follows is a presentation of standard textbook material. It is however helpful to explicitly state the concepts and steps involved in managing uncertainty because it demonstrates the level of statistical modelling that is required for event the simplest experiments.

A good experimentalist will conduct several replicate experiments, as he would expect non-biological variability in experimental data. In other words, even under nearly identical conditions the results of measurements are not exactly the same – there is a source of some (hopefully manageable)uncertainty in the experimental system. This uncertainty is frequently referred to as *noise*. What we like to do is to separate noise from the *signal*.

To ensure that this non-biological variability is small, compared to the biological phenomenon or effect investigated, we repeat an experiment. Comparing two variables, repeated measurements will then provide us with two sets of observations, referred to as *samples*, denoted A and B, each containing n values. To deal with the unavoidable uncertainty in experiments, we consider the two datasets as samples, representative of some *statistical population* which we can formalise as 'random variables'. In practice, we would determine the 'average value' of a sample and consider this to be an estimate of what we would 'expect' to observe if there was no noise. Even though this is frequently forgotten, what we have done is to invoke a *stochastic model*. The 'expected value' mis formally known as the *mean value* of a population. In probability theory a population is modelled by a probability distribution. The values of a sample are thus drawn from

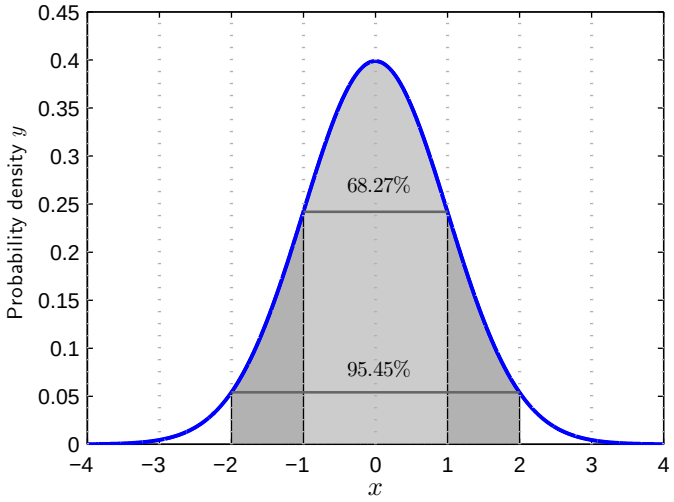


Figure 1.5: Standard normal distribution for a probability law with zero mean and unit variance. The values for x are in standard units ($\mu \pm x\sigma$). This density function serves as a model for observations we make in experiments.

that distribution with a certain probability. The 'average value' referred to is formally known as the *sample mean*

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1.1}$$

where x_i are the individual values in a sample. What we have done is to model the variability, assuming it follows a certain probability law. Technically, the sample mean is thus an estimate of an idealised mean value of some probability distribution. This distribution is formally associated with the technical term of a *random variable*. One would say that the random variable generates a sample value according to that distribution, i.e., with a certain probability. The probability distribution ('probability law') is an abstract model for how a sample is generated.

normal distribution

For a very large number of practical cases, the distribution for random variables is assumed to have the following 'gaussian shape'

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{1.2}$$

where μ denotes the *population mean*, and σ^2 is the *population variance*. Technically speaking this bell-shaped curve *Gaussian-* or *Normal Distribution* is a *probability density function* shown in Figure 1.5. The simplest normal distribution is the *standard normal distribution*. It has zero mean and unit variance. As shown in Figure 1.5, the area plus/minus one standard deviations from the mean captures 68.27% of the area. This means that we would 'expect', on average, sampled values to be in that interval. Since the total area under the curve equals 1, we can say that, the probability that an observation

is found in the interval $[-\sigma, \sigma]$ is 0.68. In general, for any interval $[a, b]$ in X, the probability $P(a < x < b)$ is calculated by the area under the curve. It is useful to remember some of the typical values for the normal distribution:

50% of the observations fall between $\mu \pm 0.674\sigma$.

95% of the observations fall between $\mu \pm 1.960\sigma$.

99% of the observations fall between $\mu \pm 2.576\sigma$.

It is often convenient to 'translate' an arbitrary Gaussian distribution to *standard units* by subtracting the mean and dividing by the standard deviation

$$z = \frac{x - \mu}{\sigma} \ . \tag{1.3}$$

Equation (1.2) is then replaced by the so called *standard form*:

$$p(z) = \frac{1}{\sqrt{2\pi}} \, e^{\frac{-z^2}{2}} \ , \tag{1.4}$$

where the constant $1/(\sqrt{2\pi})$ ensures that the area under the curve is equal to unity, which simply means that the probability of 'some' (any) outcome is one. For our discussion about the role of modelling in cell biology we notice that this density function is an abstract model of experimental observations. We *assume* that the uncertainty in our data can be described by this model. The sample statistics $\bar{X}$, (1.1), and s^2 are considered estimates of the real population μ and σ^2 respectively. The area of statistics is really the real-world interface to probability theory: using statistics there is an (often implicit) assumption of a stochastic/probabilitistic, that is, a theoretical, model. Rather than indulging in more abstract discussion, let us return to our practical situation in which we study gene or protein expression.

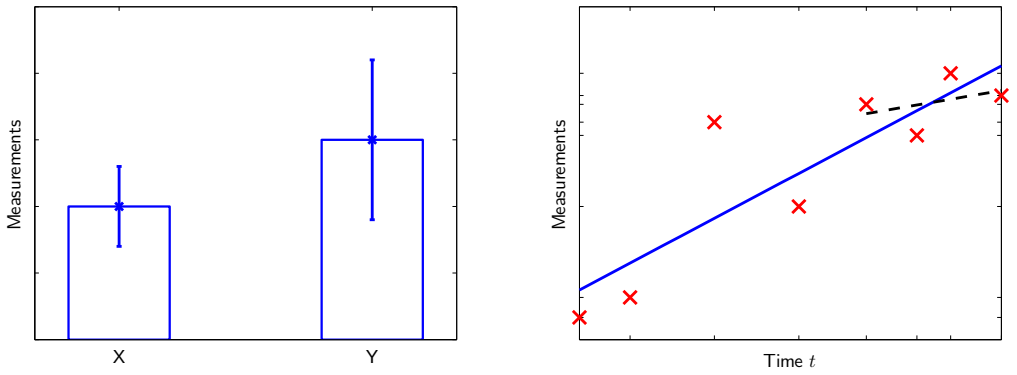
If the two sample mean values we calculated differ, what we may want to do is to argue for a difference in the two expression levels of the biological variables under investigation. The question is then how convince ourselves that the measured difference is 'significant', knowing that the estimate of the mean value itself is varying, that is, subject to some uncertainty? Our psychology makes it easy to convince ourselves of a difference but in order to convince our peers in scientific journals we would have to resort to statistical modelling, whether we like it, or not. This procedure is often less painful than it seems. We start by noting that the sample mean is an estimate, subject to variations and therefore is itself a random variable. Its variability is known and quantified as the (estimated) *standard error* [SR94]

standard error

$$\overline{\mathrm{SE}} = \frac{s}{\sqrt{n}} \ , \quad \text{where} \quad s = \sqrt{\frac{n}{n-1}\sigma_n^2} \ , \quad \text{and} \quad \sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{X})^2 \ .$$

Our sample means would in fact follow a t-distribution, which allows us to calculate a 95% confidence interval (CI)

$$95\%\ \mathrm{CI}(mean) = \bar{X} \pm \left(t_{(n-1)}(5\%) \times \overline{\mathrm{SE}}\right)$$



(a) (b)

Figure 1.6: (a): Bar plot to compare differences in samples. (b): Time series measurements and linear models fit to the data. The solid line is a linear regression model through all data points, while the dashed line is a model fitted only through the last four data points and is used to decide whether the system has reached a steady state.

using some statistical table of critical values for the t-statistic. The *critical value* $t_{(n-1)}(5\%)$ is the number of estimated standard errors $\bar{\text{SE}}$ away from the estimate of the population mean $\bar{X}$, within which the real mean value will be found 95 times out of hundred, i.e., with probability 0.95. Our analysis so far could be visualised with a barplot of the kind shown in Figure 1.6. As can be seen, while the two sample mean values are different, we cannot be certain that the difference is significant. The confidence interval tells us that in a repeated experiment there is a 'chance' that the perceived difference is not real. To improve on our analysis, we can employ what is called a t-test. In general there are three possibilities to test differences with the t-test: testing the difference between a sample and an expected value (one-sample t-test), testing the difference between two samples from the same population (paired t-test), and testing the difference between two samples from two populations (two-sample t-test).

t-test

analysis of variance

Using the two-sample[2] t-test to decide whether the means of two sets of measurements, sampled from two independent populations, are significantly different from each other, we find an answer by following these steps

Step 1: The null-hypothesis is that the mean of the differences *is not* different from zero. In other words, the two groups X and Y from which the samples were obtained have the same mean.

Step 2: The test statistic t is given by the following formula:

$$t = \frac{\text{mean difference}}{\text{standard error of difference}} = \frac{\bar{X} - \bar{Y}}{\bar{\text{SE}}_d}$$

[2]For more than two variables the ANOVA (analysis of variance) test is available.

The standard error of the difference $\overline{\mathrm{SE}}_d$ is more difficult to calculate because this would involve comparing each member of the first population with each member of the second. Assuming that the variance of both populations is the same, we can estimate $\overline{\mathrm{SE}}_d$ using the following equation:

$$\overline{\mathrm{SE}}_d = \sqrt{\left(\overline{\mathrm{SE}}_X\right)^2 + \left(\overline{\mathrm{SE}}_Y\right)^2} \ ,$$

where $\overline{\mathrm{SE}}_X$ and $\overline{\mathrm{SE}}_Y$ are the standard errors of the two populations.

Step 3: Calculate the significance probability P that the absolute value of the test statistic would be equal to or greater than t if the null hypothesis were true. There are $n_X + n_Y - 2$ degrees of freedom, where n_X and n_Y are the sizes of samples X and Y.

Step 4: Using a statistical software package,

If $P < 0.05$, reject the null hypothesis, the sample means are significantly different from each other.

If $P \geq 0.05$, there is no evidence to reject the null hypothesis, the two sample means are not significantly different from each other.

Step 5: The 95% confidence interval for the mean difference is given by

$$95\%\,\mathrm{CI(difference)} = \bar{X} - \bar{Y} \pm \left(t_{(n_X+n_Y-2)}(5\%) \times \overline{\mathrm{SE}}_d\right) \ .$$

Where are we then with our question why a molecular biologist should use mathematical modelling? In order to demonstrate that an experimentally determined difference in expression levels is significant, we saw that one has to resort to *statistical modelling*. All of the concepts are elementary level statistics. The question of why statistical modelling is useful is usually not asked in this context. Instead it is accepted that statistical modelling is part of the scientific method, specifically as a means to judge uncertainty. The statistical models involved are in this case the probability distributions from which we assume the samples were taken or which describe the variation in the sample means. The sample mean, sample standard error etc. are then parameters of these models, which we obtain from data. In other words, if you want to convince others of a significant difference that reflects the presence/absence or under-/overexpression of some gene/protein, then one cannot avoid statistical modelling. The steps described in the following sections will take us naturally to an example that demonstrates why for virtually all pathways, we cannot avoid *mathematical modelling* of the dynamic interactions of the proteins involved. In future, this part of cell biology research will embrace mathematical modelling and simulations in the same way statistical modelling is used in analyzing non-temporal data. As we shall see, it is simply not possible to understand the behavior of a nonlinear dynamic system without mathematical modelling.

1.1.1.2 Associations and regression

With advancing measurement technologies, we are now in many experiments able not only to detect the presence/absence of genes and proteins (testing differences) but we may even be able to manipulate a variable to different levels and monitor changes in others. Such experiment allow us to establish *associations* using the (product-moment) *correlation coefficient*

correlation

$$r_{x,y} = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2 \sum_i y_i^2}} \ .$$

The correlation coefficient is a value between -1 and 1, describing whether the sizes of two variables are independent or whether they covary. A value of -1 means that if one variable is large, the other is small and the other extreme of $r_{x,y} = +1$ means that if one variable is large the value of the other is as well. A correlation coefficient near zero means that there is no association in the sizes of two variables. In other words, using correlation analysis we can test whether the size of one variable coincides with the level of another. As with the test for differences, the correlation coefficient is subject to variations, which we could describe with a confidence interval. This can be looked up in the literature.

Correlations do not imply *causal entailment* and in order to predict the level of one variables as a function of others, we need a different conceptual framework that helps us to establish *lawful* relations among variables. Towards this end *regression models* are a frequently employed framework to analyze experimental data. Let us denote with y the *dependent variable*, which we wish to predict on the basis of the independent variables $x_1, x_2, \ldots, x_m$. Note that we now speak of a *variable*, i.e., a changeable quantity of (natural) system. These variables can assume values, obtained through experiments. A variable is thus a concept, say the 'expression level' of a gene or protein, and the sampling or measurement gives us a value (or sample of values) for this variable. Using small capitals to denote variables as well as sample values can be confusing, which is why the statistical literature often uses capital letters to denote variables. This notation is however not very common in systems theory where Y would be used to denote the space of possible values the variable y can take. The difference of a sample value and a variable should be clear from the context.

regression

There are two main purposes of multiple *regression analysis*: One is to establish a linear prediction equation that will enable a better prediction of y than would be possible by any single independent variable x_j. The aim is to define a set of independent variables that predict a significant proportion of the variance in y. The second purpose of multiple regression is to estimate and fit a *structural model* to 'explain' variation in the observations of *response variable* y in terms of ("as a function of") the independent *regressor variables* x_j. Such 'functional relationship' is formally expressed by a mapping[3], which we denote by f, such that for given values of $x_1, \ldots, x_m$, these are "mapped into"

[3] In mathematics, a 'mapping' is more often called a 'function'. Since this term can be confused with other biological meanings, we use the term 'mapping' (which is formally correct and understood in mathematics as well).

a value y. The mathematician's notation for this is $x_1, \ldots, x_m \mapsto y$, which is the same as writing

$$y = f(x_1, \ldots, x_m) \ , \tag{1.5}$$

which is read as "y is a function of $x_1, \ldots, x_m$". The notation can be further simplified if we combine variables $x_1, \ldots, x_m$ in a vector $x = (x_1, \ldots, x_m)$. Let us denote by X and Y the space of values the variables x_j, respectively y can take. The mapping h is said to establish a 'relation' between variables x_j and variable y. An equivalent representation of model (1.5) is then

$$\begin{aligned} f \ : X_1 \times \cdots \times X_m \ &\rightarrow \ Y \\ (x_1, \ldots, x_m) \ &\mapsto \ y = f(x_1, \ldots, x_m) \ . \end{aligned}$$

The symbol $\times$ is used to combine the value spaces of the variables x_1 to x_m. A model (that is, a mapping) establishes therefore a *rule* assigning to each element $(x_1, \ldots, x_m)$ of $X_1 \times \cdots \times X_m$ an element y in Y. While this notation may on the surface not look very practical, it is however a rather compact representation of what science is all about: establishing relations between system variables!

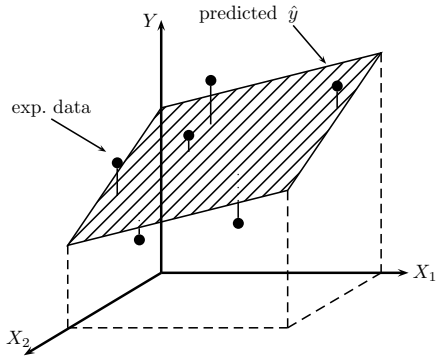


Figure 1.7: Multiple regression fits a hyperplane through data points. This process is guided by an objective function, which tries to minimise the prediction error. The objective function is usually some least squares criterion, minimising the squared difference between the plane and the experimental data point for all datapoints in the training sample.

In order to relate this model to a real system, we need to establish a link to experimental data. The relationship between independent and dependent variables, expressed through f, can be of a linear or nonlinear nature[4]. For mathematical convenience a linear relationship is preferable (but not necessarily realistic). Towards this end, the *linear multiple regression model*, as a special case of (1.5), is given by the following equation

linear regression

$$\hat{y} = \theta_0 + \theta_{y_1} x_1 + \theta_{y2} x_2 + \cdots + \theta_{ym} x_m \ , \tag{1.6}$$

[4]See also the discussion about the definition of nonlinearity on page 27 in Section 1.1.

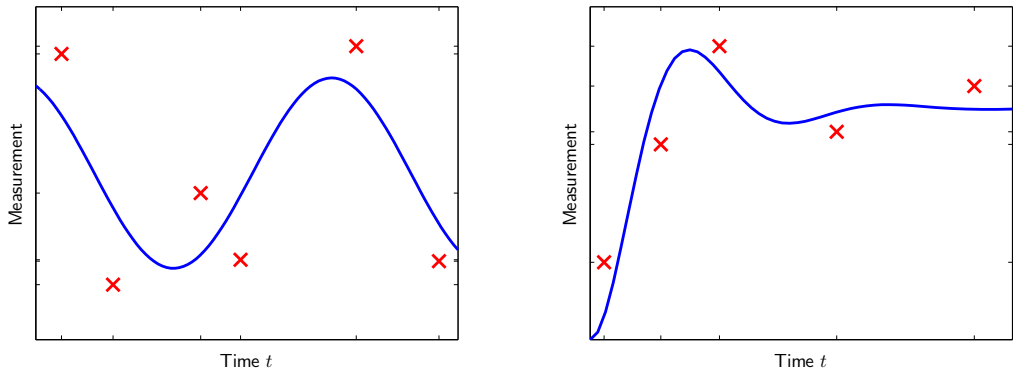
where the estimate of the dependent variable y is a function of m independent variables $x_1, x_2, \ldots, x_m$. $\hat{y}$ denotes a predicted value of y. The coefficients θ_{yj} denote the *partial regression coefficient*, that is, these would be the regression coefficients of y on variable x_j, if all other variables could be held constant in the experiment. More generally, the θ are referred to as *model parameters*. While variables are by definition quantities that change, are expected to change in an experiment, one usually considers parameter values to be constant. They are characterising the system's behavior and should thus not change with time. Considering only two independent variables and plotting data as well as predictions of y as a function of x_1 and x_2, we see that the linear regression model fits a plane through the data (Figure 1.7). Just as in the case of testing for differences in simple comparative experiments, the estimation of parameter values from experimental data is likely to involve some noise, which is often accounted for by adding a noise term w to the regression model:

$$y = \theta_0 + \theta_{y_1} x_1 + \theta_{y2} x_2 + \cdots + \theta_{ym} x_m + w$$

One can either think of w as being some *measurement error/noise*, linked to the instruments, or we can think of w being the residual influence of independent variables that are not among the x_1 to x_m chosen to represent the natural system (as predictors of y). In either case, one considers w to be a random variable with a definite probability distribution. This in turn means that y is a random variable as well. The model is geometrically a surface that is fitted through the data. The estimation of parameter values is thus a process by which we 'fit' a surface through the data points, so as to minimise the prediction error. There are well established algorithms and tools available for this purpose [Wol01].

1.1.1.3 Time series analysis

What we have so far assumed for all of our experiments is that measurements of gene/protein expression levels do not depend on the point in time they are taken. Many experiments are however of a stimulus-response nature, that is, we induce a change in one variable and observe the subsequent change in one or more other variables. For example, in cell signalling we would stimulate a pathway by adding ligands to a culture and in many other typical experiments we study the response to stress, including changes in pH levels, changes in temperature or oxygen supply or modifying the level of small molecule concentrations. In order to use any of the statistical tools described above, we must be sure we can assume that in these before-and-after experiments the system is at an initial point in time (say t_o) at rest (in a steady state) and following the stimulus (perturbation, stress) the system will eventually settle into another steady state. In other words, we must demonstrate that our before-and-after measurements are independent of time. This can only be done by a series of later measurements to test whether there is no further change. Measurements at only two time points would therefore rarely suffice. Consider the time series plot in Figure 1.6(b), the dashed straight line is fitted to the data of the last four time points to test whether there is any significant change. The



(a) (b)

Figure 1.8: Cyclic and transient dynamics which cannot be studied by trend analysis.

model is given by the regression equation

$$\hat{y}(t) = \theta_0 + \theta_1 t \ ,$$

where θ_1 is the regression coefficient and θ_0 determines the intercept of the straight line with the y-axis. If the slope is close to zero, we could argue that there is no significant further change to variable y, the system would be in a steady-state. From the given time series, using only the last few time points to test whether the system has reached a steady state, we used the regression model to account for some of the variation of y in terms of variation in the independent variable, which is here time t. As for the t-tests we can now determine the standard errors of regression statistics to test for significance. We do not pursue this further as this is a standard technique in the literature (e.g. [SR94]).

Studying cell functions, including cell growth, cell differentiation, apoptosis and the cell cycle, we notice that these are dynamic processes where for an understanding of the principles that underly observed data time, respectively change matters. The simplest approach to accept an influence of time is to consider a simple trend model. This leads us to the second use of a regression model, illustrated by the solid line in the plot on the right in Figure 1.6(b). We fitted a curve (straight line) through the time series to test whether there is a temporal trend or whether the system has reached a steady state (in which case time would not matter). Such simple trend-model will however only be adequate for rather simple systems and will have little explanatory value. The values of the dependent variable y are only explained in terms of the independent variable, time t. In Figure 1.8 we see two typical dynamic responses which would not sufficiently well modelled by the straight line regression model. There are now two possible avenues to proceed: We introduce a forecasting model in which the current value of y at time t depends on past values of y and secondly we introduce a state-space model based on rate equation and state-variables (next section).

If time series observations are available data from the past should contain information about the future. Let $y(t)$ denote the value of the variable of interest in period t. A

forecasting

forecast for the period $t+h$ made at the end of period t, has the form

$$\hat{y}(t+h) = f\big(y(t), y(t-1), \dots\big)$$

where $f(\cdot)$ denotes a suitable (linear or nonlinear) functional relationship (mapping) that explains $y(t+h)$ in terms of past values of y. For virtually all realistic cases, the behavior our system under consideration will not be determined by one gene or protein but many others. (Additional variables may be introduced by considering the activation state of proteins). Based on a multiple time series, $y_j(t)$, $j = 1, \dots, q$, a forecast of the j-th variable is then expressed as

$$\hat{y}_j(t+h) = f_j\big(y_1(t), \dots, y_q(t),\ y_1(t-1), \dots, y_q(t-1), \dots\big)\ . \tag{1.7}$$

The aim of modelling is then to determine suitable mappings $f_1, \dots, f_q$. Let us conclude this short introduction to forecasting by stating the standard conceptual framework for time series analysis. Assuming a prediction horizon of $h = 1$ time steps and linear mappings, we get

$$\hat{y}(t+1) = \theta_0 + \theta_1 y(t) + \theta_2 y(t-1) + \dots + \theta_k y(t-k+1)\ , \tag{1.8}$$

where $k \leq q$ is the chosen depth of past values that are considered sufficient for predictions of y at time $t+1$. Because the predictions of variables are based on their past values, these models are refereed to as *autoregressive models* [Lüt91]. The resemblance of this time series model with the multiple regression model (1.6) indicates the kind of methods that are employed to estimate parameters of such models.

The formalism above may appear to be a *deterministic* representation but can also be interpreted as a q-dimensional (multivariate or vector) *stochastic process*

$$y\ :\ I \times S\ \rightarrow\ Y_1 \times \dots \times Y_q\ ,$$

where I is an index set (e.g. the set of all integers or positive integers) such that for each fixed $t \in I$, $y(t,s)$ is a q-dimensional vector of random variables, denoted by y_t. S denotes the set of all elementary random events and s is an element of S, i.e. $s \in S$. The underlying probability space is usually not mentioned. A *realisation* of a (vector) stochastic process is a sequence (of vectors) $y_t(s)$, $t \in I$, for a fixed s. A realisation of a stochastic process is thus a mapping $S \rightarrow Y_1 \times \dots \times Y_q$ where $t \mapsto y_t(s)$. A time series, generated by this stochastic process is usually denoted $y_1, y_2, \dots$, leaving out the arguments in the brackets. Referring to the established literature (e.g. [KS03]) one notices that in successful approaches either the nonlinearity of the natural system under consideration to be a small perturbation of an essentially linear stochastic process, or they regard the stochastic element as a small 'contamination' of an essentially deterministic nonlinear process. Methods for genuinely nonlinear stochastic processes are not well established.

1.1.1.4 Kinetic modelling

The general regression model (1.6) has three possible uses/interpretations [Fre97]:

(i) A representation or summary of a dataset.

(ii) To predict the value of y from $x_1, \ldots, x_m$.

(iii) To predict the value of y as a *causal* consequence of *changes* in x_j.

For the first use, equation (1.6) describes the expected value of y for a given data set (assuming a linear relationship amongst $x_1, \ldots, x_m$ and using a least-squares criterion for fitting this equation to the data). For this use of (1.6) it is not necessary to interpret variables x_j as causes of y. The second possible use of the regression model is to predict values of y, based on data that were not included in the dataset used to estimate the parameter values. Here again no causal link between the x_j and y is required but instead it is necessary to assume the process, which generates the data, is time-invariant or stationary in space and time. The first two interpretations are a *data fitting* exercise. The third use and interpretation of (1.6) does assume a causal link between independent and dependent variables. Causal inference in this context, means that it is assumed that a value of y is the result or consequence of *changes* in x_j. Causal entailment does therefore imply some form of intervention (stimulus, perturbation), the consequence of which is a change (in space and/or time).

Autoregressive models (1.7) and (1.8) may be suitable for time series forecasting but the predictions are not based on equations that encode the principles or mechanisms of a system which generate the data we observe in experiments. In other words, these models have little explanatory value, they are not suitable if the structure of the equations should reflect our hypothesis about the nature of the biological interactions among system variables. In cell biology we are interested in interactions, in particular regulatory mechanisms, which leads us to state-space models. The concept of a *state*, denoted with the letter x, of a system assumes that the behavior of the system at any point in time can be encapsulated by a set of states, together with some transition rules which specify how the system proceeds from one state to another (over time). A dynamic system is then considered to have inputs or control variables that provide a stimulus (denoted with the letter u) and there are observable output or response variables. The outputs of the system, denoted by the letter y, may be identical to the state variables but may not. For example, in cell signalling the activated and inactivated form of a protein would be considered as two separate state variables, x_1 and x_2. However, in immunoblotting experiments one may only be able to measure the total concentration, i.e., the sum of the two state variables. This situation can then conveniently represented by the $y = x_1 + x_2$.

causation

Causation manifests itself only through changes in states. What we require is a formalism in which changes are encoded, naturally leading us to the formulation of differential (rate) equations:

$$\frac{d}{dt}x_i = f_i\big(x_1(t), \ldots, x_n(t); u(t)\big)$$
$$y(t) = h\big(x_1(t), \ldots, x_n(t)\big)$$

where the operator d/dt denotes the rate of change of system variable x_i, which is a function of the values of variables x_j at time t. In this formal representation, u, y could denote vectors, comprising several output variables and h is a mapping that relates state-variables with measured outputs. Rather than describing this mathematical formalism in greater detail, we consider a practical example from which it arises. To keep things simple, we assume a network of biochemical reactions (i.e., a pathway) with one independent input variable S (providing the stimulus to the system), one internal state-variable E and one output or response variable R. We know/hypothesise that this system has generated the time series data in Figure 1.9. To begin with we do not know the structure of the biochemical network under consideration but have available experimental data sets of the kind shown in Figure 1.9. Our example is going to demonstrate that a relatively simple system can display rather complex behavior, which we would not

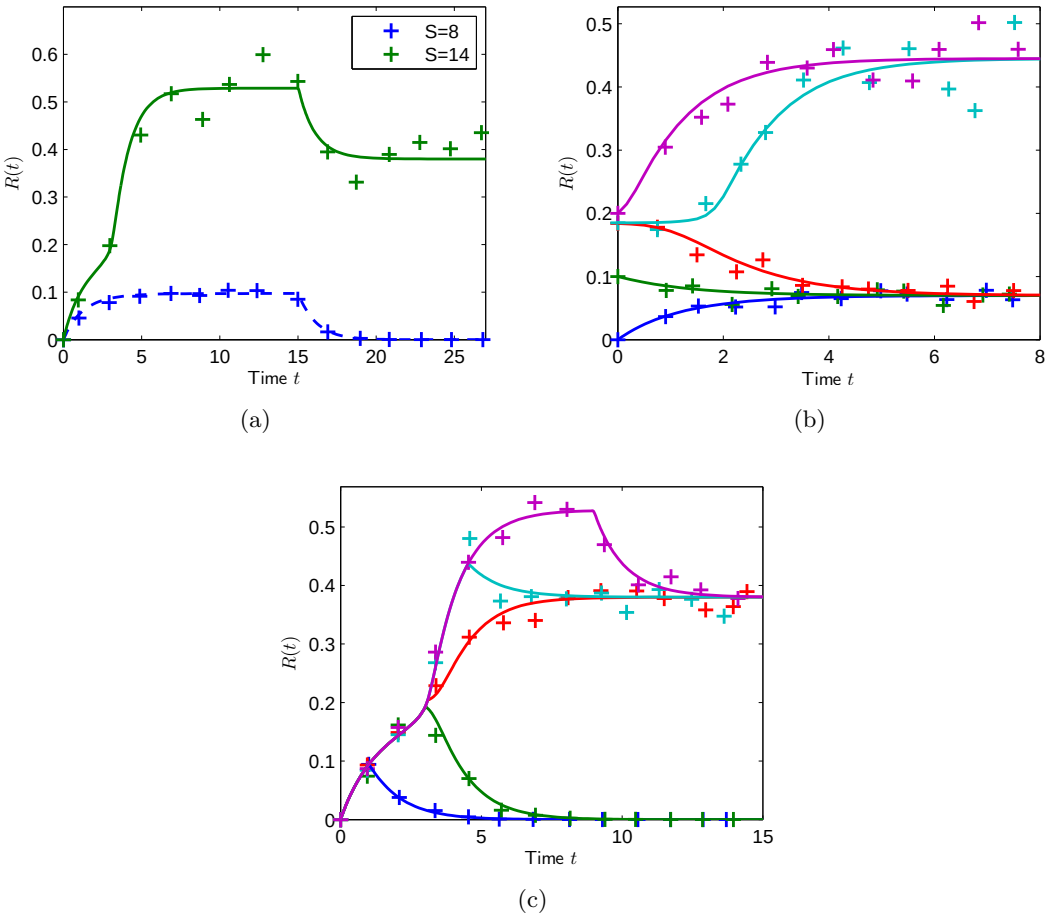
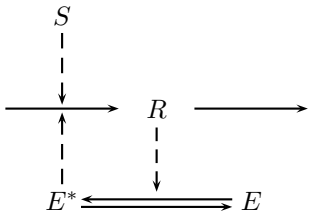


Figure 1.9: (a): Response of two different stimuli, the first subcritical and the second supercritical. (b): Relaxation into the steady state for different initial response signals R_o. (c): Response to a supercritical signal $S = 14$ of different duration.

be able to understand without mathematical modelling. Say we are conducting a first experiments in which we stimulate the system at a level of $S = 8$. The stimulus is kept constant until $t = 15$, when it is set to a level of $S = 0$. The response is shown as a dashed time curve in the plot in Figure 1.9(a). Looking at this simple time plot a modeler could suspect a relatively simple linear model. The nonlinear nature of the underlying system surprises us however if we stimulate the system again at a level of $S = 14$ (shown as the solid line time plot in the same figure). What appears surprising here is not only the curve that describes the transient behavior of the system into a new steady state but when the stimulus is removed the system does not return to its previous steady state: This system is *bistable*. How can we explain these experimental observations? One might suspect that if we are dealing with one and not two systems, that the system undergoes a change, that the kind of proteins involved is different or that the nature of interactions is altered. As we shall see, it is possible for is one and the same system, one set of equations with a set of fixed parameter values, to display such behavior.

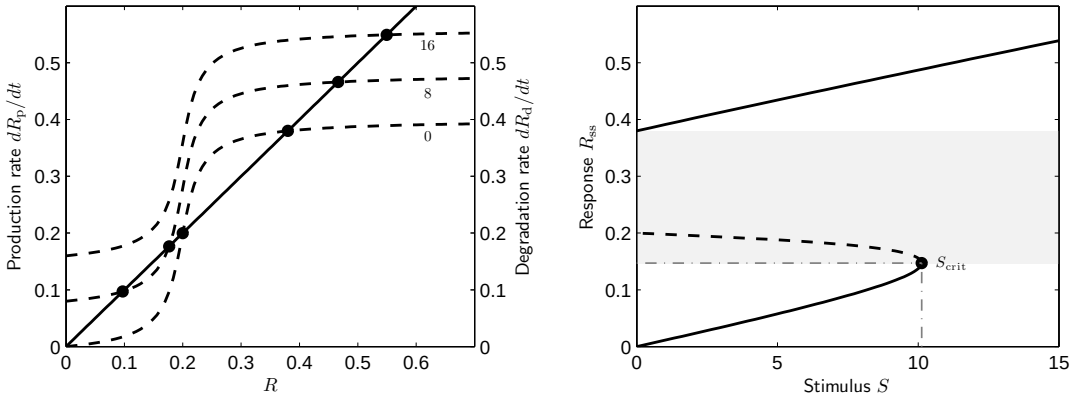
Consider the desirable situation with a particularly curious experimentalist and a lot of funding for experiments who is the position to conduct several other experiments on our system. As shown in the plots in Figure 1.9(c), taking one level of stimulus but considering different initial conditions for the response variable R, we obverse that the experimentally observed behavior depends on the initial conditions as well. Finally, conducting experiments with different durations for the stimulus, we again observe a range of curves – all of which are generated by exactly the same system. All simulations shown in Figure 1.9 were generated by one single equation and one fixed set of parameters. The message of this example is that if we are to understand the interactions of this (rather simple) three-component system we require a model that encodes hypotheses about the nature of the interactions of the system variables. In our example, a positive feedback loop is responsible for the bistable behavior. A diagram of the system illustrates this:



In this system, an increase in variable R activates E (say through phosphorylation). E in turn contributes towards the synthesis of R, which means that there is a positive feedback loop in the system. Following the drawing of a pathway map, the next step is to devise a model. Without explaining the details (described later in the text) the rate equation for this system is

$$\frac{d}{dt}R(t) = k_0 E^*(R) + k_1 S(t) - k_2 R(t)$$

where $E^*(R)$ describes the activation of enzyme E and for which we have chosen a mathematical formulation introduced by Albert Goldbeter and Daniel A. Koshland [GKJ81].



(a) (b)

Figure 1.10: Bistable mutual activation switch. (a): Comparison of the rate of degradation (solid line) and the rate of production for different signal strengths (dashed lines). (b): Signal-response curve or bifurcation plot.

We have omitted this equation since the form of this expression does not matter for our discussion. To complete our story, mathematical tools from dynamic systems theory provide us with techniques to devise such models and analyze (or simulate) their behavior. The rate equation for $R(t)$ consists of two parts: the rate of production, $k_0E^*(R) + k_1S(t)$, and degradation $k_2R(t)$. Plotting these parts separately we obtain the plot shown on in Figure 1.10(a). The intersections of these two curves are the steady-states of the system, in which the rate of degradation and the rate of production are balanced. As can be seen, the number of steady states of the system varies for different levels of stimulus. The plot in Figure 1.10(b) shows the stimulus-response curve. For each dot in this curve a stimulus was applied and the value of R in steady state was determined. This pictures shows us the critical level of the stimulus at which the system displays its *switching* behavior. Once the system has reached the upper branch it does not return to lower levels of R even if S returns to zero. Such graphical tool to investigate the behavior of nonlinear dynamic systems is referred to as a bifurcation plot.

If a molecular biologist is given sufficient research funding to conduct all experiments necessary to obtain the time series shown in Figure 1.9, would he be able to understand the system's behavior without mathematical modelling? Our example shows that pathways involved in realising the cell functions are nonlinear systems in which variables interact dynamically. If we believe that such processes as cell differentiation, proliferation and apoptosis are relevant for an understanding of diseases, then we must consider these as dynamic systems. A consequence of our analysis is that systems biology experiments are far more time consuming and expensive than those we have considered at the beginning of our discussion. Furthermore, whether we like it or not, nonlinear (dynamic) systems theory is going to become part of cell biology, sooner or later.

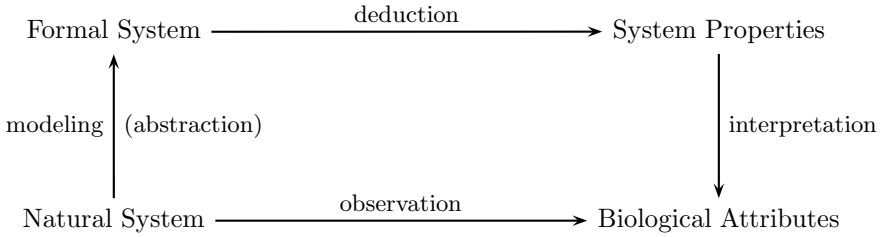


Figure 1.11: Mathematical modelling is a process by which we establish an abstract representation of a natural system. For a model to be a valid representation, it is necessary to relate it with observable attributes of a biological system. Diagram adapted from [Mes68].

1.1.2 Systems Theory

The definition of a system as a set of objects with relations defined among those objects is not quite a proper mathematical expression and will require refinement. Instead of focussing on *things*, we are going to emphasise *processes*[5]. The two things are not necessarily different: The existence or recognition of something as an object is at the same time a process. Appearances or phenomena are entities equivalent to the act by which they are apprehended. Likewise, a set is equivalent to the process or mapping that identifies elements of a set. A thing is identified by distinguishing it from something else.

process

We are going to distinguish two kinds of systems: a natural system, which is an aspect of the phenomenal world under consideration and a formal system, which is the mathematical framework employed to represent the natural system (Figure 1.11). A formal system, theory or mathematical model is a collection of concepts. Establishing relationships between concepts is process of *modelling*. Mathematical modelling is the process by which we interpret a natural system using a mathematical model. Establishing the validity or truth of concepts requires the *a priori* concepts of space and time. This may not be obvious but if we, for example, consider the concept of a derivative, we find that any explanation of this using pen, paper or words, will use the ideas of movement, change, or rate of change. A mathematical model, or model for short, is thus a realisation of the formal system employed. It should not escape our notice that the process of reasoning is a system itself. Since the world of objects is conditioned on the subject, science is not dealing with an independent reality. It is therefore desirable for us to look out for a conceptual framework in which not only to represent natural and formal systems but also the process of modelling itself. I believe this may be found in the context of category theory.

modelling

As simple as our definition of a system, as a set of objects with relations among these, may appear, since mathematics can be reduced to set-theoretic principles, our definition of a system is in fact as rich as the field of mathematics itself. Since we are going to derive every aspect of our understanding from this pair of a set and relation, we also realise the necessity or *a priori* nature of mathematics to establish truth in the phenomenal world.

[5] A process implies succession and so we are going to describe a dynamic system as a process, i.e., a sequence of events.

That there is something else than the world of phenomena or "Wirklichkeit" which we can experience, can be seen from the fact that every argument we can establish has to have an absolute minimum of one premise and one rule of procedure, e.g. IF p, THEN q, before it can begin, and therefore begin to be an argument at all. So every argument has to rest on at least two undemonstrated assumptions, for no argument can establish either the truth of its own premises or the validity of the rules by which itself proceeds. We are therefore limited by the third class of objects that forms one of the roots for Schopenhauer's principle of sufficient reason. The world as we know it, is our interpretation of the observable facts in the light of the theories we invent. The world of objects is thus conditioned by the subject: there is something that is grasped and something else that grasps it. In line with Kant and Schopenhauer, the entire world of phenomena or appearances is the world of representation in time and space, connected by causality. The world of phenomena is the self-objectification of the *noumena*. The noumena is what things are in themselves, or from our systems perspective, things in themselves are understood as things apart from *relation*. The existence of things independently of human representation or cognition, the unknowability of things, is the non-relational aspect of things. Schopenhauer described this as *will*. Before we further enter the slippery territory of philosophical arguments, we return to our definition of a system as a set of objects with relations, and refine it in mathematical terms.

noumena

general system

With the definitions of sets and relations at hand, we rewrite our definition of a *general system*[6], which is now considered as a relation on variables/indicators/items defined in set theoretic terms

$$\mathcal{S} \subset \times\{O_j\} \ , \tag{1.9}$$

or equivalently

$$\mathcal{S} \subset O_1 \times O_2 \times \cdots \ ,$$

where the curly brackets denote a list or set, $j = 1, 2, \ldots$ and $\times$ denotes the so called Cartesian product, that is, a combination of sets. A *complex system* is a relation on systems/subsystems, i.e.,

complex system

$$\mathcal{S} \subset \times\{\mathcal{S}_j\} \ ,$$

such that there is a distinct behavior of the complex system while the integrity of the subsystem is preserved. The characteristics of a complex system include multilevelness and hierarchies. This definition of a complex system is more specific than the usual interpretation of molecular and cell-systems as being complex due to the fact that usually we are dealing with a large number of variables and nonlinear relationships among those. More generally we consider complexity

- a property of an encoding, i.e., the number of variables in a model.
- an attribute of the natural system under consideration, e.g., the connectivity, nonlinearity of relationships.

[6] The notion of a general system is due to Mihajlo Mesarović, who developed the most general and most complete conceptual framework for general systems [MT75]. For a comprehensive overview of systems theory see [Kli91].

- related to our ability to interact with the system, to observe it, to make measurement or generate experimental data.

A mathematical *model* is a realisation of a formal system, an abstract representation of some aspect of the real world which is called the natural system under consideration. The process of modelling identifies and defines the variables and relationships among them. For our purposes, a mathematical model is subsequently identified with a set of parametric equations. The process of *simulation* is an 'execution' of, or 'experiment' with a model. For example, in case of differential equations models, a simulation refers to numerical integration, as the process of finding a solution to the set of equations. In case of mathematical models that use probabilities rather than direct values, a simulation run generates *a single* realisation of the stochastic process.

modelling

simulation

The cell is a complex system and in studying the cell, we are considering *processes* and *components*. The components interact through or in processes to generate other components. We define the *organisation* of a system as the pattern or configuration of processes. Pathways are therefore an example of describing organisation in cells. The *structure* of a system is the specific embodiment (implementation) of processes into material components. In modelling there is therefore a duality between abstract concepts and physical entities. Even the most basic concepts by which we make the world plausible, 'space' and 'time' have no material embodiment and the coexistence between the physical or material real and the mathematical or abstract should not be a real problem. In the present text we are trying to exercise this exciting aspect of modelling in the context of cell signalling, where the interactions of molecules lead to changes in protein concentrations, which define a signal that in turn carries or *transduces* information. We are going to consider two kinds of dynamics: *intracellular* dynamics and *intercellular dynamics*. Intracellular dynamics are the processes by which the genome and proteome act, react and interact within the context of the cell so as to bring about its bio-physical structure and to realise its function (role) in a larger whole. Intercellular dynamics are the interactions of cells to realise higher levels of structural organisation, including tissue or organs. It is obvious that both concepts are tightly coupled. Each cell has a copy of the entire genome and there are processes that run quite independently of other cells, cell-cell interactions are necessary for an individual cell to realise its role in a larger whole that is some tissue, an organ or the entire organism.

organisation

structure

inter-/intra-cellular dynamics

The key to successful modelling is that there has to be some correspondence between the causal structure of the natural system and the formal system. The decoding of a natural system into a mathematical model needs to be validated through an encoding that allows predictions about the natural system. In case of molecular systems modelling is hampered by *complexity*, and *observability*, i.e., the difficulties in making direct observations and measurements. Both, the complexity and observability of such systems lead to *uncertainty*.

With all these abstract formalisms, one may ask what the practical use of systems theory is. Although not simple, once it is mastered systems theory, mathematical modelling and simulation of dynamic systems provides a conceptual framework in which to discuss the following questions:

What is the influence inputs have on the system?

What do the outputs tell about the system?

What are the consequences of differences in the model structure?

More specifically, related to signal transduction pathways we wish to gain a better understanding of the following questions:

How does the physical interaction of molecules create signals, store and transmit information?

How are signals integrated and turned into decisions?

What is the role of space, location and translocation?

How are decisions, actions and reactions of cells integrated into higher levels of structure, organisation and response pattern?

This discussion can proceed at different levels. An abstract but also most general discussion of properties of general systems, including issue such as coordination, hierarchies, multilevelness etc. can be conducted in the context of (1.9) [MT75]. We hereafter follow a different strategy. We first argue the case for differential equations as a suitable approach to represent signal transduction pathways and thereby concentrate on a special case of the general system (1.9). Once we have shown that nonlinear state-space models are an appropriate modelling framework, we generalise our analysis of a particular pathway model to a class of pathways.

Let us first demonstrate how we get from the formal definition of a system as a relation of objects to some representation of dynamic systems. The purpose is not to frighten the reader with complicated mathematics but to show that there is a general systems theory which generalises many of the models that have been published in systems biology. We are going to lower the level of abstraction immediately after this short excursion. Beginning with the general system

$$\mathcal{S} \subset O_1 \times O_2 \times \cdots ,$$

when (sub)systems interact they do this through defined interfaces, which may refer to as inputs and outputs. In cell signalling, membrane receptors suggest themselves as inputs that receive a stimulus in form of ligands binding to them. The expression level of a target gene may be defined as the response, respectively output of the pathway as a system. More formally, a signal transduction pathway may thus be described as the system

$$\mathcal{S} \subseteq \Omega \times \Gamma ,$$

where Ω and Γ related to the stimulus and response respectively. Most important for our philosophy of systems biology is that we understand a pathway not as a static graph but as a network of biochemical reactions, that is, a dynamic system which establishes a

causal connection between stimulus and response. This means that we define a pathway formally as the mapping

$$\begin{aligned} \sigma \,:\, \Omega \;&\to\; \Gamma \\ \omega \;&\mapsto\; \gamma \end{aligned}$$

where a stimulus ω (respectively response γ) is defined as a temporal sequence of events

$$\Omega = \{\omega \,:\, I \;\to\; U\} \;, \qquad \Gamma = \{\gamma \,:\, I \;\to\; Y\} \;.$$

At any point in time $t \in I$, our system $\mathcal{S}$ receives a stimulus $u(t)$ to which it responds at time t with $y(t)$. We assume that stimuli and responses take their values from constant sets U respectively Y.

A key concept in the systems biological approach to cell signalling is that of a signal $\omega\colon]t_0, t_1] \to U$ acting on the system between time t_0 and t_1, generating a response $\gamma\colon]t_0, t_1] \to Y$. In general, we denote the set of acceptable stimuli in terms of the mapping

$$\omega \,:\, I \to U \;,$$

and for the response

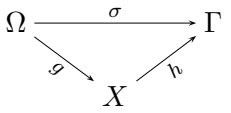
$$\gamma \,:\, I \to Y \;,$$

where for say concentrations we assume that data are described by positive real numbers such that $U \subseteq \mathbb{R}_+^m$ and $Y \subseteq \mathbb{R}_+^q$. For $m = 1$ and $q = 1$ the vector-valued notation reduces to a single signal or time series. If $I = \mathbb{Z}_+$ we have a discrete-time system model, which corresponds to the collection of experimental data, and in which case we could consider Ω and Γ as finite-dimensional vector spaces, encoding sequences of events

$$\begin{aligned} \omega &= \big(u(0), u(1), \dots\big) \;, \\ \gamma &= \big(y(1), y(2), \dots\big) \;. \end{aligned}$$

In modelling it often makes sense to assume a signal that is continuous in value and time. For $I = \mathbb{R}_+$ a continuous-time system, with $\omega\colon]t_1, t_2] \to U$ in Ω and $\gamma\colon]t_2, t_3] \to Y$ in Γ. The entire sets of stimuli and responses that the cell can realise form the objects Ω and Γ of our definition of a stimulus-response system.

The description of a pathway σ as a mapping that takes an input sequence to a response pattern, is an external description, without consideration for the interactions that generate this input-output behavior. Extending our abstract model with a state-space X, we have the following state-space representation:



Dynamic pathway modelling is the process in which we identify and characterise the mathematical objects Ω, Γ, X as well as the mappings that put these objects in relation to another. At any point in time, the behavior of the system is thus encapsulated by

the state $x \in X$. The temporal evolution of the state, $x(t)$, implies the existence of a state-transition map

$$\varphi : I \times I \times X \times \Omega \ \rightarrow \ X$$

whose value is the state $x(t) = \varphi(t; t_0, x, \omega)$, that is, an element of the state-space X. In this setting, the state x at time t arises from an initial state $x_0 = x(t_0) \in X$ at some initial time $t_0 \in I$ under the action of stimulus $\omega \in \Omega$. The graph of φ in $I \times X$, called the *trajectory* and describes the temporal evolution of the system. If we are to investigate a pathway or cell function in experiments we assume that, at least for the experiment, the system is *time invariant*, i.e.,

$$\varphi(t;\ t_0,\ x,\ \omega) = \varphi(t + s;\ t_0 + s,\ x,\ \omega')$$

for all $s \in I$. A dynamical system is *continuous-time* if I is a set of real numbers and *discrete-time* if I are integers. $\mathcal{S}$ is *finite dimensional* if X is a finite-dimensional space and we speak of a *finite-state* system if X is a finite set. A finite system is more commonly known as a *automaton*.

Depending on the nature of the biological process under consideration but also motivated by personal preference and mathematical convenience, one can choose among a number formalisms by which to translate biological understanding into a mathematical model. To name but a few, we can we can distinguish between mass-action-, power-law or S-system-, and Michaelis-Menten models that are based on differential equations. The computer scientist tend to find the setting of automata theory, machines and formal languages (π-calculus, Petri-nets, process algebras) more preferable. For any formalism mentioned, one question is always the role or relevance of randomness. If the answer is that random fluctuations cannot be ignored, we are led to stochastic models, based on Markov-processes, Langevin-, or Chapman-Kolmogorov equations.

By far the most frequently employed approach is to represent temporal changes of protein concentrations as differential (rate) equations

$$\begin{aligned} \dot{x} &= V\big(x(t), u(t)\big) \\ y(t) &= h\big(x(t)\big) \end{aligned}$$

Here $\dot{x}$ describes the rate of change in state vector $x = (x_1, \ldots, x_n)$ at time t. Here V is again a map that in the context of dynamic systems theory is referred to as a vector field. The map h describes the observations we make on the states of the system. A special case of the above representation is a class of nonlinear systems models for which there exists some experience

$$\dot{x} = f\big(x(t)\big) + \sum_{i=1}^{m} g_i\big(x(t)\big)u_i(t)\ , \qquad y_j = h_j\big(x(t)\big)\ , \qquad 1 \leq j \leq q\ .$$

Variable u is considered a state-independent external 'control'-input to the system. Not surprisingly, most of the research around this formalism has emerged from the control engineering community.

We are going to consider a dynamic systems approach to understand inter- and intracellular processes. To identify a system's behavior we require stimulus-response time course data. For most molecular and cell biological experiments it is not straightforward to generate sufficiently rich and quantitative data sets that satisfy the theoretician. The richest set of system-theoretic methods is available for *time-invariant linear* systems. Time invariance means that although the system variables change over time, the relationships among the variables do not. If we were to repeat an experiment the same mathematical relationships would be identified. The definition of linearity deserves attention as different scientific communities have different interpretations. For example, the standard model of a MAP kinase pathway is a *linear* cascade of three modules (cf. Figures 4.33 and 4.34). Linearity in this context refers to a *series* connection of modules. Any feedback loop branching of one of these modules and influencing a variable further up in the pathway is occasionally described as *nonlinear feedback*. This is rather unfortunate and should be avoided. Let $y_t(\theta, u_t)$ be the output of the model[7] with parameters θ at time t and which is due to the input $u(\tau)$, which has been applied from initial conditions between time zero to t, $0 \leq \tau \leq t$, $t \in \mathbb{R}_+$. A model is said to be *linear in its inputs* (LI) if the outputs satisfy the superposition principle with respect to the inputs, i.e., if

time-invariant systems

linearity

$$\forall(\alpha, \beta) \in \mathbb{R}^2,\ y_t\big(\theta, \alpha u_1(t) + \beta u_2(t)\big) = \alpha y_t\big(\theta, u_1(t)\big) + \beta y_t\big(\theta, u_2(t)\big)\ . \tag{1.10}$$

A system is thus nonlinear if the output from the system is not proportional to the input. If we draw a graph of the output against the input on the abscissa, a linear system would define a straight line while a nonlinear system would diverge from the straight line. While this definition is common in engineering and applied mathematics, statisticians usually refer to a different kind of linearity: A model is said to be *linear in its parameters* (LP) if its outputs satisfy the following superposition principle with respect to its parameters:

$$\forall(\alpha, \beta) \in \mathbb{R}^2,\ y_t\big(\alpha\theta_1 + \beta\theta_2, u(t)\big) = \alpha y_t\big(\theta_1, u(t)\big) + \beta y_t\big(\theta_2, u(t)\big)\ . \tag{1.11}$$

For example, the simple straight line equation $y = \theta_1 x + \theta_2$ is LI and LP. We are going to return to a discussion of the difference between linear and nonlinear systems on page 224.

If $\mathcal{S}$ is linear and time invariant, we can express the relationship between dependent and independent variables by the following equation:

$$y(t) = \sum_{k=0}^{t-1} \Theta_{t-k} u(k), \quad t \in T\ . \tag{1.12}$$

where $\Theta_t \in \mathbb{R}^{p \times m}$ denote the coefficient matrices which characterise the process and we have assumed a time-discrete system, i.e., $T \subset \mathbb{Z}_+$. For each t, (1.12) specifies a set of q

[7] Note that we slipped the time dependence from the brackets, $y(\theta, u(t), t)$ into the subscript y_t. This is to simplify the notation with no other meaning.

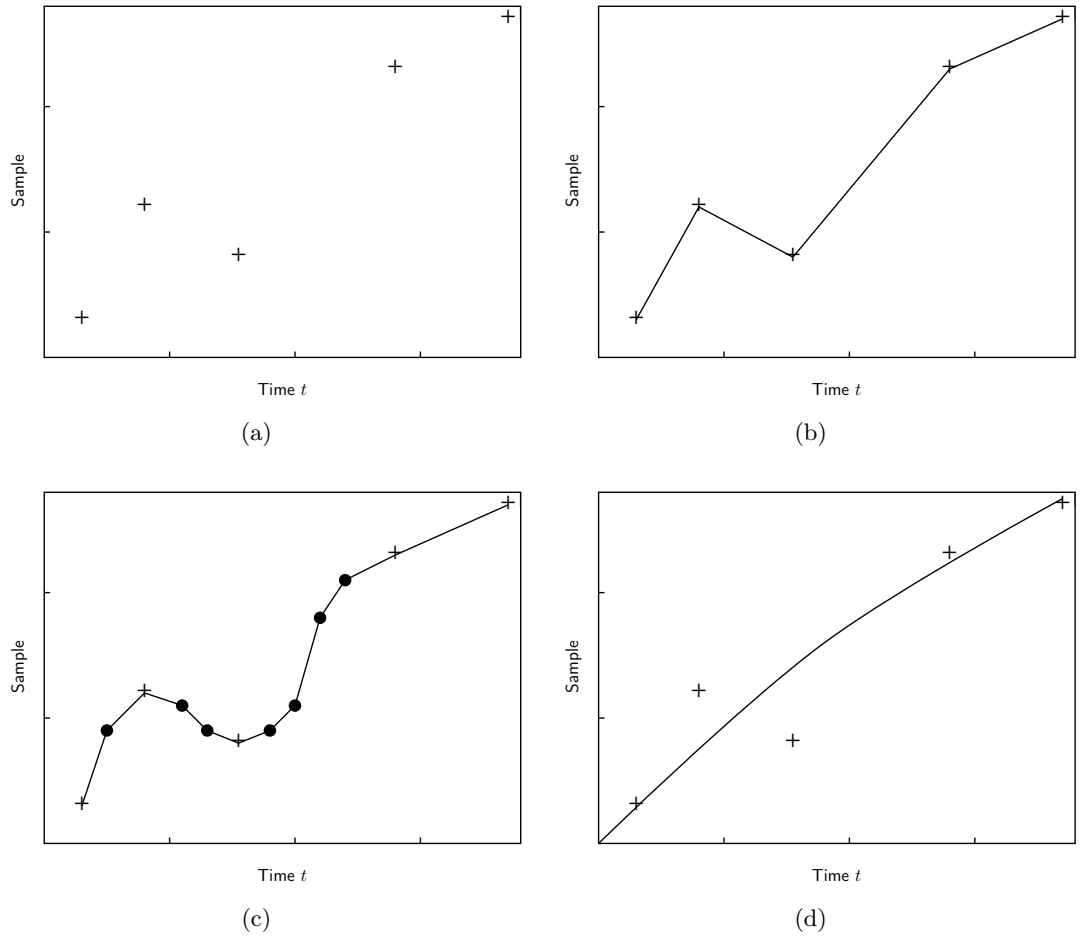


Figure 1.12: Sampling and approximation of signals. (a): Experimental data define only points and in biological experiments measurements are frequently taken at equidistant time points. (b): The common representation of data as a line plot implies a model of linear interpolation, ignoring any possible random fluctuations. (c): For system identification and parameter estimation, it is usually assumed that measurements are sampled at equally spaced intervals. (d): In some cases it may be feasible to fit a curve through the points and model changes with a continuous-time model. It could however be the case that the dip at the third time point is not an outlier but rather an important biological aspect. Which assumption is correct depends on the context in which the data are generated.

equations in $q \cdot m$ unknowns of the matrix Θ_t. We find that for the linear system there exist a one-to-one correspondence:

$$\mathcal{S} \cong \{\Theta_1, \Theta_2, \Theta_3, \ldots\} \ .$$

For the example of a linear discrete time-invariant system, the relationship between input $u(t)$ and output $y(t)$ is *linear*. Let $U \subset \mathbb{R}^m$, $Y \subset \mathbb{R}^q$, $\Theta(t) \in \mathbb{R}^{p \times m}$, the system can then also be represented in a *canonical* form using matrices $F \in \mathbb{R}^{n \times n}$, $G \in \mathbb{R}^{n \times m}$

and $H \in \mathbb{R}^{q\times n}$

$$\Theta(t) = HF^{t-1}G \quad t = 1, 2, \ldots$$

The problem of modelling is then to define the dimension of X, for which the sequence $\Theta_1, \Theta_2, \ldots$ is uniquely determined; leading to the discrete-time state-space model[8]:

$$\begin{aligned} x(t+1) &= Fx(t) + Gu(t) \\ y(t) &= Hx(t) \qquad\qquad x(t_0) = x_0 \ . \end{aligned}$$

Given the present state $x \in X$ defined by $x(t)$ and input $u(t) \in U$ the map φ determines the next state and for every state x, the output map h determines an output $y(t)$. It is usually assumed[9] that X is equal to or a subset of the Euclidean space of real numbers, $\mathbb{R}^n = \mathbb{R}_1 \times \cdots \times \mathbb{R}_n$, and thereby any state can be represented as a point in X (see page 26). Note that the concept of state is a general notion, defining a set of n state-variables such that the knowledge of these variables at some initial point in time $t = t_0$ together with the knowledge of the input for $t \geq t_0$ completely determines the behavior of the system for any time $t \geq t_0$. State variables need not be physically measurable or observable quantities. The state-space representation is well established and forms the basis for automata theory and control theory. An *automaton* is a discrete-time system with finite input and output sets U and Y, respectively. We say the automata is finite if X is a finite set[10]. Automata theory has been used to model numerous systems including gene networks. However, we note that the finiteness of spaces, in which the inputs and outputs take their values, may require a quantisation of measurements and discretisation. With typically short time series and a lack of replicate measurements this may imply an unreasonable loss of information.

automaton

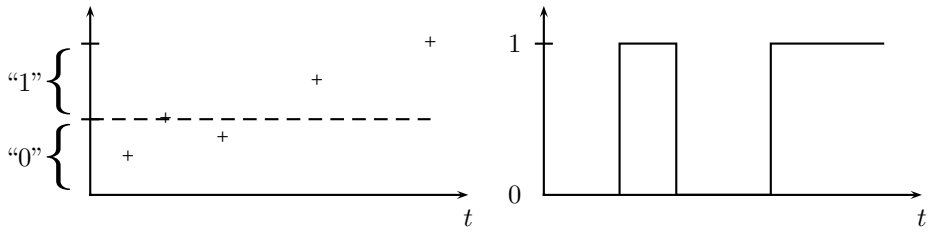


Figure 1.13: Quantisation of signals. Using a threshold on the amplitude of a signal, we can convert the data into a binary signal suitable for modelling boolean networks.

The state-space representation introduced above may look fairly general but there are more general cases to consider. For example, consider the state-space system

$$\begin{aligned} \dot{x}(t) &= V\big(x(t), u(t), w(t), t\big) \ , \qquad x(t_0) = x_0 \\ y(t) &= h\big(x(t), u(t), m(t), t\big) \ , \end{aligned} \tag{1.13}$$

[8] Any control engineering textbook will provide further reading on properties of such systems, and how the matrices can be identified from experimental data.

[9] There are a number of mathematical requirements associated with the definitions and reasoning in this section. We leave these details for later chapters and refer to the extensive literature in mathematical systems and control theory, including for example [Son98], [Nv90], [Isi89] and [Bel90].

[10] The state of a linear dynamic system, continuous-time or discrete-time evolves in $\mathbb{R}^n$, whereas the state of an automaton resides in a finite set of *symbols*.

where V can be changing with time and $\{m(t)\}$, $\{w(t)\}$ are stochastic processes. Representation (1.13) is usually too general for a detailed mathematical analysis of a specific model. The first step to a tractable model is by assuming that ϕ is not dependent on time, i.e., the system is *autonomous*[11] or time-invariant

autonomous systems

$$\dot{x}(t) = V\big(x(t), u(t)\big) \ .$$

Let us look at an example of the system above. In subsequent chapters, we are considering molecular populations that change as the result of chemical reactions. Under the hypotheses that all elementary reactions obey first-order kinetics and the compartment in which the reaction takes place has a constant temperature, the generalised mass action model is given by the set of coupled equations

$$\begin{aligned}
\frac{dx_1}{dt} &= -\theta_1 x_1 + \theta_2 x_2 \\
\frac{dx_2}{dt} &= \theta_1 x_1 - (\theta_2 + \theta_3) x_2 \\
\frac{dx_3}{dt} &= \theta_3 x_2 \ .
\end{aligned}$$

The structure of this mathematical model is given by prior knowledge or hypotheses about the system. The parameters θ_i are kinetic rate constants of the elementary reactions, and the state variables x_1, x_2, and x_3 are the concentrations of the reacting chemical species. All of them therefore have a precise meaning and interpretation[12]. These kind of models, describing *observed* changes, are therefore also referred to as *phenomenological* or *knowledge-based models*. Identifying the parameters of such a model from experimental data is called *parameter estimation* , and for nonlinear differential equations relies on sophisticated statistical tools. The decision of a model structure and parameter estimation, together are referred to as *system identification*.

phenomenological models

One of the question that is going to arise is how we deal with stochasticity or randomness. When we are using differential equations we not arguing that the underlying physical process is deterministic. One interpretation is that we describe the mean value of what is intrinsically a random process, arising from the interactions of molecules in the cell. There may however be situations in which we model a process involving only a relative small number of molecules in a larger volume. In this case we may prefer a stochastic model. Here again we have a selection of modelling approaches to choose from. Randomness of the states leads to *stochastic differential equations*:

stochastic differential equation

$$\dot{x} = V\big(x(t), w(t)\big) \ ,$$

[11] The term *autonomous* is more frequently used in the context of differential equations, while *time-invariance* is more commonly used in the context of applications of differential equations to natural systems.

[12] In subsequent chapters we are going to use a different notation for biochemical reactions. The one based on x_i is commonly used in applied mathematics and in the context of differential equations. In biochemistry capital letters and square brackets, $[S_i]$, are used to denote the concentration of a molecular species S_i.

If $w(t)$ is considered to a Gaussian process this representation is called *Langevin equation*:

$$\dot{x} = V\big(x(t)\big) + G\big(x(t)\big)w(t) \ ,$$

where G is a matrix. We are not going to pursue this approach further but will look at Markov models and stochastic master equations in later chapters.

There are however also random fluctuations in measurements that are not related to the nature of molecular interactions but have to do the measurement itself. Adding an error model for such *measurement noise* $m(t)$ in the observation of the system gives us

measurement noise

$$\begin{aligned} \dot{x}(t) &= V\big(x(t), u(t)\big) \ , \\ y(t) &= h\big(x(t), u(t)\big) + m(t) \ , \end{aligned} \tag{1.14}$$

For measurement noise one usually assumes a Gaussian process $\{m(t)\}$, for which the joint probability distribution of $\{m(t)\}$ is multivariate *normal*. We are not going to consider such noise models. The reason is that at present it is rather difficult to validate such models with experimental data. This is however not to say that the consequences of perturbations to chosen parameter values are not an important part of the modelling process. To the contrary, an analysis of the sensitivity of the system's behavior on parameter variations is an important aspect of systems biology.

1.1.3 Differential equations

While the physical object we focus on is the cell, the most fundamental process we consider is that of a biochemical reaction. In these reactions the concentration of a molecular species, referred to as the *substrate*, is changed. We are going to describe networks of coupled reactions with the help of equations. As an example let us consider the frequently occurring *Michaelis-Menten equation*:

$$V = \frac{V_{\max} \cdot S}{K_M + S} \ , \tag{1.15}$$

where the variable on the left-hand side of the equation is the *dependent variable* or 'output'. $V_{\max}$ and K_M are fixed numbers, i.e., constants or parameters, while S denotes the concentrations of the substrate and is our *independent variable*. A graphical representation of equation (1.15) is shown in Figure 1.14. This equation, which is widely used in the biological sciences, is 'hiding' the fact that it is derived from a *differential equation*. Differential equations describe the rate of change of a variable over time, and are thus ideally suited to describe changes in concentrations in biochemical reactions. We are going to use differential equations extensively throughout and the present section serves as a gentle introduction to this most useful conceptual framework.

Pathways are the concept by which knowledge of interactions of proteins in cell functions is organised. A *pathway map* exhibits the names of the molecular components, whose interactions govern the basic cell functions. This leads us to a definition of pathways as biochemical *networks*. One motivation for systems biology is to bring these static

pathways

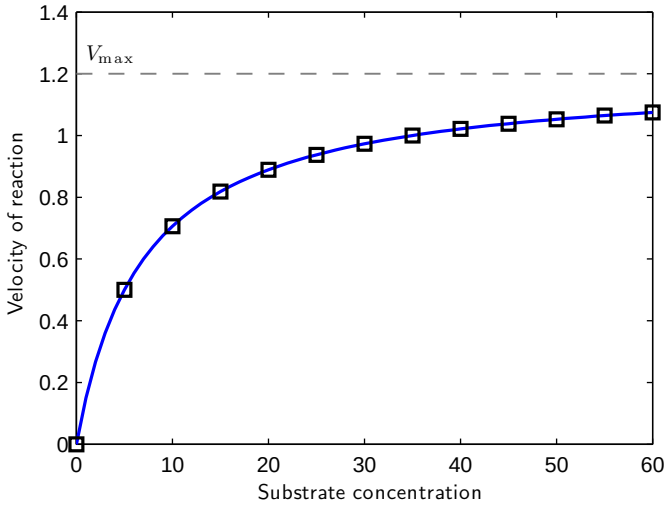


Figure 1.14: Michaelis-Menten plot describing an enzyme kinetic reaction.

diagrams to life by modelling and simulating the biochemical reactions that underlie cell function, development, and disease.

Although a pathway or pathway map describes molecules, their physical state and interactions, it is an *abstraction* and has no physical reality. An *abstraction* is a *representation* of an aspect of the real world, so as to reason about it. A *model* is the consequence of this process, may it be by graphical means (e.g. the pathway map), by natural language, or through mathematical formalisms.

pathway map

A pathway map is thus a model, and the first step in the *art* of modelling is to identify which proteins need to be included in the model. One approach to support this process is to consider a pathway as a network of biochemical reactions. If we denote the chemical species and/or the modifications by capital letters, the following collection of biochemical equations would formalise a pathway:

$$R_\mu \colon l_{\mu 1} X_1 + l_{\mu 2} X_2 + \cdots + l_{\mu n} X_n \xrightarrow{k_\mu} \cdots \tag{1.16}$$

where X denotes a chemical species participating in reaction channel R_μ, the '+' signs represents a combination, and the arrow a transformation. The coefficients $l_{\mu j} \geq 0$ indicate how many molecules of reactant species X_j are involved, and k_μ describes the rate at which the reaction proceeds. The reader unaccustomed to biochemical equations should not worry, we are going to introduce and discuss these representations in greater detail throughout the text.

While in the biochemical reaction equation participating components are denoted by X, in the mathematical equations we use small letters x to denote variables of the model. For example, the pathway

$$X_1 + \alpha X_2 \xrightarrow{k_1} \beta X_3 \xrightarrow{k_2} \alpha X_2 + \gamma X_4 \ , \tag{1.17}$$

can be split into two *reaction channels*

$$R_1 \ : \ X_1 + \alpha X_2 \xrightarrow{k_1} \beta X_3 \ ,$$
$$R_2 \ : \ \beta X_3 \xrightarrow{k_2} \alpha X_2 + \gamma X_4 \ .$$

When a reaction occurs, the changes to molecule populations is determined by the stoichiometric coefficients. For example, in reaction channel R_1, X_2 looses α molecules, while in R_2 it gains α molecules.

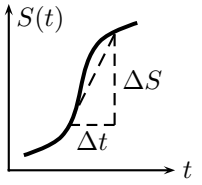
The framework we are going to adopt to model biochemical reaction networks is that of nonlinear differential equations. I consider differential equations a *natural* choice for the following reasons. Causation is the principle of explanation of *change* in the realm of *matter*. For anything to be different from anything else, either *space* or *time*, or both have to be presupposed. Causation is a *relationship*, not between things, but between changes of *states* of things. As we going to demonstrate, differential equation models are an ideal means to realise this philosophy. Let consider the simplest of biological examples to demonstrate the view of causation as an explanation of change. Studying a protease cleaving peptide bonds in a substrate protein, we stipulate that

> "The rate of proteolysis is somehow proportional to amount of substrate."

The purpose of mathematical modelling is to translate a hypothesis into a set of equations. These equations have parameter values, which we obtain from experimental data. In our example, a direct translation of the statement above is the following differential equation

$$\frac{d}{dt}S = -k_p S(t)$$
$$\qquad\quad \text{parameter (fixed)} \quad \text{variable (changes)}$$

The operator d/dt is used to represent the *rate* of change of the substrate concentration $S(t)$ over time. The rate is thus also a velocity. As such it is related to the slope of the concentration profile, determined as the limit of Δt going towards zero. This is illustrated in the following diagram:



$$\frac{d}{dt}S = \lim_{\Delta t \to 0} \frac{\Delta S}{\Delta t}$$

As indicated, we call a *parameter* a value of the model that does not change for the time interval of interest, while a *variable* changes. The former are typically rate constants while the latter are concentrations. Such *rate equation* are mathematically speaking differential equations with $dx/dt = \ldots$ describing a rate of change over time. If the mathematical expression on the right-hand side is a "usual one", not involving

rate equation
differential equations

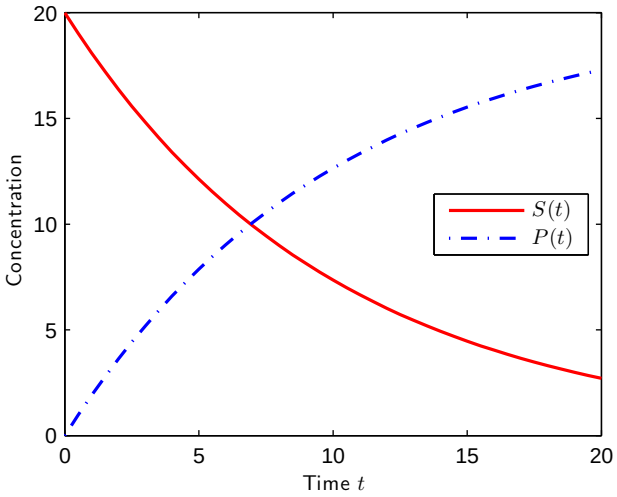


Figure 1.15: Simulation of the differential equation model for the proteolysis.

distributions like Dirac-distribution or the Heaviside step-function. These are then referred to as *ordinary* differential equations (ODEs). They consider changes only over time and not space. If diffusion across an area of the cell has to be considered we would end up with a description using *partial* differential equations (PDEs). Accounting for different regions of the cell, e.g., the cytosol and the nucleus, can be realised with ODEs by introducing different variables for the same protein, located in different regions. A more serious threat to the differential equation approach comes from the translocation of proteins, e.g., nucleocytoplasmic export. Time delays in a feedback loop more often than not have a significant effect on the dynamics of a system. An explicit representation of such phenomena leads to *delayed* differential equations. Needless to say that the theory for partial and delayed differential equations is more complicated than for ordinary differential equations.

In our example, the *state* of the system is fully determined by the equation that describes the value of the substrate S at any time t:

$$S(t) = S_0 \cdot e^{-k_p t} \ , \tag{1.18}$$

where S_0 denotes the initial concentration of S at time $t = 0$. This equations also called the *solution* to the differential equation above. Although there is also the "product" concentration $P(t)$, the result of the proteolysis, its value can be directly determined from $S(t)$ using the conservation relation $S_0 = S(t) + P(t)$

$$P(t) = S_0(1 - e^{-k_p t}) \ . \tag{1.19}$$

The *simulation* of this mathematical model produces plots of (1.18) and (1.19) (Figure 1.15). As trivial as this example may seem, modelling arbitrary complex pathways is a straightforward extension of the approach demonstrate there.

Let us return to our pathway example (1.17) and consider the following set of ordinary differential equations with dynamic variables $x_1, \ldots, x_4$, corresponding to chemical species $X_1, \ldots, X_4$:

$$\begin{aligned}
\frac{d}{dt}x_1 &= -k_1 x_1(t) x_2^{\alpha}(t) \\
\frac{d}{dt}x_2 &= -\alpha k_1 x_1(t) x_2^{\alpha}(t) + \alpha k_2 x_3^{\beta}(t) \\
\frac{d}{dt}x_3 &= \beta k_1 x_1(t) x_2^{\alpha}(t) - \beta k_2 x_3^{\beta}(t) \\
\frac{d}{dt}x_4 &= \gamma k_2 x_3^{\beta}(t) \ .
\end{aligned} \tag{1.20}$$

Looking at the structure of these equations, we recognise the generalised representation for (1.16):

$$\frac{d}{dt}x_i(t) = \sum_{\mu=1}^{M} \nu_{\mu i} k_\mu \prod_{j=1}^{n} {x_j}^{l_{\mu j}}(t) \qquad i = 1, 2, \ldots, n \tag{1.21}$$

where the units of the concentrations x are mol per liter, M=mol/L. For simplicity, we omit the commonly used square brackets [ ] to denote concentrations. The set of nonlinear differential equations (1.21) describes a large class of systems but is by no means the most general representation.

The aim of this section was to give a glimpse on how biological and experimental information is translated into mathematical expressions. We have omitted various questions but will return to these in due course.

1.1.4 Dynamic Systems Theory

The mathematical model (1.21), as general as it may seem, remains a particular choice for a conceptual framework in which to model biochemical reaction networks. We are going to discuss various other approaches to represent intracellular processes. The choices involved in selecting one or the other approach are guided by practical or experimental considerations, mathematical convenience and personal preference. To guide a discussion of different approaches it is sometimes helpful to see whether modelling approaches can be generalised into a more abstract framework. An increased level of abstraction can mean that one looses the link of the model to a particular experimental setting but it can also help in the development and understanding of a mathematical model. Another motivation is that dynamic systems theory has developed various techniques and tools within other disciplines such as control engineering, applied mathematics and physics. By generalising a biological model we can use these tools and techniques for our purposes. To illustrate this, let us consider a generalisation of (1.21), which frees us from the discussion of how the right-hand side of the differential equation should be constructed. Towards this end we consider the ODE as a *mapping*

$$f \ : \ X \times \mathbb{P} \rightarrow \ \mathbb{R}$$

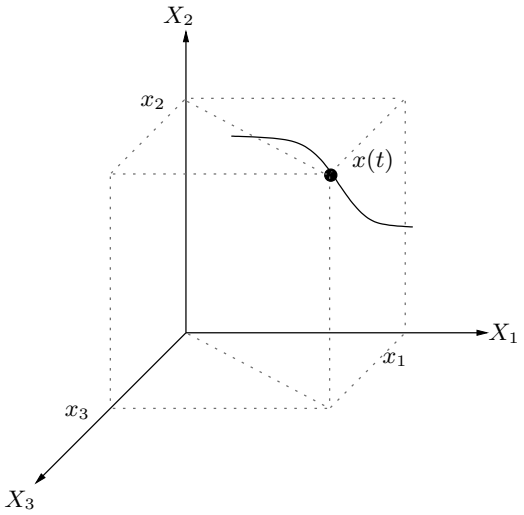


Figure 1.16

where X denotes the *state-space* and $\mathbb{P}$ a parameter space. At any point in time $t \in I$, the concentrations are collected in a vector, called the *state* of the system and denoted

$$x(t) = \big(x_1(t), \ldots, x_n(t)\big) \ .$$

The variables x_i are called state-variables and the space in which they take their values is the state-space. For a particular set of parameter values $\Theta \subset \mathbb{P}$, the current state $x(t)$ is associated with or mapped to a rate of change in the set of real numbers $\mathbb{R}$. Any particular parameter value is denoted as $\theta \in \Theta$. The model of ordinary differential equations (1.21) can then be generalised as

$$\dot{x} = f\big(x(t), \Theta\big) \ , \tag{1.22}$$

where $\dot{x}$ is short for dx/dt and f is a n-valued mapping. The vector-valued mapping f determines the dynamics of the system[13]. If one imagines the state (vector) of concentrations as a *point* in state-*space* $X = \mathbb{R}^n_+$, the temporal evolution of the system describes a curve, called *trajectory*, in $X = X_1 \times \cdots \times X_n$. The analysis of the dynamics of the system may thus be conducted in *geometric* terms, as illustrated in diagram 1.16. The system (1.22) is said to be *unforced* (or *autonomous* since there is no independent *input* $u(t)$ inside the bracket. We are also going to assume that rate coefficients are rate constants, i.e., the parameters of the system are not changing over time. One speaks of time-invariant systems, although their behavior is of course dynamic, changing over

[13] In subsequent sections various variations of (1.22) will be discussed. The mapping f may alternatively be denoted V (for *velocity*). The parameter vector is often omitted despite the fact that it is always present.

time. If there were an explicit dependence of the model on time, the brackets of the right-hand side of (1.22) should include a t.

Since the dynamics arising from (1.22) can be rather complex a common approach is to decompose a complex system into simpler subsystems or *modules* (see for example [TCN03, SJ04]). Under the headings of *bifurcation analysis*, *phase-space analysis*, *stability analysis*, *reachability*, *observability*, *controllability*, and *realisation theory*, dynamic systems theory provides graphical and mathematical tools to predict the behavior of inter- and intra-cellular dynamics. The challenges for the modelling lie in the relatively large number of variables, in the inherent nonlinearity of interactions and the difficulties in getting quantitative time series data.

Graphical tools, such as phase plane and bifurcation analysis which help us to visualise the dynamic behavior of a system are restricted to systems in two or three dimensions. Even then, a mathematical analysis is focussing only on specific parts of the phase-plane and reduces in those areas the analysis of nonlinear dynamics to a linear systems theory. Although biological systems are for most cases driven by nonlinear interactions we must acknowledge the fact that mathematical tools are for most cases only tractable if we revert to linear systems theory. This does not mean that we treat a nonlinear system as a linear one. Let us illustrate this approach.

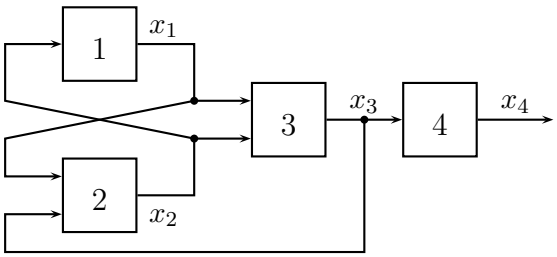


Figure 1.17: Decomposition of the pathway model (1.20) into subsystem described by (1.23).

As can be seen from (1.21) the nonlinear properties of pathways are mostly determined from simple interactions among the system variables. One intuitive approach is therefore to decompose the network (1.22) into subsystems

$$\dot{x}_i = f_i\big(x_i(t), \bar{x}_i(t), \Theta\big) \ , \tag{1.23}$$

where x_i is now a scalar and $\bar{x}_i$ is an input-vector consisting of the other state-variables x_k, $k \neq i$, of all other subsystems. Figure 1.17 illustrates the decomposition for the pathway model (1.20).

Many techniques in systems theory focus on the long-term behavior of a system, as $t \to \infty$. For example, an important question is the stability of the system. Stability theory helps us in this case to determine whether state variables remain within specified bounds as time proceeds. In dynamic pathway modelling we are however particularly interested in dynamic and short-term changes. For example, in cell signalling we wish to establish the response of a pathway to a stimulus of ligands binding to cell surface receptors. For signalling pathways, transient changes are of particular interest, while

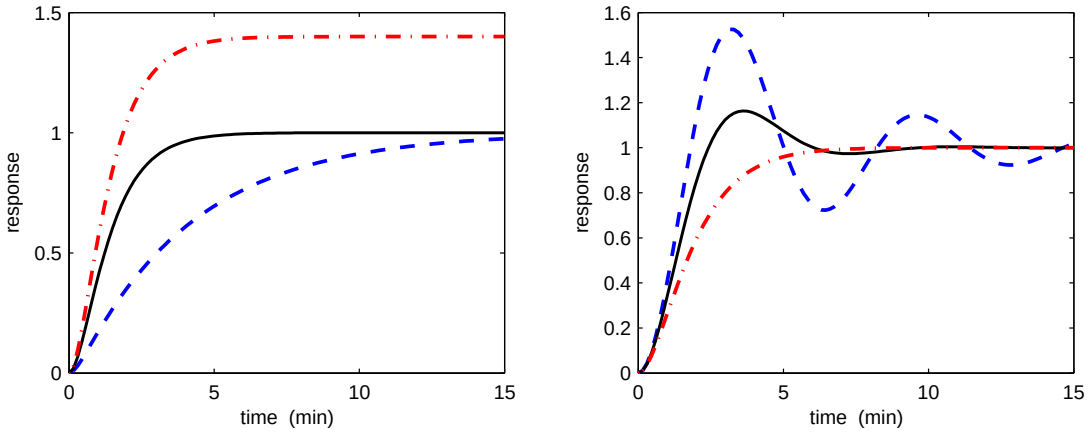


Figure 1.18: The two plots illustrate the qualitative analysis that is a main purpose of dynamic pathway modelling. The aim is to establish the consequences or causes of parameter changes and/or the removal/introduction of feedback loops. The plots show typical changes in form of an acceleration/deceleration, amplification/suppression of signals.

steady states are particular relevant for metabolic pathways. The plots in Figure 1.18 show typical responses encountered in dynamic pathway modelling. The plot shows on the left monotonic responses, while on the right there are damped oscillations. Another important case are sustained oscillations as they are known from the cell cycle, calcium and glycolytic systems. More recently, sustained oscillations are also discussed in the context of cell signalling. In Figure 1.18 the systems depicted reach a *steady state* after about 15 minutes. A steady-state is thus reached when there are no observable changes in the concentrations[14]. This observation corresponds to the mathematical condition where the rates of change on the left-hand side of the different equation are equal to zero:

$$0 = f\big(x(t), \Theta\big) \ . \tag{1.24}$$

critical points

For a given and fixed set of parameter values Θ the points in $X(t)$ for which this condition is met are called variously *steady-states*, *fixed points*, *critical points* or *equilibrium point*. This diversity in definitions is the result of various disciplines dealing with dynamic systems. For a new field like systems biology we thus have a choice of tools developed elsewhere but the different interpretations of the same mathematics may also be confusing to start with.

Stability analysis provides tools to characterise the dynamics of the system the transient behavior leading up to them and thereafter. It is then possible to predict whether the system will display cyclic changes of concentrations or whether they remain constant. Furthermore, it is possible to predict whether for small changes or perturbations the system remains stable with the critical point behavior or not. Bifurcation analysis is used to predict how system dynamics change as a function of parameter values or

[14] We are going to return to the question of steady-states and the assumptions involved throughout the text.

changes to them. For most practical purposes these analyzes are conducted by local linearisation of the nonlinear system around critical points obtained from (1.24) and then use well established tools from linear systems theory. Ignoring the subscripts denoting a subsystem in (1.23), we linearise the system (1.22) around the critical point x^* (now including all x_i and x_k) by considering small perturbations to the steady-state

$$x(t) = x^*(t) + \Delta x(t)$$

where powers $(\Delta x)^p$, $p > 1$, are considered "very small" compared to Δx. This is indicated by the notation

$$(\Delta x)^p \doteq o(\Delta x) \ .$$

A classical technique to approximate a function around a point is by means of a Taylor series. Assuming $f(\cdot)$ is sufficiently smooth such that derivatives exist, the Taylor series expansion around x^* is given by

$$f\big(x(t), \Theta\big) = f\big(x^*(t)\big) + Df_{x^*}\Delta x(t) + o(\Delta x) \ ,$$

where

$$Df_{x^*} = \left(\frac{\partial f_i\big(x(t), \Theta\big)}{\partial x_j}\right)\Bigg|_{x^*} , \quad \text{with } i, j = 1, 2, \ldots, n \ .$$

is the Jacobian matrix of first partial derivatives

$$Df_{x^*} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}$$

of the mapping

$$f = \big(f_1(x_1, \ldots, x_n), \ldots, f_n(x_1, \ldots, x_n)\big) \ .$$

The linearised system can now be written as

$$\Delta\dot{x} = Df_{x^*}\Delta x(t) \ , \tag{1.25}$$

where $\Delta x(t) = x(t) - x^*$ denotes deviations from the steady state. The Jacobian matrix is of particular interest. An off-diagonal element of the Jacobian describes the change of one variable, relative to another. Positive or negative entries correspond to *activation*, respectively *inhibition* of one variable by another[15]. If we denote the Jacobian matrix by A and consider the diagonal matrix $\tilde{A} = \mathrm{diag}(A)$, which contains the diagonal entries of A, then the decomposition of the linearised system (1.25) into one-component subsystems (1.23) is given by

activation
inhibition

$$\Delta\dot{x} = \tilde{A}\Delta x(t) + (A - \tilde{A})\Delta\bar{x}(t) \ ,$$

[15] The eigenvalues and eigenvectors of the Jacobian provide valuable information about the behavior of the dynamic system. There are many books available that describe this kind of analysis (e.g. [GH83, JS07, Str00a, HSD04]).

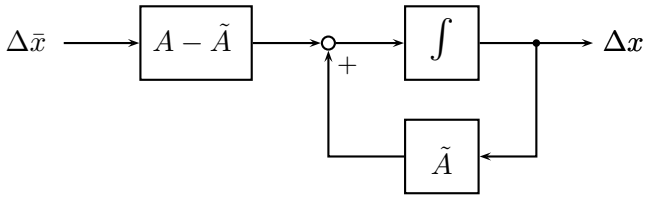
where the connections among linear subsystems are now denoted by

$$\Delta x(t) \doteq \Delta \bar{x}(t) \ .$$

In terms of individual subsystems this corresponds to

$$\Delta \dot{x}_i = a_{ii} \Delta x_i(t) + \sum_{j=1,\, j \neq i}^{n} a_{ij} \Delta x_j(t) \ ,$$

where the first term on the right-hand side corresponds to self-regulation and the last term corresponds feedback-type relations with other components. Interpreting $\Delta \bar{x}$ as an input and Δx as an output, we can study the behavior of the entire system, around the critical point x^*, using well established tools from systems and control theory (e.g. [SJ04]). In this setting the system can be represented by the following block-diagram:



Let us summarise what has been said so far. Beginning with the very simple biochemical reaction of proteolysis we introduced differential equations for modelling changes of molecular concentrations in the cell. We showed how sets of differential equations can be generalised and the dynamics analyzed using dynamic systems theory. From this brief introduction we can summarise the primary tasks in dynamic pathway modelling:

1. *Realisation Theory:* Characterise model structures that could realise given stimulus-response data sets.

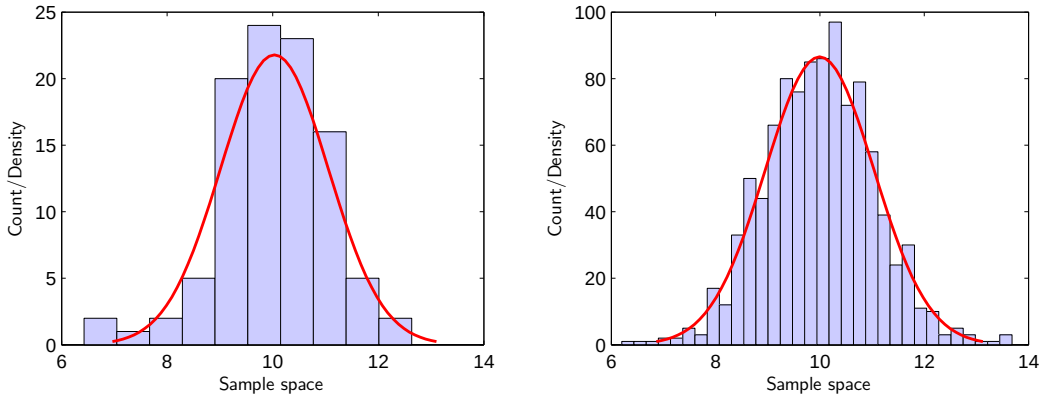
2. *System Identification:* Determine values for model parameters; using experimental data or simulation studies.

3. *Control Analysis:* Predict the consequence of changes to a pathway; in particular modifications to parameters, introduction and removal of feedback loops.

Figure 1.18 illustrated typical time course data and the kind of qualitative analysis the experimentalist is interested in. The main aim is to determine the causes of changes to parameters and the removal or introduction of feedback loops. The analysis is qualitative in the sense that exact values of the curves do not matter. More often we are interested in whether a response is 'accelerated' or 'decelerated', whether a signal is suppressed or amplified.

1.1.5 Dealing with uncertainty

Uncertainty in modelling arises for various reasons, including randomness in the biological process and measurement noise from observations. For most practical situations we



(a) (b)

Figure 1.19: As the sample size n, i.e., the number of repeated experiments increases, the shape of the histogram approaches a distribution which changes little and approximate the probability density function (solid line) from the sample values were drawn: On Figure (a) a sample of 100 values is used and 1000 for the plot on Figure (b)

cannot expect the measured data to match the model perfectly. In other words, observations are usually subject to *random variations*. If we were to repeat the experiment, for a particular point in time t, we would obtain a set, called the *sample* of measurements. The purpose of statistics is to characterise this sample. The most intuitive approach to investigate a sample of data from a repeated random experiment is to visualise the *distribution* of the data in the *sample space*. Such a plot is called a *histogram*. In Figure 1.19 we show a frequency histogram. Dividing the heights of the bars by the total number of elements we obtain the *relative frequency histogram*. Dividing the relative frequency by the bar width, the total area sum of all areas the bars equals 1. This is then called *relative frequency density*. If the sample size is increased and the bar width reduced, the relative frequency density function approaches gradually a curve, called *probability density function*, denoted $p(x)$, where we used the letter x_i to denote an element of the sample space X. The area under the probability density function is equal to one, i.e., the probability that *any* value will occur is one. Note that a probability density function is a model of what we observe through statistics. We can therefore abstract from a sample to obtain a stochastic model of the process that is underlying the data. Instead of the histogram one could characterise the sample (statistical model) or experiment (stochastic model) by some *characteristic* properties describing effectively the curve drawn by the histogram. For this we first consider the repeated data as referring to some real world variable (e.g. count of molecules at time t). If the measurement or observation is subject to random variations, it would make sense to speak of a random variable, say x. A description of a tendency for the data to cluster around a particular point, is called the *mean value*. From a statistical sample $\{x_i\}$ with n elements, the mean is estimated as the *sample average*:

sample
histogram
density function
mean value
sample average

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \ .$$

variance

Similar, a measure of variability around the mean value is obtained by the *variance*. The variance is an average distance to the mean value and given a sample, we can estimate the variance $\mathrm{Var}[x]$ as

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

where the $\hat{}$ is to denote the fact that this is an estimate. The squaring is necessary to avoid an influence of the sign of the difference on the average. The problem is then that if the values are a physical measurement with a unit, the variance would not be in this unit. This problem can be solved to taking the square root of the variance, leading to

standard deviation

what is known as the *standard deviation*

$$\mathrm{Std}[x] \doteq \sigma \ .$$

We should remind ourselves of the difference of the mean value and the sample mean or sample average. One has to do with a statistical experiment and the other with a stochastic model of the process that generates such data. In this sense, statistics is the real-world interface for probability theory. The mean value may also be considered an *expected value*, if we are to repeat the random experiment many times we would expect this value, on average. If our random experiment can be modelled by a probability density $p(x)$, where each possible value x is effectively *weighted* by the distribution or

expectation

density p, we could define the mean value as the expectation, $E[x]$, of random variable x

$$E[x] = \sum_{i=1}^{n} xp(x_i) \quad \text{if } x \text{ is discrete,}$$
$$E[x] = \int xp(x)dx \quad \text{if } x \text{ is continuous.}$$

Similar, the variance can be defined as an expectation

$$E[(x - \bar{x})^2] = \int (x - \bar{x})^2 p(x)dx \ .$$

covariance

For two random variables, x and y, the *covariance* is defined as

$$\sigma_{x,y} \doteq E[(x - \bar{x})(y - \bar{y})] \ .$$

If $\sigma_{x,y} = 0$, the two random variables are said to be *independent*. A bounded measure

correlation

of how two variables co-vary is the *correlation coefficient*

$$\rho_{x,y} \doteq \frac{\sigma_{x,y}}{\sigma_x \sigma_y} \ ,$$

such that $-1 \leq \rho \leq 1$. A positive correlation means that as one variable increases/decreases the other increases/decreases. In contrast, negative correlation means that as one variable increases/decreases, the other decreases/increases.

Let us now instead of individual elements of X consider an *event* or set $A \subset X$ defined by its characteristic map. If we extend the concept of an expectation

$$E[1_A] = \int 1_A(x)p(x)dx \ .$$

This in effect determines the likelihood or probability of event A, and noting that taking the interval over X, the characteristic map 1_A in effect restricts the interval to A

$$\begin{aligned} 1_A \ : \ X \ &\rightarrow \ \{0,1\} \\ x \ &\mapsto \ 1_A(x) = 1 \text{ if } x \in A, \text{ otherwise } =0 \ . \end{aligned}$$

This leads us to an intuitive definition of *probability*: probability

$$P(A) = E[1_A] = \int_A p(x)dx \ .$$

In terms of a sample of experimental data, it would seem plausible to define the probability then as the relative frequency:

$$P(A) = \frac{\text{number of outcomes in } X \text{ where } A \text{ occurs}}{\text{number of elements in } X} \ .$$

This is called the *relative frequency interpretation* for probabilities. There are however various other approaches to probability theory and it was the Russian mathematician Kolmogorov who put probability theory on a firm footing by linking probability theory to measure theory. The probability of an event A is then the measure of the area of the probability density function that overlaps with the subset $A \subset X$. For our purposes the mathematical details of these definitions are not of central importance and we refer to the vast literature on this. We do however note that a probability measure should satisfy the following *axioms*:

1. $0 \leq P(A) \leq 1$ for every event A.
2. $P(X) = 1$
3. $P(A \cup B) = P(A) + P(B)$, if A and B are mutually exclusive, i.e., $A \cap B = \emptyset$.

Remember that we started off by fixing a particular point in time and repeat an experiment. This generated a sample for which we now describe a stochastic model as illustrated in Figure 1.20.

A further refinement of our random experiment is to distinguish between the random mechanism and an observation. Denote the sample space of an experiment with random outcomes as the set Ω. This consists of possible individual elementary outcomes $\omega \in \Omega$. These outcomes are mutually exclusive, i.e., only any one of the possible outcomes can occur. A collection of elements of Ω is called a *random event* and is denoted $A \subset \Omega$. sample space random event

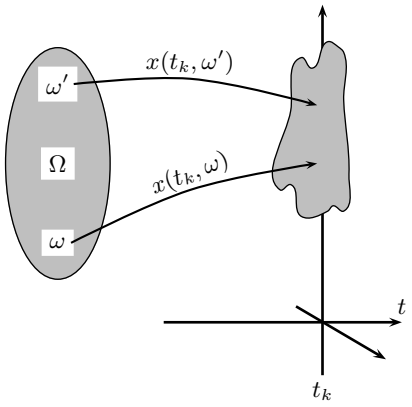


Figure 1.20: Stochastic process $x(t, \omega)$ as a t-dependent random variable.

probability space

random variable

We denote by $P(A)$ the probability that the event A will occur at each realisation of the experiment. The collection of events or subsets of Ω is mathematically defined as a σ-algebra and denoted $\mathbb{B}$. The triple $(\Omega, \mathbb{B}, P)$ of a sample space, sigma algebra and probability measure is then referred to as a *probability space*. The variable which is associated with the random experiment, for example the measurement of a protein concentration at a particular point in time, is referred to as a *random variable*. If Ω is a continuous set, x is referred to as a continuous random variable and if Ω is a set of discrete outcomes, we speak of a discrete random variable and discrete probability distribution. A random variable x is a real-valued map defined on Ω such that for each real number α, $A_\alpha = \{\omega \in \Omega \mid x(\omega) \leq \alpha\} \in \mathbb{B}$. A_α is an event for which the probability is defined in terms of P. A random variable is neither random, nor variable, it is simply the mapping

$$\begin{aligned} x \,:\, \Omega &\rightarrow X \\ \omega &\mapsto x(\omega)\ . \end{aligned}$$

Again we will drop the ω from $x(\omega)$ in most cases to simplify the notation, especially if x is a signal that is also a function of time. Since the experiment is associated with some random variable x, we write

$$p(\omega_i) = P(x = \omega_i) \text{ and } p(x) \text{ or } p_x \text{ for } P(x = \omega)\ ,$$

probability distribution

mass function

density function

where p denotes the *probability distribution*, *probability mass function*, or *probability density function*. We use the term 'mass function' for discrete sample spaces and density function for continuous sample spaces.

Talking of signals, considering a time set $I \subseteq \mathbb{Z}$, a time-varying process $x(t)$ is called a *random process* if for each t we cannot determine a precise value for $x(t)$, but instead have to consider a range of possible values with an associated probability distribution describing the relative likelihood of each possible value. More formally a *stochastic process* is a mathematical model of a random process, defined by the real-valued function

stochastic process

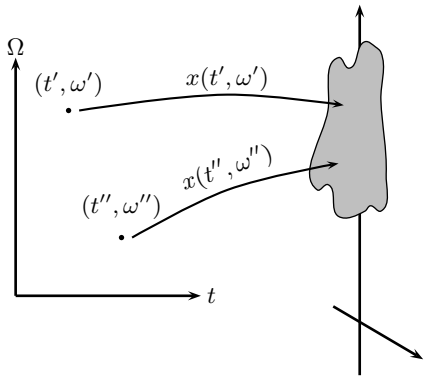


Figure 1.21: Stochastic process $x(t, \omega)$ as a joint function of t and ω.

$$x \colon I \times \Omega \to X$$

such that for each fixed $t \in I$, $x(t, \omega)$ is a random variable. A stochastic process (Figure 1.21) is subsequently a sequence of t-dependent random variables

$$\begin{aligned} x \; : \; I \times \Omega \; &\to \; X \\ (t, \omega) \; &\mapsto \; x_t \; . \end{aligned}$$

For each fixed $\omega \in \Omega$ the mapping from index set I into X describes a sequence of vectors $x_t(\omega)$, which is called a *realisation* or *sample function* of the process. More commonly, we refer to the realisation of a stochastic process is a *time series*, i.e., a sequence of observations and for which an observation at time t is modelled as the outcome of a random variable. The collection of all possible realisations is called the *ensemble*. All elements $x_t \equiv x(t)$ of a stochastic process $\{x(t)\}$ are defined on the same probability space. A stochastic process is in principle described by the joint distribution functions of all finite subcollections of $x(t)$'s but since these distributions will usually be unknown, most approaches will restrict themselves first and second order moments of the distributions, i.e., means, variances and co-variances.

realisation

time series

ensemble

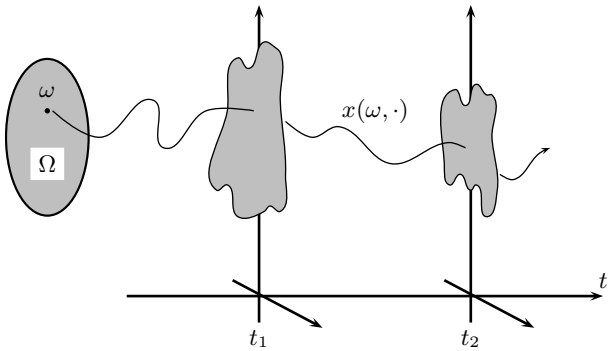


Figure 1.22: Realisation of the stochastic process $x(t, \omega)$, ω-dependent random variable.

We referred to a stochastic process as a model of a random process and should add

that a stochastic model can take various forms. A dynamic model that determines or predicts for any t a precise value $x(t)$ is called *deterministic* and any model that accounts for random variations is called *stochastic*. A differential equation model such as equation (1.15) is thus deterministic. If we however add an additive noise term to this ODE model, we could refer to this as a stochastic model. In the application of dynamic systems theory to molecular and cell biology one generally has to make a decision whether to regard the process as a deterministic nonlinear system but with a negligible stochastic component or to assume that the nonlinearity is only a small perturbation of an essentially stochastic process. A theory of nonlinear stochastic processes has so far not found many applications to time-series analysis. We find that there are a large number of alternatives and combinations and the choice to whether account for uncertainty in the model or not will depend on various issues, including convenience and realism amongst others. This discussion how to translate a biological process into a mathematical model makes applied mathematics so interesting. The reader should avoid getting confused by the many alternatives and assumptions one can consider and see them as part of the *art* of modelling. The great artist Pablo Picasso once said "Art is a lie that makes us realise the truth", to which we might add that science is the art that makes us realise reality.

The present section introduced the basic toolkit to describe systems as sets of objects with relationships defined on them. For the rest of this text all we do is to further refine and extend the concepts introduced here without actually adding more fundamental concepts than those introduced here.

1.2 Dynamical Analysis

1.2.1 Overview - Dynamical Analysis

Under the heading of "modelling" in the systems biology workflow various subtasks and types of analyses are included. We here summarise key concepts in a concise manner, providing links to the literature for further consideration. While some concepts and definitions are of general nature, a number of techniques are developed for mathematical models based on ordinary differential equations.

First the observable properties of a system have to be identified. In a second step system (state) variables are defined, parameters, input and output variables. Given a large-scale network, various genomics and bioinformatics tools can be used to support the identification of a suitable (sub)system (network or pathway) to focus on a more detailed or refined analysis.

For the following overview we assume for most concepts a mathematical model based on differential equations. The goal of this survey is a summary of the most important types of analyses that are available for the investigation of dynamical systems.

1.2.2 Stability Analysis

A system is in a stable state when a small disturbance of that state at a given time only alters the state at all subsequent times by a corresponding small amount. If an

arbitrary small disturbance to the state of a system has the effect that it produces a large change in the state at all subsequent times, then the system is said to be in an unstable state. Stability and instability are qualitative features of solutions of differential equations. Stability and instability is defined in terms of fixed points and trajectories [Jef93, Per98, JS07, Str00b]. In systems biology, stability analysis is used as a tool to estimate boundaries of parameters for a given system. It is done in such a way that these parameters will keep the system in stable state. Stability analysis is an important tool used in bifurcation analysis and robustness.

Stability of fixed points [Jef93, Per98, JS07, Str00b]

Stable fixed point: fixed point x^* in the state space is said to be stable if all trajectories which at initial time are close enough to the x^* remain close to it as time approaches infinity (see Figure 1.23a).

Asymptotically stable fixed point: x^* is said to be asymptotically stable if all trajectories which at initial time are close enough to the x^* eventually approaches to fixed point x^* as time approaches infinity (see Figure 1.23b).

Unstable fixed point: A fixed point which is not stable is said to be unstable or repelling. Its geometrical interpretation will be: all trajectories which at initial time are close to x^* diverge from x^* as time approaches infinity. (see Figure 1.23c)

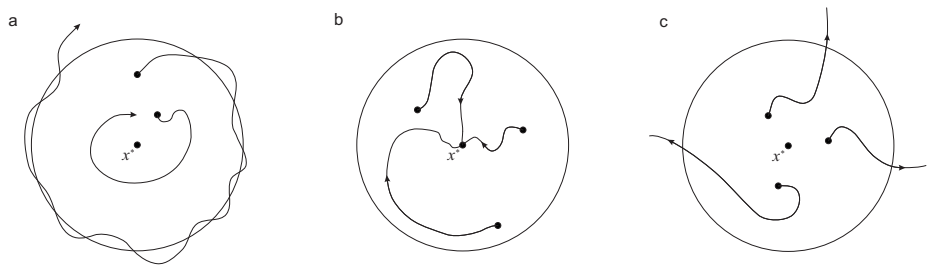


Figure 1.23: Stability of fixed points: a) stable fixed point x^*, b) asymptotically stable fixed point x^*, and c) unstable fixed point x^*.

Formal methods for stability analysis focus mostly on autonomous systems of differential equations $d\mathrm{x}/dt = f(\mathrm{x})$ where $\mathrm{x} = [x_1, x_2, \ldots, x_n]^T$ are the state-variables of the system and the mappings f encode the relationships between the state variables. The precise form of the mathematical terms for f depends on the chosen framework describing the considered biochemical processes like mass-action kinetics, power-law models, or Michaelis-Menten kinetics.

To investigate the behavior of trajectories in the neighborhood of a fixed point x^* linearization of differential equations about fixed points can be applied The eigenvalues of the corresponding Jacobian matrix (i.e. unstable states of linear equations are characterized by at least one eigenvalue with a positive real part) characterize the stability of the fixed point x^*.

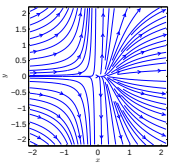
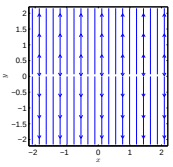
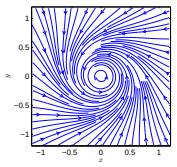
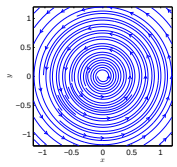
$\det(J) = 0$		purely imaginary eigenvalues	
$\frac{dx}{dt} = x^2$ $\frac{dy}{dt} = y$	$\frac{dx}{dt} = 0$ $\frac{dy}{dt} = y$	$\frac{dx}{dt} = -y - x^3 - xy^2$ $\frac{dy}{dt} = x - x^3 - x^2y$	$\frac{dx}{dt} = -y$ $\frac{dy}{dt} = x$
			

Table 1.1: Phase portraits of a system with $\det(J) = 0$ (left) and purely imaginary eigenvalues (right) and their around the fixed points linearized versions, respectively [Bet01].

Linearization: Given the system of differential equations $f(\mathrm{x}, t)$, the first step is to calculate the fixed points x^* of the system such that $0 = f(x^*)$. Then, the corresponding Jacobian matrix J

$$\mathrm{J} = [J_{ij}] = \left[\frac{\partial f(x_i)}{\partial x_j}\right]_{x^*}, \quad i, j = 1, 2, \ldots, n \tag{1.26}$$

is calculated. The linear system $d\mathrm{x}/dt = \mathrm{Jx}$ with the Jacobian matrix J is called the linearization of $d\mathrm{x}/dt = f(\mathrm{x})$ at the fixed point x^* [Per98, Str00b, JS07]. The eigenvalues λ of the Jacobian matrix J at the fixed point x^* given by the characteristic equation

$$\det(\mathrm{J} - \lambda \mathrm{I}) = 0,$$

where I is the identity matrix, are then used to characterize the stability of the fixed point P. An extensive discussion is given in Chapter 4.5, Phase-plane analysis, where this methods is applied to two-dimensional systems. This technique can only be applied in a small neighbourhood of the fixed point P [Jef93].

There are some exceptional cases where the linearization technique may fail namely when either $\det(J) = 0$ or when J has purely imaginary eigenvalues. In these cases, the phase portraits for the nonlinear and the linear system may or may not be similar. In Table 1.1 two examples illustrate the case when the two phase portraits do not resemble [Bet01, JS07].

Lyapunov/Liapunov stability theorem: If the linearization of a nonlinear system is not possible, the Lyapunov function is an alternative method to investigate the stability of a fixed point. Its existence guarantees the stability of the fixed point [Per98, Bet01].

Workflow:

1. Find a Lyapunov function $V(\mathrm{x})$ for fixed point x^* such that $V(x^*) = 0$ and $V(\mathrm{x}) > 0$ for $x \neq x^*$. (i.e. the Lyapunov function is not a unique function)

2. which additionally fulfills the condition

$$\frac{V(\mathrm{x})}{dt} = \left[\frac{\partial V(x_i)}{\partial x_j}\right] f(\mathrm{x}), \qquad i,j = 1,2,\ldots,n$$

3. Compute $\dot{V}(\mathrm{x})$
   a) If $\dot{V}(\mathrm{x}) \leq 0$ for all x, then x^* is stable
   b) If $\dot{V}(\mathrm{x}) < 0$ for all x, then x^* is asymptotically stable
   c) If $\dot{V}(\mathrm{x}) > 0$ for all x, then x^* is unstable

Biological interpretation of stability analysis: Bistability (when a system is able to exist in either of two steady states) can imply hysteresis, which is interesting for some regulatory mechanisms, e.g. sporulation in Bacteria, or irreversibility of apoptosis. Bistability can be an interesting and relevant phenomenon, in a single cell culture in which two distinct subpopulations of cells show discrete levels of gene expression [SVK07].

1.2.3 Network-Topological Analysis

A system described by ordinary differential equations can be locally characterized by a Jacobian matrix. This matrix can be visualized by a graph whose topology or structure gives hints to the underlying system dynamics. In contrast to stability analysis the focus here is not on the numerical values of the elements of the Jacobian but only their signs. This matrix is then referred to as the "interaction matrix" (interaction graph or digraph).

Chains of interactions, encoded by the elements of the interaction matrix, are referred to as "circuits" [Tho94, KST07] or "cycles" [MR07]. Their analysis provides conditions for qualitative instability, and a necessary condition for the existence of several stationary states [Tys75, Tho94, KST07, MR07].

We point out below some theorems which connect the structure of the graph with necessary conditions for oscillations, multistationarity/multistability, and other nontrivial behaviour.

Main ideas:

1. Tyson [Tys75] shows the necessary and sufficient conditions for "qualitative stability" of steady state solutions of nonlinear differential equations through the interaction description of the matrix. The main goal of the following theorem is to classify instability: The real irreducible Jacobian matrix A is qualitatively stable if and only if (i) $a_{ii} \leq 0 \ \forall \ i$; (ii) $a_{ii} \neq 0$ for some i; (iii) $a_{ij}\, a_{ji} \leq 0 \ \ \forall \ \ i \neg j$; (iv) $a_{ij}\, a_{jk} \ldots a_{pq}\, a_{qi} = 0$ for any sequence of three or more indices $i \neq j \neq k \neq \ldots \neq p \neq q$; (v) $det(A) \neq 0$.
2. Thomas et al. [Tho94, KST07] further elaborated the idea and resulted into the following major conjectures that are made about the behaviour of a dynamical system from Jacobian matrix to circuits and nuclei. Conjecture 1

(Thomas): The presence of a positive circuit (somewhere in phase space) is a necessary condition for multistationarity. Conjecture 2 (Kaufman): Multistationarity requires either the presence of a variable nucleus or else the presence of two nuclei of opposite signs. Conjecture 3 (Thomas): The presence of a negative circuit of length at least two (somewhere in phase space) is a necessary condition for stable periodicity. Conjecture 4 (Thomas): A chaotic dynamics requires both a positive and a negative circuit.

3. Mincheva and Roussel [MR07] applied and elaborated Ivanova's theorems (specifically stoichiometric network analysis [SNA]) in correlating the bifurcation structure of a mass-action model to the properties of bipartite graph. They continued by integrating some theorems mentioned above. In addition, they extended these graph-theoretical methods to delay-differential equation models. There is even a one-to-one correspondence between the nonzero terms in the determinant and all subgraphs of the bipartite graph. The subgraphs corresponding to negative terms in a coefficient of characteristic polynomial imply the structure in the graph responsible for the potential instability.

Tyson [Tys75] introduced the concept of representing each term of the Jacobian matrix by sign only; and he used it at the level of steady state. Wherein Thomas [Tho94, KST07] used it everywhere in state space; i.e. moving the steady state at will. They had shown in their studies that the presence of appropriate circuits (or unions of disjoint circuits) is a necessary (sufficient in particular) condition to generate steady states since only those terms of the Jacobian matrix that belong to a circuit are present in the characteristic equation. According to [MR07], the graph-theoretic condition for instability is more easily applicable to large biochemical networks since it obviates the need to compute a Jacobian. Note that some of the Jacobian elements are sums of positive and negative entries, making the signs of the weight functions undetermined. They show how these disadvantages of the digraph can be avoided by using bipartite graph.

1.2.4 Bifurcation Analysis

Bifurcation analysis is the study of qualitative changes in the behavior of dynamical systems under parameter variation. Qualitative changes here mean the appearance of new solutions of the differential equations where the behavior of the system changes significantly. Bifurcation analysis provides insight to stability changes of a system around a critical value (of a parameter) that determines such qualitative changes (referred to as "bifurcations"). One of the areas where bifurcation theory is widely applied, is physics, for example, quantum physics, dynamics of the nuclear reactor, electric power systems.

For changes to a single parameter, a nonlinear system is linearized around a fixed point (steady state). The linearised system is characterized by a Jacobian matrix whose eigenvalues λ will inform us of the local stability of the system, see also Stabililty Analysis 1.2.2. The stable dynamical system's eigenvalues have negative real parts.

Bifurcations can be classified into two groups:

Local bifurcation which can be detected by fixing any small neighborhood of the fixed point.

- at least one real eigenvalue crosses zero with a change of the parameter. There are some of such kind of bifurcations:
  - transcritical bifurcation
  - saddle-node bifurcation (fold, or turning-point, or blue sky bifurcation)
  - pitchfork bifurcation (supercritical, subcritical)
- at least one eigenvalue crosses the imaginary axis [FMWT02].
  - Andronov-Hopf bifurcation

Global bifurcation which are hard to detect because they involve large regions of the phase plane rather than just the neighborhood of a single fixed point. Some examples are:

- homoclinic bifurcation
- heteroclinic bifurcation

To determine bifurcations, stability analysis has to be performed first. Then to determine a type of bifurcation, that is how the behavior of a system changes as a parameter varies, different types of software might be used, for instance Auto 2000 [ol09a], CANDY/QA [ol09b, ol09c], MATCONT [DGKS04, DGKS06].

1.2.5 Reachability (of states) and Controllability (of systems)

A state of a system is reachable from another state if the system can be moved from the later to the former in a finite time by some input (or control signal) [Son98]. Note that the system may not necessarily stay in a reachable state.

A system is controllable if each state is reachable from every other state. Thus, a controllable system can be steered from any initial state to any other state in a finite time by a suitable input. Physically, this means that, directly or indirectly, the input independently influences all parts of the system [DTB97].

Uncontrollable systems have subsystems (e.g. state variables) that are unaffected by the input. This usually means that those subsystems of the system are physically disconnected from the input. Most of the theory and tools for reachability and controllability analysis are developed for linear systems. Algebraic tests for controllability including rank test and Grammian test have been developed over time [FPEN05] and implemented in MATLAB^{TM} Control Systems Toolbox [232].

Some authors define controllability and reachability of a system as its ability to reach the origin and an arbitrary state, respectively [KFA69]. In some texts the term output controllability is defined as the ability to steer the system output to all possible values in the outputs space [Son98, AS04].

1.2.6 Distinguishability (of states) and Observability (of systems)

Two states of a system are distinguishable if it is possible to differentiate between them on the basis of input/output experiments [Son98]. Thus the outputs resulting from the same input but two different initial conditions will be (same) different if the two states, used as the initial conditions, are (in)distinguishable.

A system is observable if every two distinct states of the system are distinguishable. For linear systems, the input has no role in determining observability. A linear system is observable if all the state variables can be reconstructed from measurements made at the system's output. In an observable system, directly or indirectly, all the state variables influence the output in one way or another [DTB97]. Unobservable systems have subsystems that have no influence on the output. Algebraic tests for observability have been developed over time [FPEN05] and are implemented in the MATLAB™ Control Systems Toolbox [232].

1.2.7 Robustness

Robustness is the property that allows a system to maintain its function against internal and external perturbations [Kit07, Hun09] or under conditions of uncertainty [SSS+04]. In general the robustness can be classified into "absolute robustness", representing the average functionality of the system under perturbation, and "relative robustness" quantifying the impact of perturbations on the nominal behaviour [RBFS09]. In biology the concept of robustness is closely related to the notions of "stability" and "homeostasis" [SSS+04]. While robustness is a general concept, homeostasis and stability are its particular instances which are identical if the function to be preserved is one maintaining the system state [Kit07]. Different strategies to design a robust system exist:

Redundancy: Alternative ways to carry out the function that the component performs [Kit07, SSS+04].

Feedback Control: Enables the system to regulate the output by monitoring it [SSS+04].

Modularity: Modules constitute semi-autonomous entities [Kit07, SSS+04].

Hierarchies and Protocols (Decoupling): Protocols encompass the set of rules underlying the efficient management of relationships. This decoupling isolates low-level variation from high-level functionalities [Kit07, SSS+04].

Kitano [Kit07] proposed a formal definition of robustness:

$$R^S_{a,P} = \int_P \Psi(p)\, D^S_a(p)\, dp \tag{1.27}$$

where R is the robustness of system S with regards to function a against a set of perturbations P, $\Psi(p)$ the probability for perturbations p to take place, $D(p)$ the evaluation function under perturbation p. A system S_1 can be said to be more robust than a system S_2 with regard to a function a against a certain set of perturbation Y when $R^{S_1}_{a,Y} > R^{S_2}_{a,Y}$.

Measuring the robustness of a system means to determine the behaviour as a function of the input [SSS+04]. An analysis of robustness in complex systems requires detailed specifications of functions, disturbances and robustness measures considered [SSS+04].

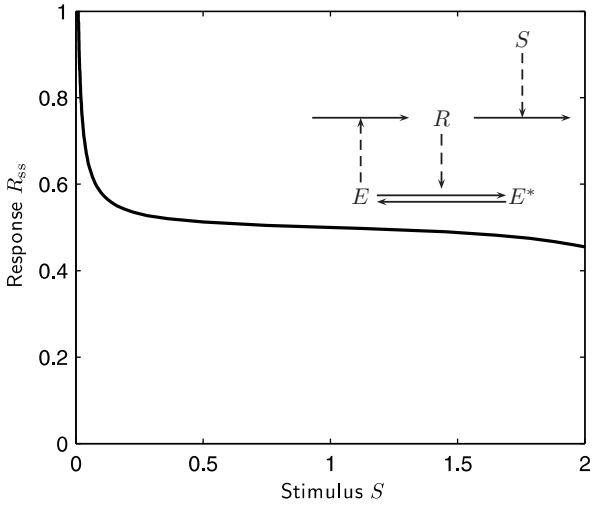


Figure 1.24: Stimulus-Response curve for a homeostatic system with negative feedback [TCN03]. The corresponding wire-diagram is shown as inset.

The coordination of physiological processes which maintain most of the steady states in organisms is called homeostasis [Kit07]. Using a negative feedback, the response counteracts the effect of the stimulus. One example is shown in Figures 1.24, described in [TCN03]. Homeostasis can be also described as stability through constancy with the purpose of physiological regulations to clamp each internal parameter at a set-point by sensing errors [Ste04]. In contrast, "allostasis" is stability through change. The goal of regulation is not constancy but rather fitness under natural selection [Ste04]. Figure 1.25 shows both types of robust systems in a condensed way.

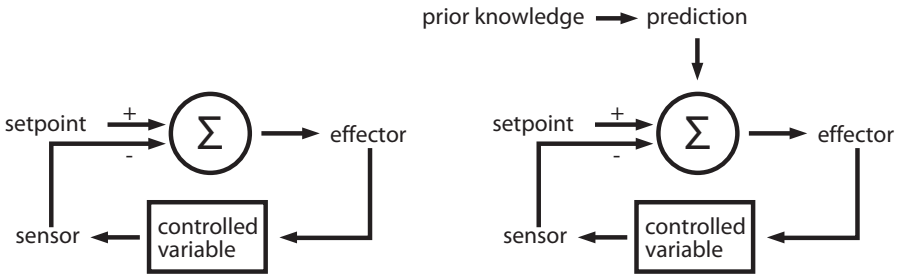


Figure 1.25: Homeostasis vs. allostasis model. Allostasis change the controlled variable by predicting, what level will be needed. [Ste04]

1.2.8 Sensitivity analysis

Sensitivity analysis is the study of how the variation in the output of a model can be apportioned, qualitatively or quantitatively, to parameter variations, and how the given model depends upon the information fed into it [SCS00]. In systems biology sensitivity analysis yields sensitivity values that tell us how a single parameter or a group of parameters affect the dynamics of a given variable, and these two pieces of information (parameter-variable) must accompany any sensitivity value. In general two different approaches can be distinguished, i) local and ii) global sensitivity analysis.

Local Sensitivity Analysis [SCS00]**:** One parameter at a time is varied within a small interval around a nominal value. This is repeated with the same perturbation for all parameters. The local sensitivity approach is practicable when the variation around the midpoint of the input factors is small; in general, the input-output relationship is assumed to be linear. When significant uncertainty exists in the input factors, the linear sensitivities alone are not likely to provide a reliable estimator of the output uncertainty in the model. When the model is nonlinear and various input variables are affected by uncertainties of different orders of magnitude, a global sensitivity method should be used.

A common approach to determine local sensitivities is the "finite difference approximation". This is done by changing one parameter k at a time where a 1% change is a good practical choice in most cases, see also Table 1.2, and measuring the change in a target variable Y. Other methods include the direct method which uses the Jacobian matrix J and parametric Jacobian matrix F, Table 1.2.

Global Sensitivity Analysis [SCS00]**:** The aim is to apportion the uncertainty in the output variable to the uncertainty in each input factor. Distributions that represent our degree of knowledge for each parameter provide the input for the analysis. A sensitivity analysis is considered global when a) all the parameters are varied simultaneously and b) the sensitivity is measured over the entire range of each input parameter. Global sensitivities are variance based where the test data is either generated with Monte Carlo methods (e.g. Sobol [SCS00]), or with a Fourier Amplitude sensitivity test (FAST). Another approach is the High Dimensional Model Representation (HDMR) by Rabitz and co-workers [FHC+04] which aims to facilitate analysis of systems with many parameters. The disadvantage of global sensitivity analysis is that the dynamics of the system, and therefore biologically significant behavior, is likely to change over large parameter perturbations. This means that sensitivities can change in different dynamic regimes and global sensitivities might therefore not reflect biologically significant sensitivities.

1.2.9 Model Reduction

The goal of model reduction is the simplification of analysis of multivariate systems with several parameters. During the procedure of model reduction the number of parameters

$S_k^Y = \frac{\Delta Y}{\Delta k}$	$S = -\frac{F}{J}$
Principle of finite difference approximation of local sensitivity analysis for variable Y and parameter k.	Direct method approach where F is the parametric Jacobian matrix and J the Jacobian matrix.

Table 1.2: Illustration of finite difference approximation and direct method.

and/or variables is reduced. This reduction leads to different dynamic behaviors of the model and the art is to eliminate those parameters/variables that maintain essential properties of the system. There are different routes that model reduction [SMDC08]:

1. lumping of variables
2. sensitivity/identifiability based model reduction
3. time scale separation.

Lumping: Lumping is mostly based on biological intuition of the process and describes the combination of different variables into one new variable. The drawback of lumping is that the new variables might not have a representation with the biological system, therefore biological credibility is lost.

Sensitivity/Identifiability approach: The variables can be separated into three classes: important, necessary and redundant. Redundant variable variables can be detected using the normalized Jacobian of the system:

$$\tilde{J}_i = \left(\sum_{j=1}^{N} \frac{y_i \, \partial f_j}{f_j \, \partial y_i} \right)^2$$

Redundant variable variables can be detected by considering only the N important variables and calculating $\tilde{J}_i$ which expresses the strength of direct effects of variables j on i-th variable. The variables j with the highest effects are called necessary variables and $\tilde{J}_i$ is determined for them. Those variables that remain are the redundant variables [SCS00].

Parameters for reduction can be determined using the normalized parametric Jacobian (cf. Sensitivity). A principal component analysis of the normalized parametric Jacobian, considering the important and necessary variables, reveals all parameters that can be eliminated from the model [SCS00]. It is also possible to reduce the mathematical rate laws using identifiability theory and principal component analysis. The resulting rate laws are reduced in their number of parameters and have an increased identifiability [SMDC08]. In general it is wrong that if the sensitivity of a parameter is small for all important variables, that this parameter could be

eliminated from the model [SCS00]. The reason is due to indirect effects: setting a parameter value to zero may change the trajectory of the system which will in turn extent to other variables.

Time scale separation: Some processes within a model are very fast allowing for the application of quasi-steady-state approximations or rapid equilibrium approximations. Very slow processes can be neglected leading to the occurrence of conservation laws. For both cases some of the differential equations are replaced by algebraic equations which can be used to express several dependent variables as functions of independent variables. The latter variables are determined as solutions of a smaller set of differential equations not containing the dependent variables [KH04].

1.2.10 Sloppiness/ Sloppy Parameter Sensitivity

Many multi-parameter models are 'poorly constrained' or 'ill conditioned', meaning that various parameter sets can exhibit the same dynamical behaviour. Such a poorly constrained model is called a 'sloppy' model [BS03]. The behaviour of a sloppy model is very insensitive to many parameter combinations, but also conversely very sensitive to some other parameter combinations.

This is of particular importance in systems biology models, where some or many of the model parameters are either unknown or significantly uncertain. Additionally, kinetic models of biological regulation exhibit renormalized interactions (simplified dynamics) and tentative topology (uncertain connectivity of proteins) [BS03]. Sloppiness is especially relevant to biology, because collective behaviour of most biological systems is much easier to measure in vivo than the values of individual parameters. Sloppy sensitivity spectra have been demonstrated in various systems biology models [GWC+07].

Workflow: The change in model behaviour is quantified as parameters θ varied from their published/measured values θ^* by the average square change in molecular species time course:

$$\chi^2(\theta) = \frac{1}{2\,N_c N_s} \sum_{s,c} \frac{1}{T_c} \int_0^{T_c} \left[\frac{y_{s,c}(\theta, t) - y_{s,c}(\theta^*, t)}{\sigma_s} \right]^2 dt\,,$$

a kind of continuous least-square fit of parameters θ to data simulated from published data θ^* (experimental measurements or fitted "Perfect Data"). $y_{s,c}(\theta, t)$ is the time course of molecular species s given parameters θ in condition c, and T_c is the "measurement" time for that condition. The species normalization σ_s is set to be equal to the maximum value of species s across the conditions considered. The sum runs over all molecular species in the model and over all experimental conditions considered for the model. The total number of conditions and species are denoted as N_c and N_s, respectively.

To analyze the model's sensitivity to parameter variation, the Hessian matrix corresponding to χ^2 is considered:

$$H^{\chi^2}_{j,k} = \frac{d^2\chi^2}{d(\log\theta_j)d(\log\theta_k)}\,,$$

The second derivative is calculated, since the first derivative will vanish for the best fit (minimal cost function). The derivative is taken with respect to $\log\theta$ to consider relative changes in parameter values, because biochemical parameters can have different units and widely varying scales. Analyzing H corresponds to approximating the surface of constant model behaviour deviation to be N_p dimensional ellipsoids. The principal axes of the ellipsoids are eigenvectors of H. The width of the ellipsoids along each principal axis is proportional to one over the square root of the corresponding eigenvalue. The narrowest axes are called "stiff" and the broadest axes "sloppy". In general, very few principal axes are aligned to the bare parameter axes; the ellipses are skewed from single parameter directions. The eigenvectors tend to involve significant components of many different parameters.

Definition of a sloppy model: In a sloppy model, the characteristic parameter sensitivities

- evenly span many decades
- are skewed from bare parameter axes

What causes sloppiness?

- Transformation in parameter space between the bare parameters natural in biology and the eigenvalues controlling system behaviour. This parameterization is fundamental to the system, not an artefact of the modelling process.
- Sloppiness depends not just upon the model, but also on the data it is fit to.
- Sloppiness may arise due to underdetermined systems, proximity to bifurcations, and separation of time and concentration scales.
- Sloppiness emerges from a redundancy between the effects of different parameter combinations.

Consequences of Sloppiness: Direct parameter measurements must be both precise and complete to usefully constrain predictions in sloppy systems. By contrast, collective parameter fitting in general yields tight predictions with only a modest number of experiments. Concrete predictions can be extracted from models long before their parameters are even roughly known and, when a system is not already well-understood, it can be more profitable to design experiments to directly improve predictions of interesting system behaviour rather than to improve estimates of parameters.

Robustness, see also Section 1.2.7, is often assumed to be an emergent evolved property, but the sloppiness natural to biochemical networks offers an alternative

non-adaptive explanation. Conversely, ideas developed to study evolvability in robust systems can be usefully extended to characterize sloppy systems [Dan08].

1.2.11 Identifiability of Parameters

Identifiability deals with the uniqueness of the parameterization for a given model [VR94]. This is an important question as identifiability is often assumed in systems biology models, and parameters are sought without first establishing whether these can be deduced from the set of measurements [AP09].

The parameters can be non-unique in two different ways. In the first case, there exist two or more sets of parameter values for which a model generates identical simulations for the observable quantities (deterministic non-identifiability). In the second case, two substantially different parameter values yield simulated trajectories for the observables (output variables) that are not identical but too close to be discriminated. (practical non-identifiability) [VR94].

For the problem of parameter identification for systems which are not in a special formalism, there exist no general solution to the problem [FFE+06]. Early ideas included the use of Taylor series, Generating series, Differential algebra and different transformations [WP96, WP97]. These approaches are restricted to low dimensional systems. Most today available approaches make use of experimental design and include statistical properties [FFE+06]. Solutions of global optimization problems arising from these approaches are given by numerical optimization algorithms [FFE+06]. An approach based on assumptions of optimality of biological systems has been proposed in 2005 [GGD05]. In [FFE+06], conditions that guarantee local identifiability of biochemical reaction networks are derived. In this work the problem of identifiability was approached by linearizing the model. In a recent work by August and Papachristodoulou an a priori method for establishing identifiability for nonlinear dynamical systems is described [AP09]. This approach uses a connections between parameter identifiability and state observability.

1.2.12 Distinguishability of Model Structures

Distinguishability questions the uniqueness of a model structure within a class of competing models [VR94].

Assuming two models, M and M' with two associated parameter vectors p and p', one can make the following definition: M will be structurally distinguishable from M' if, for almost any feasible value p of the parameters of M, there is no feasible value p' of the parameters of M' such that $M'(p') = M(p)$ [WP97].

It is easy to prove that the identifiability of two structures is neither necessary nor sufficient for their distinguishability [WP97].

In a practical or nondeterministic framework a mechanism is never unique, since it is always possible to postulate additional steps that involve unstable intermediates which occur rapidly enough so that their inclusion does not significantly alter the computed magnitudes of the observable species [VR94].

1.2.13 Generalized Modelling

In biological systems it is often difficult to identify a unique model parameterisation (cf. Identifiability of Parameters 1.2.11) or which processes one need to consider and specify (cf. Model Reduction 1.2.9 and Identifiability of Parameters 1.2.11). To avoid this specification, generalized models describe a class of systems exhibiting a similar structure. The processes that are taken into account are neither restricted to specific functional forms nor specific parameterisations [GF06].

The approach is based on a normalization procedure that is used to describe the natural parameter of the system. The Jacobian matrix in the normalized steady state is then derived as a function of these parameters. The eigenvalues of the Jacobian determine the stability of the system.

The investigation of generalized modelling enables us to generalize insights from specific models. In particular, generalized models can be used to investigate local dynamics, to identify important parameters and to draw conclusions on the global dynamics.

By investigating randomized parameterizations, this procedure allows us to find stable and unstable regions within the parameter space without specifying the considered processes. This approach of generalized models has already been applied to such diverse areas as socioeconomic, laser physics and ecological food webs [GF06]. Moreover, it has been used to analyze the stability and robustness of states in metabolic networks [STSB06, GSB+07], where generalized models are referred to as structural kinetic modelling. The underlying procedure however, is the same.

Workflow:

1. State and describe the interactions/general functions considered, e.g. $F(X)$
2. Formulate the system of ODEs using the general functions
3. Normalized the model
   a) assume there is at least one steady state
   b) denote the dynamical variables in steady state X^*
   c) define normalized state variables $x = X/X^*$, where $x^* = 1$
   d) define normalized general functions $f(x) = F(X^* x)/F^*$, where $f^* = 1$
4. Substitute the definitions in (3.c) and (3.d) into (2)
5. Consider the steady state of the system: $dx/dt = 0$
6. Define the scale parameters and substitute in (4)
7. Compute the Jacobian of the normalized model and thereby define exponent parameters
8. Rewrite the Jacobian in terms of scale and exponent parameters
9. Randomly set parameter values and calculate stability of each parameterization

10. Calculate correlation between every parameter and the stability of the system and thereby identify important control parameters in the system
11. Perform a bifurcation analysis for these parameters

Limitation: The scale and exponent parameters only capture the local behaviour of the generalized model. The bifurcation analysis is therefore limited to the investigation of local bifurcations. Nevertheless, this analysis enables us to draw certain conclusions on the global dynamics of the generalized model.

1.3 The Organisation of Living Cells

In this book, our aim is a study of how cell functions are governed by biochemical signals. The emphasis is on the "functional organisation" of the cell and the present section is to describe what this means. This survey is a mini review and for a comprehensive description one should consult the literature (e.g. [AJL+02]) or the Internet (e.g. `www.wikipedia.org`). A non-technical introduction is [Har01] and one designed for engineers and computer scientists is [TB04]. For the area of cell signalling, specialist literature include [Han97, Hel05, Gom03, Bec05]. We describe physical components and processes they are involved in. The emphasis of this text is very much on processes and physical objects and their properties are somewhat neglected. It should be here sufficient to know that a protein has a three-dimensional structure that influences the mechanisms by which it interacts with other molecules. How these proteins fold and their detailed chemical composition is largely ignored. One does not always get away with a high level of ignorance towards the biochemistry of intra- and inter-cellular processes but then this chapter is only to introduce the most important concepts needed to access the specialist literature.

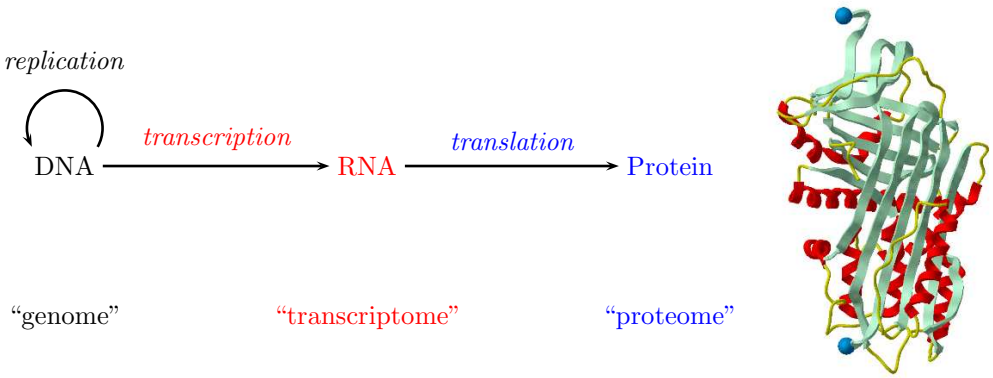


Figure 1.26: Information flow to generate a protein (protein synthesis). The picture on the right visualises the three-dimensional structure of a protein. The entire process is also referred to as *gene expression*.

macromolecule

The most fundamental physical object we are dealing with is this that of a (macro)molecule,

i.e., a chemical substance with a defined three-dimensional structure in which atoms are held together. While the composition and structure of a molecule largely determines how it can directly interact with other molecules, we are trying to avoid any detailed discussion of how these molecules are created. Molecules interact through the formation and breakage of chemical bonds. Two important binding processes are *covalent* and *non-covalent bonding/modification*. Covalent bonding is an intermolecular bonding characterised by the sharing of one or more pairs of electrons between two molecular species, producing a mutual attraction that holds the resultant molecule together. Such bonds are always stronger than the intermolecular *hydrogen bond*[16] and similar in strength to or stronger than the *ionic bond*[17]. There are four types of non-covalent binding processes that bring molecules together in a cell: ionic bonds (electrostatic attractions), hydrogen bonds, van der Waals attractions and hydrophobic force[18].

bonds

A *macromolecule* is a molecule with a large molecular mass. A *polymer* is a generic term used to describe a very long molecule consisting of structural units and repeating units connected by chemical bonds. The key feature that distinguishes polymers from other molecules is the repetition of many identical, similar, or complementary molecular subunits in these chains. These subunits, called *monomers*, are small molecules of lower molecular weight, linked to each other during a chemical reaction called *polymerisation*. There are various important kinds of molecules we are interested in, including nucleic acids, metabolites, proteins, enzymes, hormones.

A *nucleic acid* is a macromolecule that conveys genetic information. The most important nucleic acids are *deoxyribonucleic acid* (DNA) and *ribonucleic acid* (RNA). Nucleic acids are found in all living cells and viruses. DNA is made of four distinct monomers.

nucleic acid
RNA

Metabolites are the intermediates and products of metabolism. The term metabolite is usually used to refer to small molecules.

metabolite

Proteins are polymers composed of larger numbers of monomers that are called *amino acids*. In proteins, there are twenty different kinds of amino acids. The protein sequence of amino acid monomers largely determines the protein's shape and hence the interaction with other molecules. The process by which a protein is generated involves first transcribing (copying) the information of the DNA into an intermediate molecule RNA and then translating it into its final form (Figure 1.26). The process by which the information, made available in the genome, is facilitated in the cells function is also referred to as *gene expression* (see also Figure 1.27). In multicellular organisms each cell carries the same genome and yet cells can realise specialist functions in different tissues. The process by which cells of different types are generated is called *cell differentiation*.

protein

Most chemical reactions in biological cells do not happen by themselves or not as fast if enzymes were not involved. An *enzyme* is a protein that catalyzes (facilitates, speeds

enzyme

[16] A *hydrogen bond* is one mechanism by which parts of the same molecule are held together. It is an attractive force between two partial electric charges of opposite polarity.

[17] *Ionic bonds* are a type of chemical bond based on electrostatic forces between two oppositely charged *ions*. An *ion* is an atom or group of atoms with an electric charge.

[18] A *hydrophobic* molecule is repelled by water. They are not electrically polarised and unable to form hydrogen bonds. Thermodynamics favor hydrophobic molecules clustering together, even though hydrophobic molecules are not actually attracted to another.

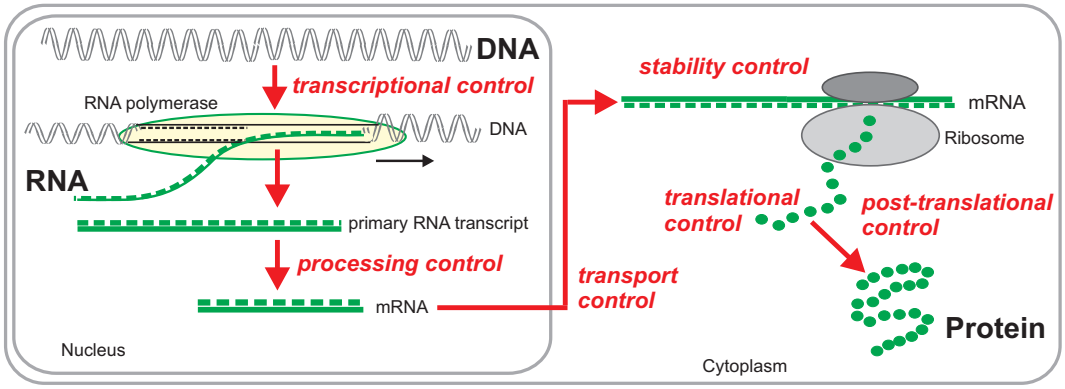


Figure 1.27: Gene expression is the process by which information in the genome is facilitated to realise the cell's functions. The entire process is regulated, controlled and coordinated to ensure a stable and fast operation that reacts sensitively to environmental cues.

mutation

catalyst

up) a chemical reaction. A malfunction (mutation, overproduction, underproduction or deletion) of a critical enzyme is a potential cause for the malfunctioning of the cell. Like all catalysts, enzymes work by lowering the activation energy of a reaction, thus allowing the reaction to proceed much faster (e.g. by a factor of millions).

The classic model of an enzyme-kinetic reaction is that a substrate combines with the enzyme to form an intermediate complex before releasing the desired product and the enzyme molecule. The enzyme is thus only temporarily involved and continues to function after it has been involved in a reaction. This leads to the frequent assumption of the total concentration, that is, the sum of free enzyme molecules and those bound in complexes is constant. Such assumption tends to simplify mathematical equations substantially.

effector molecules

inhibition

Enzyme activity is effected by temperature, pH levels and other *effector molecules*. An effector molecule is a regulatory or small molecule that can bind to a protein and alter its activity. One distinguishes between *inhibitors* and *activators*. The *active site* of an enzyme is the *binding site* where catalysis occurs. The structure and chemical properties of this active site allow the recognition and binding of the substrate. The active site in many enzymes can be inhibited or suppressed by the presence of another molecule. There are three primary modes of reversible inhibition. In *competitive inhibition*, the active site itself is blocked when a molecule chemically similar to the substrate binds to the active site but cannot be processed by the enzyme. In *noncompetitive inhibition*, the inhibitor binds to the enzyme at another site, the *allosteric site*, and this causes a structural change in the enzyme such that the active site is rendered useless. *Uncompetitive inhibition*, is similar to noncompetitive inhibition except that the inhibitor can only bind the enzyme-substrate complex rather than the free enzyme. *Allosteric control* is the process in which an enzyme's or protein's behavior is modified by binding an effector molecule at the protein's *allosteric site* (that is, a site other than the protein's active site). Effectors that enhance the protein's activity are referred to as *allosteric activators*, while those

that decrease the protein's activation are called *allosteric inhibitors*.

The two most important classes of cellular processes in which proteins are the basic physical objects are metabolism and signalling. *Metabolism* is the biochemical modification of chemical compounds in living organisms and cells. This includes the *biosynthesis* of complex organic molecules (*anabolism* – consuming energy and their breakdown (*catabolism* – generating energy. Metabolism usually consists of sequences of reaction steps, catalyzed by enzymes. Networks of such reaction steps are refereed to as *metabolic pathways*. To realise metabolic functions, the cell requires energy. In thermodynamics,*free energy* is a measure of the amount of work that can be extracted from a system. In this sense, it measures not the energy content of the system, but the 'useful energy' content. *Gibbs free energy* is an energy measure, calculated from the system's internal energy, the pressure, volume, temperature and entropy. Is the dissipation of Gibb's free energy through a cell's network of biochemical reactions zero, the system is in *thermodynamic equilibrium* and the cell dies. A living cell thus operates away from thermodynamic equilibrium.

biosynthesis

free energy

equilibrium

After we have introduced the most important physical objects and briefly mentioned metabolism, we now consider the physical structure of the cell before considering cell signalling as a key aspect of the cell's functional organisation.

1.3.1 The structural organisation of the cell

The most obvious aspect of the structural organisation of the (eucaryotic) cell (Figure 1.28) is given by the outer membrane and the inner membrane that defines the *nucleus*. While the *procaryotic cell* (microorganisms, bacteria etc. is characterised by only one compartment, the *eucaryotic cell* has the inner membrane that defines the nucleus. The nucleus of the cell contains the genetic material or *genome* in form of the double-stranded DNA molecule with its characteristic double helix structure. The genetic material is packed into *chromosomes*. A *gene* is a generic term to describe the role of information- and protein-coding regions in the genome. The medium between the nucleus and the outer membrane is the intracellular fluid *cytosol*. The area between the outer and inner membrane, including all of the components therein is called *cytoplasm*. The *cytoskeleton* is a meshwork providing structural support for the cell. As part of the cytoskelton, *microfilaments*, made of *actin*, provide mechanical support (and participate in some cell-cell or cell-matrix interactions), while *microtubules*, made of *tubulin* act as a transport system for molecules.

structural organisation

membrane

nucleus

genome

chromosome

cytosol

cytoskeleton

In addition to the two main compartments (nucleus and cytoplasm), eucaryotic cells have *organelles*, that are smaller compartments with a membrane and which contain a set of specific enzymes. Material can pass through the membranes directly or through gates. More specifically, there are three kinds of proteins that are embedded in the outer cell membrane and organelle membrane to allow material import and export:

organelles

pore

ion channel

transport

pump

> *Pores:* are made of pore-forming proteins, called *porins*. Pores allow small *hydrophilic molecules* (molecules that can transiently bond with water) to pass through the membrane of an organelle.

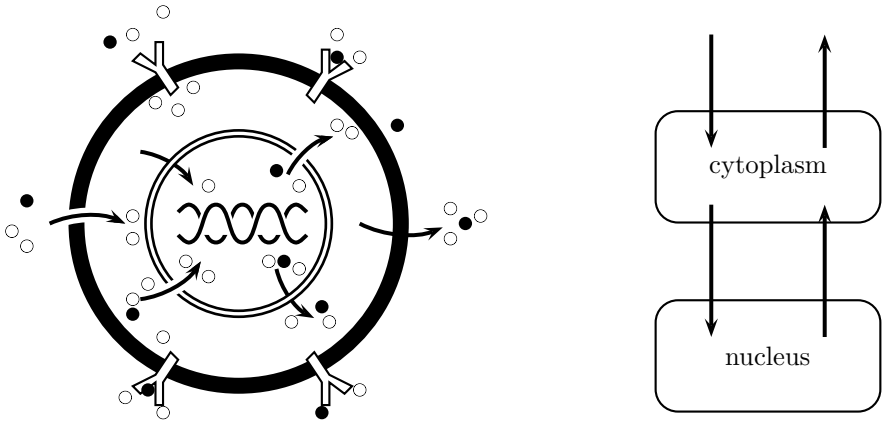


Figure 1.28: Structural organisation of the (eucaryotic) cell. The defining boundary structures are the outer membrane and the inner membrane that defines the nucleus. Material, signals and information can pass through the membranes directly, through gates, or through receptors. What the drawing does not show are two important structural elements: the cytoskeleton (providing structural support and transport mechanisms) and other organelles in the cytoplasm that fulfill specialised roles in the processing of proteins.

Ion Channels: are membrane proteins that allow specific inorganic ions (e.g. - Na^+, Potassium - K^+, Calcium - Ca^{2+}) to pass through membranes. *Ions* are electrically charged particles. For example, in nerve cells ion channels are responsible for electrical signaling, but they are also involved in muscle contraction, respiration, hormone release etc. Ions either diffuse passively in and out of the cell along a electrochemical gradient or they are actively transported by pumps.

Pumps: are proteins that actively transport ions and other molecules across cellular and intracellular membranes. Pumps can work against an electrochemical gradient.

The transport of molecules via pores, channels and pumps is obviously important for the normal functioning of the cell. The transport of molecules occurs by different mechanisms, which can be separated into those that do not consume energy in the form of ATP (*passive transport*) and those that do (*active transport*). *Adenosine triphosphate* (ATP) is the 'molecular currency' of intracellular energy transfer; that is, ATP is able to store and transport chemical energy within cells. ATP also plays an important role in the synthesis of nucleic acids and in signal transduction pathways, ATP is used to activate protein, which subsequently leads to the transmission of a signal. ATP *hydrolysis* is the reaction by which chemical energy that has been stored and transported in ATP is released to produce work. Metabolic processes primarily uses ATP hydrolysis, hydrolysis of *guanosine triphosphate* (GTP) is another important energy source that is used to drive the synthesis of peptides, and which plays an important role in cell signalling. GTP hydrolysis is catalyzed by enzymes known as GTPases.

ATP

hydrolysis

GTP

One distinguishes between *unicellular* and *multicellular* organisms (microorganisms or microbes being an example for the former, humans for the latter). However, even bacteria

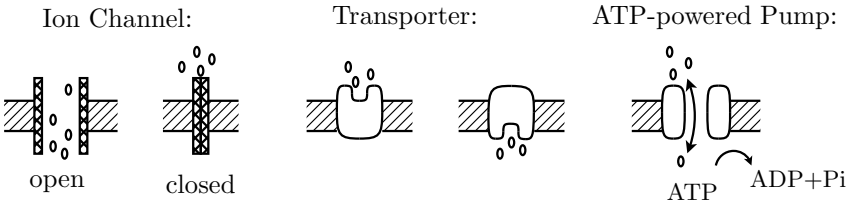


Figure 1.29: Illustration of three important information/material transducers: Channels, transporters and pumps.

form colonies in which cells communicate. A cell must therefore not be looked at in isolation but always be considered with respect to the environment in which it realises its function. In higher levels of structural organisation, say organs, the *extracellular matrix* (ECM) is any material part of a tissue that is not part of any cell. The extracellular matrix is the defining feature of connective tissue. The ECM's main component are various *glycoproteins* (macromolecules composed of a protein and a carbohydrate). In most animals, the most abundant glycoprotein in the ECM is collagen. The ECM also contains many other components, including proteins such as *fibrin* (a protein involved in blood clotting), *elastin* (a protein for the elasticity of connective skin tissue), minerals, or fluids such as blood plasma or serum with secreted free flowing *antigens* (a substance that stimulates an immune response, especially the production of antibodies[19]). The ECM's main role is to provide support and anchorage for the cells. An *integrin* is a receptor protein in the plasma membrane that transmits information (e.g. mechanical stimuli) from the ECM to the cytoskeleton.

extracellular matrix

This very brief overview of key structural aspects of the cell demonstrates that the cell is a highly structured and dynamic environment with materials continually moved around within the cell, between organelles as well as in and out of the cell. We next consider the functional organisation of these processes. In Figure 1.30 key components and concepts characterising the structural organisation of cells are summarised.

1.3.2 The functional organisation of the cell

Cell functions are an emergent, rather than an immanent or inherent, property of molecules. Although it arises from molecular interactions, it cannot be reduced to it. Structural organisation is usually easy to recognise through observation (e.g. microscopy). The functional organisation of a system is realised through molecules and constrained by the cell's structural organisation but is otherwise an abstract concept. The search for universal principles or mechanisms in molecular- and cell biology is to a large extent the study of functional organisation. In any particular context, the role of a cell may be to grow, to specialise, to divide, or to die. These processes all require the transfer of information. Signal transduction, cell signalling or simply signalling is the study of the principles by which this transfer of biological information comes about.

functional organisation

[19] An *antibody* is a protein used in the immune system to identify and neutralise foreign objects like bacteria and viruses.

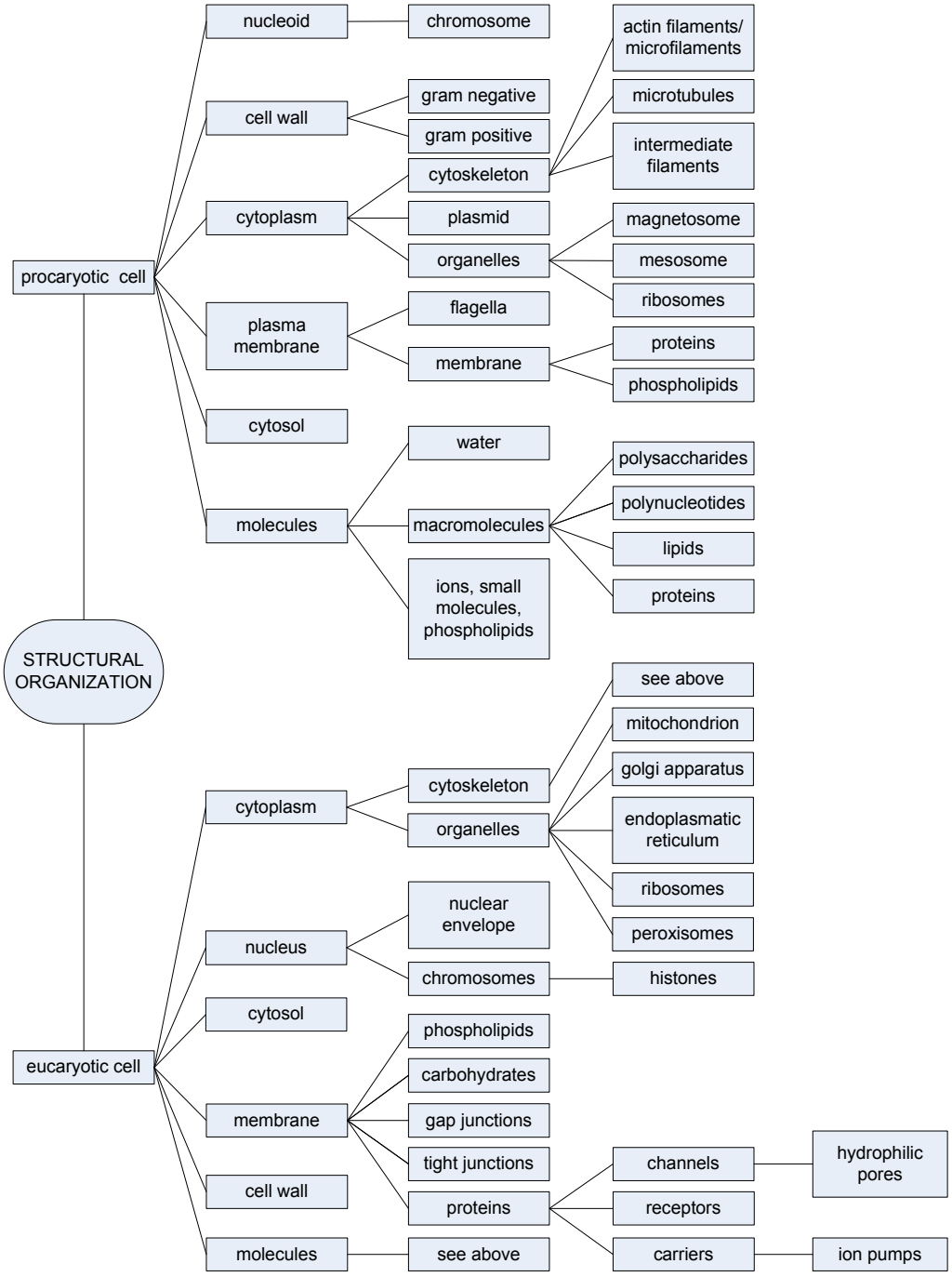


Figure 1.30: Key components and concepts characterising the structural organisation of cells.

intercellular signalling

Intercellular communication is accomplished by extracellular signalling and takes place in complex organisms that are composed of many cells. Intercellular signalling is subdivided into the following types:

- *Endocrine signals* are produced by *endocrine cells*[20] and travel through the blood to reach all parts of the body.
- *Paracrine signals* target only cells in the vicinity of the emitting cell (e.g. *neurotransmitters*[21]).
- *Autocrine signals* affect only cells that are of the same cell type as the emitting cell (e.g. in immune cells). Autocrine signalling coordinates decisions by groups of identical cells.
- *Juxtacrine signals* are transmitted along cell membranes via protein or lipid components integral to the membrane and are capable of affecting either the emitting cell or cells immediately adjacent.

Cells constantly receive and interpret signals from their environment (including other cells). Most of the molecules that enable signalling between the cells or tissues are known as *hormones*[22]. They realise extracellular signals. Another mechanism by which neighboring cells can share information are *gap junctions*. These junctions directly connect the cytoplasms of joined cells so that small intracellular signalling molecules (*mediators*) such as Ca^{2+} and cyclic AMP (cAMP) can pass through.

hormone
gap junction
mediator

Cell signalling is the basis for multicellularity and organises the cell's function, including *apoptosis* (programmed cell death), cell differentiation (e.g. specialisation of stem cells) and cell division. The functional organisation of cell signalling is realised through

- signals encoding extra-cellular balances and information:
  - physical ECM-to-cell and chemical cell-cell contacts related to adhesion, tissue formation and development.
  - transmission of information to regulate gene expression.
- signals encoding intra-cellular balances and information:
  - transmission of information to regulate cell growth and cell cycle.
  - homoeostasis of pH level, temperature and water imbalances.

A *receptor* is a protein on the cell membrane or within the cytoplasm or cell nucleus that binds to a specific molecule (a *ligand*), such as *neurotransmitters*, *hormones*, or other substances. There are various types of receptors, including:

receptor
ligand
neurotransmitter
hormone

- *Peripheral membrane proteins* that adhere only loosely to the biological membrane with which they are associated.
- Many hormone receptors and neurotransmitter receptors are *transmembrane proteins* that are embedded in the cell membrane, and which allow the activation of signal transduction pathways in response to the binding of extracellular molecules.

[20] The *endocrine system* links the brain to the organs that control body metabolism, growth and development, and reproduction.

[21] *Neurotransmitters* are chemicals that are used to relay, amplify and modulate electrical signals between a neuron and another cell.

[22] *Hormones* are simplify defined as chemical messengers from one cell to others.

- Another class of receptors are intracellular receptor proteins that can enter the cell nucleus and modulate gene expression in response to activation by a ligand.

The important role of receptors in transmitting information and effecting gene expression makes them an important target in the developments of drugs.

Transmembrane receptors are proteins that are inserted into the outer membrane of the cell, with one end of the receptor outside (extracellular domain) and one inside (intracellular domain) the cell. When the extracellular domain recognises a chemical messenger, the receptor undergoes a structural change that affects the intracellular domain, leading to further action. In this case the hormone itself does not pass through the outer membrane into the cell. This transfer of information is also referred to as *signal transduction*. There are however also transmembrane receptors that are ion channels. A ligand-activated ion channel will recognise its ligand, and then undergo a structural change that opens a gap (channel) in the outer membrane through which ions can pass. These ions will then relay the signal. An ion channel can also open when the receptor is activated by a change in cell potential, that is, the difference of the electrical charge on both sides of the membrane. If such a change occurs, the ion channel of the receptor will open and let ions pass through.

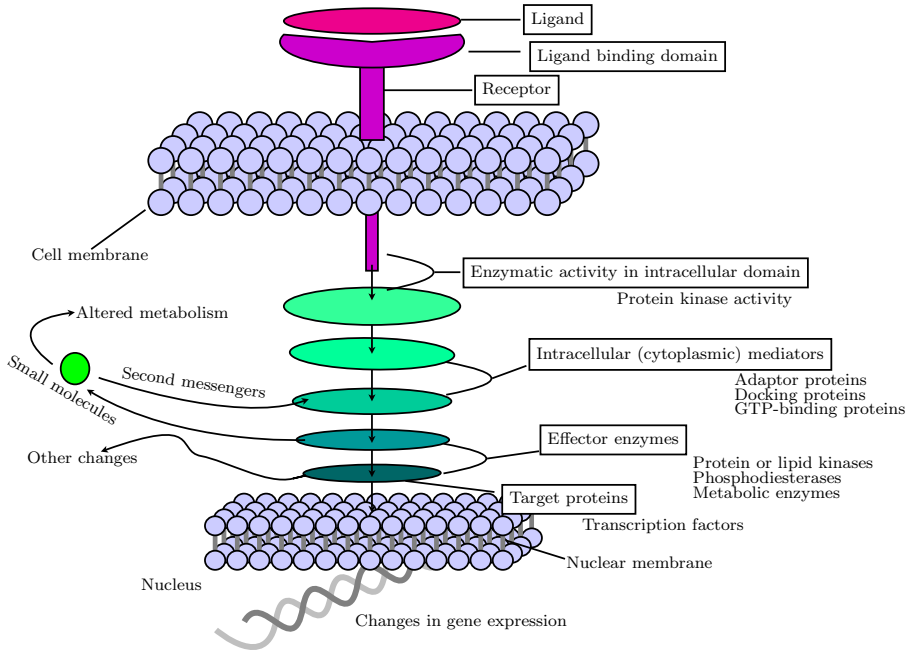


Figure 1.31: A general signalling pathway model, receiving signals in form of ligands that bind to the receptor. The binding leads to intracellular modifications and reactions that can inform or modify gene expression.

Figure 1.31 outlines a generic signalling network or pathway. Signalling starts with receptor proteins that are able to sense a change in the environment outside the cell. Transmembrane receptors are found at the cell surface, where they bind to extracellular

molecules that cannot penetrate the outer membrane. Receptors on the cell surface can bind to water-soluble signalling proteins (chemical messengers) such as *growth factors*[23] and peptide hormones which may be produced at a distant site in the body (and delivered in the bloodstream) or by neighboring cells. Peptide hormones are a class of peptides that are secreted into the blood stream and have endocrine functions.

growth factor

cytokine

As a result of ligands binding to receptors, a signal is transferred across the outer membrane. Ligand binding causes a change in the shape of the protein; this change is transmitted from the extracellular part of the receptor to the part inside the cell. Sometimes this involves the formation of *dimers*[24] of receptor molecules, i.e., two receptors bonded together. This tends to occur for receptor proteins with amino-acid chains that cross the membrane only once. For receptors that fold up so that they span the membrane several times, ligand binding may cause different parts of the molecule to reorientate themselves with respect to each other.

dimer

The stimulus to the receptor can lead to various subsequent reactions (activation of molecules) within the cell. As a consequence of signalling many enzymes and receptors are switched "on" or "off" by phosphorylation and dephosphorylation. Phosphorylation is catalyzed by various specific protein kinases, whereas phosphatases dephosphorylate. A *kinase* is a type of enzyme that transfers phosphate groups from high-energy donor molecules, such as ATP, to specific target molecules (substrates); the process is termed *phosphorylation*. An enzyme that removes phosphate groups from targets is known as a *phosphatase*. In cell signalling, the purpose of phosphorylation is to often 'activate' or 'energise' a molecule, increasing its energy so it is able to participate in a subsequent reaction with a negative free energy change. Figure 1.32 illustrates one possible illustration for the sequential activation of proteins in mitogen-activated protein kinase cascades. This type of signalling cascade is further discussed later on.

kinase

phosphorylation

The phosphorylation of substrate proteins affects their interaction with other molecules. Phosphorylated residues in a protein can act as binding sites for specific recognition domains in other proteins. A *domain* in a protein is a self-folding unit with a particular sequence and conformation, and certain domains allow proteins to recognise each other. So, as a result of phosphorylation protein complexes can assemble, resulting in changes in the localisation or activity of enzymes. Some intracellular (cytoplasmic) mediator proteins in these complexes, referred to as *adaptor proteins* or *docking proteins*, may work only to bring together other signalling molecules. Adaptor proteins tend to lack any intrinsic enzymatic activity themselves but instead mediate specific protein-protein interactions that drive the formation of protein complexes.

domain

adaptor protein

docking protein

Another aspect of *intracellular signalling* is realised through the activation of *small molecules* or *second messenger molecules*. Second messengers are low-weight diffusible molecules that are synthesised or released by specific enzymatic reactions, usually as

small molecules

[23] The term *growth factor* is sometimes used interchangeably with the term *cytokine*. Historically, cytokines were associated with hematopoietic (blood forming) cells and cells of the immune system (e.g. lymphocytes).

[24] A *dimer* refers to a molecule composed of two similar subunits or monomers linked together. It is a special case of a polymer. In a *homodimer* the two subunits are identical, while in a *heterodimer* they differ (despite being very similar in structure).

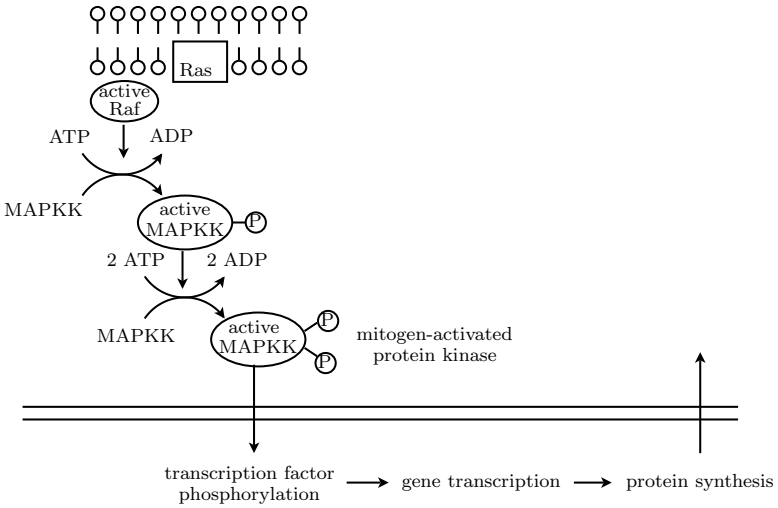


Figure 1.32: Illustration for the sequential activation of proteins in mitogen-activated protein kinase cascades. This picture is a refined depiction of what is shown in Figure 1.31. See also Section 4.8.8.

a result of an external signal that was received by a transmembrane receptor and pre-processed by other membrane-associated proteins. There are three basic types of second messenger molecules:

- *hydrophobic molecules* (repelled by water) are membrane-associated and regulate other membrane-associated proteins.
- *hydrophilic molecules* (e.g. Ca^{2+}) are water-soluble molecules that are located within the cytosol.
- gases, nitric oxide (NO) and carbon monoxide (CO), that can diffuse both through cytosol and across cellular membranes.

These intracellular messengers have some properties in common:

- They can be synthesised/released and broken down again in specific reactions by enzymes.
- Some (like Ca^{2+}) can be stored in special organelles and quickly released when needed.
- Their production/release and destruction can be localised, enabling the cell to limit space and time of signal activity.

cAMP

An important second messenger is *cyclic adenosine monophosphate* (cAMP, cyclic AMP) a molecule derived from adenosine triphosphate (ATP). cAMP is a second messenger, used for intracellular signal transduction, such as transferring the effects of hormones, which cannot get through the cell membrane. Its main purpose is the activation of protein kinases; it is also used to regulate the passage of Ca^{2+} through ion channels.

Responses triggered by signal transduction include changes in the gene-expression programme of the cell, the production of metabolic energy and cell locomotion, for example through remodelling of the cell skeleton. Genes are expressed as proteins, many of which are enzymes, *transcription factors*[25] or other regulators of metabolic activity. Because transcription factors can activate still more genes in turn, an initial stimulus can trigger via signal transduction the expression of entire set of genes and subsequently creating a physiological response of multicellular systems.

transcription factor

The activation of genes through signalling usually requires the movement (translocation) of a protein from the cytoplasm to the nucleus. Specific recognition systems ensure the import and export (shuttling) of proteins to and from the nucleus. These systems recognise sequence motifs[26] in the proteins, and the accessibility of the motifs may be altered as a result of phosphorylation or complex formation. A protein complex is a group of two or more associated proteins. Proteins can be made up of modular units, which belong to families of related structures. Modules may be either domains or smaller motifs, such as some phosphorylation sites. Modules can direct protein-protein interactions through their ability to interact with other modules. They may also have enzymatic activity, which in signalling proteins is often used to regulate other molecules. A large protein may contain several different modules, each of which behaves similarly to related modules in other proteins. A protein is thus made up of generic functional building blocks, which have been shuffled around during evolution to yield different combinations of interactions and activities.

translocation

sequence motif

The response of cells to activation of a particular signalling pathway depends on the strength of stimulus and the subsequent activation of proteins. Pathways can show graded responses - the stronger the activation of the intermediate proteins in the pathway, the stronger the final activity. In these cases, because different cells may show different sensitivities to a signal, low signal strengths might activate a subset of the responses that are activated by high signal strengths. In addition, some pathways work as on/off switches - once the signal strength rises above a certain level, positive feedback results in full activation of downstream targets. Such dynamic motifs are further discussed in later sections. The time course of a signalling pathway can also be critical. Transient activation of a pathway may have quite different effects to longterm activation. An understanding of cell signalling thus requires *dynamic* systems theory – and hence mathematical modelling and computational simulation.

What follows are illustrative examples of signalling mechanisms; there are many more described in the literature.

[25] A *transcription factor* is a protein that binds DNA at a specific promoter or enhancer region or site, where it regulates transcription. Transcription factors can be selectively activated or deactivated by other proteins, often as the final step in signal transduction.

[26] In general, a *sequence motif* is a nucleotide or amino-acid sequence pattern that has, or is conjectured to have, biological significance.

1.3.2.1 Signalling by G-protein-linked receptors

G-protein

An important mechanism of signalling is the binding of water-soluble hormones to transmembrane receptors of the *G-protein*-linked receptor family. Binding of the hormone agonist activates the receptor, which in turn activates the G-protein acting as the transducer. The G-protein dissembles upon activation, one of the subunits then interacts with the effector enzyme, which in turn catalyzes the production of cAMP second messenger molecules that relay the signal within the cell. The process by which G-protein-linked receptors function is illustrated in Figure 1.33.

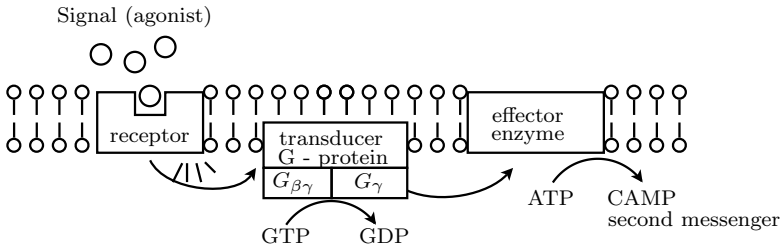


Figure 1.33: Signalling by G-protein-linked receptors as an example for the activation of signalling pathways.

1.3.2.2 Signalling by tyrosine kinase receptors

tyrosine kinase receptors

dimerisation

mitogen-activated protein kinase

epidermal growth factor receptor

The protein family of *receptor tyrosine kinases* (RTK) is another example of transmembrane signalling receptors. RTKs act as transducers of growth factor signals. These molecules posses kinase activity through a ligand dependent *dimerisation*[27].

Mitogen[28]-activated protein kinase (MAPK) pathways are an example of networks activated by RTKs. An example of a MAPK pathway is the Ras/Raf/MEK/ERK pathway. This kinase cascade controls cell differentiation and proliferation of various cell types. An example for a specific receptor is the *epidermal growth factor receptor* (EGFR). The textbook version of the signalling process describes it as a cascade (Figure 1.34) of signalling steps:

1. Ras is activated by growth factor receptors and binds to the Raf-1 kinase with high affinity if activated.
2. This induces the recruitment of Raf-1 from the cytosol to the cell membrane.
3. Activated Raf phosphorylates and activates MEK, a kinase that in turn phosphorylates and activates ERK.

[27] A *dimer* refers to a molecule composed of two similar subunits or monomers linked together. It is thus a special case of a polymer. The monomers will *dimerise*, or join together, upon the binding of a signal to the receptor of each monomer.

[28] A *mitogen* is a substance that causes a cell to begin dividing.

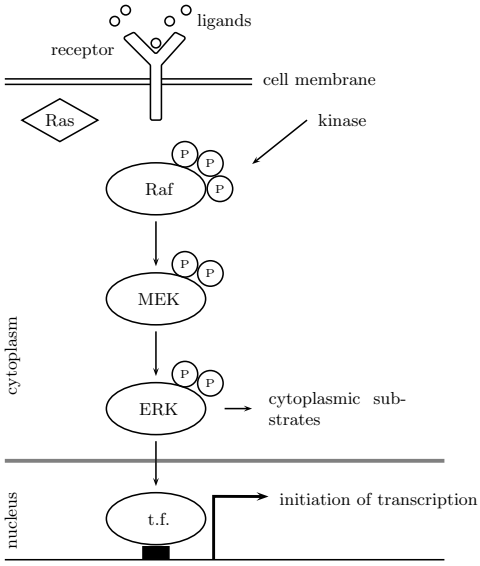


Figure 1.34: Textbook illustration of the Ras/Raf/MEK/ERK pathway as a representative of MAPK pathways. Drawing adapted from [Kol00]. This drawing is an example of the general MAPK cascade shown in Figure 1.32.

4. Activated ERK can translocate to the nucleus and regulate gene expression by phosphorylation of transcription factors.

1.3.2.3 Interaction- and pathway maps

The graphical representation of biochemical reaction networks leads to what are interchangeably called *pathway maps*, *process diagrams* or *molecular interaction maps* (MIM). There is at present no accepted standard on how to visualise reaction networks and we therefore do not argue for one or another format. The discussion of graphical representations and their translation into mathematical equations, respectively computational models is very helpful in classifying types of interactions and processes as described in the previous sections.

molecular interaction map

Noteworthy are the interaction map standard, originally devised by Kohn [Koh01, KAWP05] (for an example of a molecular interaction map of the mammalian cell cycle control and DNA repair systems see [Koh99]) and the CellDesigner software [Kit02]. In [OMFK05] this software was used to devise a comprehensive pathway map of epidermal growth factor (EGF) signaling. The map is described as a *state transition diagram* in which states of the system are represented by nodes and arcs describe state-transitions [KFMO05].

state transition diagram

In Kohn's MIMs, interactions between molecular species are shown by different types of connecting lines, distinguished by arrowheads and other terminal symbols. Molecular interactions are of two types: *reactions* and *contingencies*. Reactions include non-covalent binding and covalent bonds/modification, stoichiometric conversion[29], transcription and

stoichiometry

[29] *Stoichiometry* rests upon the conservation of mass and is often used to balance chemical equations;

degradation. Contingency arrow symbols describe stimulation, requirement inhibition and enzymatic catalysis. Reactions operate on molecular species, while contingencies operate on reactions on other contingencies.

There are 'elementary' and 'complex' molecular species, where the latter are combinations or modifications of elementary species. The *effect* of an interaction can be positive or negative. The net effect of a sequence of interactions is a *stimulation* if the number of negative effects is even, while the net effort is an *inhibition* if the number of negative effects in the sequence is odd. The consequence or *product* of an interaction is indicated by placing a small filled circle on the interaction line (not at the end). The consequence of binding between two molecules is the production of a dimer. The consequence of a *modification event* (e.g. through phosphorylation) is the production of the modified molecule. Covalent modification includes, amongst others, phosphorylation[30], *acetylation*[31] and *ubiquitination*[32]. Acetylation, phosphorylation and ubiquitination are also referred to as *posttranslational modification*, that is, a chemical modification of a protein after its translation.

phosphatase
acetylation
ubiquitination

1.3.2.4 Temporal aspects

Signalling processes take place in as little time as a few seconds (e.g. metabolites), few minutes (e.g. phosphorylation) or as long as a few hours (e.g. transcription). In Figure 1.35 the many different and wide ranging time scales of metabolic processes are summarised. An important decision in modelling is the choice of the appropriate spatial level at which to model. In systems biology the level is the cell and occasionally that of molecules. For time scales we can also make assumptions. If processes happen rather fast compared to the phenomena under investigation, one can assume these to be instantaneous. Similarly, if the level of a certain variable is rather large, changes to that variable may be somewhat small, suggesting that one might consider this variable as constant.

Not only are the processes within the cell highly dynamic, the cell itself is continuously changing. In mathematical modelling we frequently assume that the environment of a process is not changing so that model parameters are time invariant and the analysis of the dynamic behavior remains tractable. In eucaryotic cells all processes are however subordinate to the highly coordinated process of cell division or *cell cycle*. The phases of the cell cycle are:

cell cycle

- The G_0 *phase* is a period in the cell cycle where cells exist in a quiescent state, that is, a resting state of the cell in which it is not dividing.

it describes the molar proportions of components, that is, in chemical reaction diagrams it describes the number of molecules of a particular molecular species involved in that reaction.

[30] In eukaryotes, protein phosphorylation is an important regulatory event. Many enzymes and receptors are switched 'on' or 'off' by phosphorylation and dephosphorylation. Phosphorylation is catalyzed by various specific protein kinases, whereas phosphatases dephosphorylate.

[31] In living cells, *acetylation* occurs as a post-translational modification of proteins.

[32] *Ubiquitylation*, also termed ubiquitination, refers to the process particular to eukaryotes whereby a protein is post-translationally modified by covalent attachment of a small protein (ubiquitin). The ubiquitin's main function is to mark other proteins for destruction, known as proteolysis.

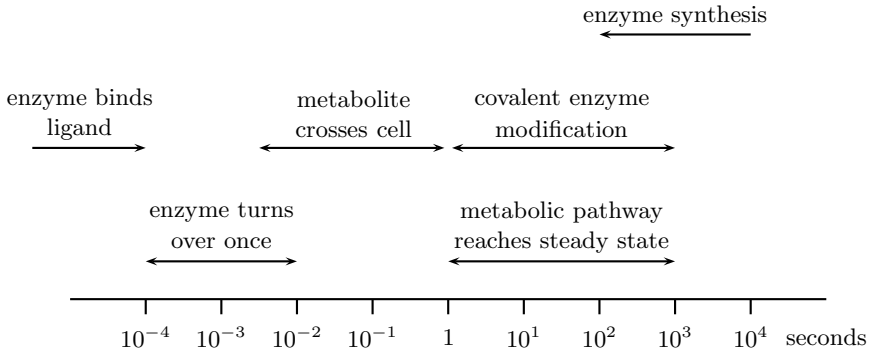


Figure 1.35: Processes in the cell occur at widely differing time scales. If a process is relatively fast, relative to the one we are considering, we may assume that changes in this process are instantaneous. If on the other hand a process is relatively slow to the one we consider, we may assume the variable to be constant. Picture adopted from [Fel97].

- The G_1 *phase* is the first growth phase. During this stage new organelles are being synthesised, resulting in great amount of protein biosynthesis. The metabolic rate of the cell will be high. G_1 consists of four subphases which may be affected by limiting growth factors, nutrient supply, and additional inhibiting factors. A cell may pause in the G_1 phase before entering the S phase and enter a state of dormancy (G_0 phase).
- *S phase*, during which the DNA is replicated, where S stands for the Synthesis of DNA.
- G_2 *phase* is the second growth phase, also the preparation phase for the cell.
- *M phase* or mitosis and cytokinesis, the actual division of the cell into two daughter cells. *Mitosis* is the process by which a cell separates its duplicated genome into two identical halves. It is generally followed immediately by *cytokinesis* which divides the cytoplasm and cell membrane. This results in two identical daughter cells with a roughly equal distribution of organelles and other cellular components.

The duration of a complete cell cycle is several hours. The cell cycle has been modelled in considerable detail, most notably by John Tyson and Béla Novák. A search for their names in literature databases provides a rich source for thorough and well written expositions of mathematical modelling in molecular and cell biology. Chapter 5 will discuss this topic further.

Figure 1.36 summarises key processes and concepts characterising the functional organisation of cells.

1.4 Cell Chemistry

The is the basic building block of which higher organisational levels such as tissues and organs and entire organisms are composed. This chapter is to review some basic

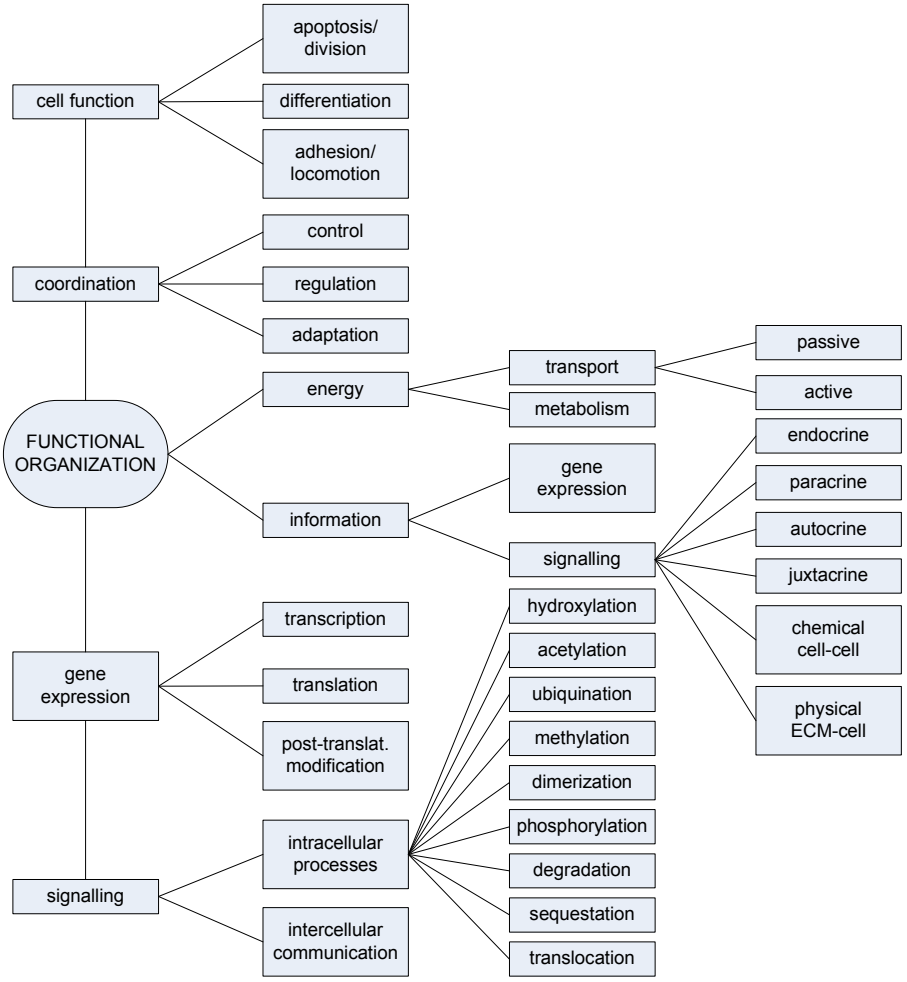


Figure 1.36: Key processes and concepts characterising the functional organisation of cells.

concepts from molecular- and cell biology. The text can however not even cover the bare minimum of the information and a comprehensive book such as the standard text [AJL+02] is strongly recommended as a reference.

The cell is a rather complex environment, consisting of many different components. Because cells are about 70% water, life depends mostly on aqueous chemical reactions[33]. These reactions occur between *molecules*, where a molecule is a cluster of *atoms*, held together by so called *covalent bonds*. The weight of a molecule is its mass relative to that of an hydrogen atom. The mass of a molecule is specified in *Daltons*, 1 Da being

covalent bonds

[33] There are alternative views that emphasise a gel-like character of the cell [Pol01]. The issue of what the inside of a cell is like should be important to us in modelling the interactions of molecules. In a somewhat brave act of modelling we later consider molecules as floating around as if they were in a gas.

an atomic mass unit approximately equal to the mass of a hydrogen atom.

$$\boxed{\text{moles} = \frac{\text{weight}}{\text{molecular weight}} \qquad \text{(a quantity)}}$$

One mole, 1 M, corresponds to $N_A \doteq 6.022 \cdot 10^{23}$ molecules of a given substance. N_A is referred to as the Avogadro's number. The molarity of a solution is defined by a concentration of 1 mole of the substance in 1 liter of solution:

$$\boxed{1 \text{ molar} \equiv 1M \equiv 1\frac{\text{mol}}{\text{L}} \qquad \text{(a concentration)}}$$

For example, 1 moles of glucose weights 180 g; a molar solution, denoted 1 M, of glucose has 180 g/L. If we dissolve 1 mol in 0.5 liters, we have 2 M solution, although the amount of substance is the same.

If molecules are clusters of atoms, held together by bonds, these bonds can be broken by violent collisions amongst molecules. Average thermal motion at normal temperatures does not break these bonds and thus the deliberate breaking and making of bonds is the fundamental process that determines the concentrations of chemical species in a reaction and subsequently cell function. This process requires energy to take place and is carefully controlled by highly specific catalysts, which in biological systems are called enzymes. How fast a reaction occurs is a matter of *kinetics*, defined by the rate of a reaction. In general, energy is the ability of a system to perform work, which is also why one speaks of *energetics* in this context.

kinetics

energetics

There are two principle types of biochemical reactions: *catabolic pathways*, breaking down foodstuff and thereby generating energy and smaller building blocks. Secondly, biosynthetic or *anabolic pathways* use energy to synthesise molecules. Both sets of reactions together constitute what is called the *metabolism* of the cell.

catabolic pathways

anabolic pathways

metabolism

Apart from water, nearly all molecules in a cell are based on carbon. Carbon-based compounds are used in the cell to construct macromolecules, including the nucleic acids (DNA, RNA), and proteins. Proteins are particularly versatile, having various roles in maintaining the function of a cell and the organism as a whole. Many proteins serve as *enzymes* that are catalysts that control kinetic (bond-breaking and -making) reactions. Other proteins are used to build the structural components that make up the cell, or they act as motors and produce force and movement. Enzymes catalyze reactions by binding one or more ligands which are also called substrates, and converting them into one or more chemically modified products, without changing themselves. Enzyme-catalyzed reactions happen faster by a factor of a million or more than a non-catalyzed reaction. They are therefore an important mechanism by which the cell can respond to changes and regulate its functions. A typical enzyme will catalyze the reaction of a thousand substrate molecules every second. The enzyme therefore requires sufficient amounts of substrate around it. The motion caused by collisions and thus heat energy ensures that molecules are rapidly moving about a confined area but can also move (diffuse) wider distances. The cell is a crowded environment and yet a small organic molecule can diffuse the entire distance across a cell in a fraction of a second.

enzymes

Enzymes move much more slowly than substrates, and the rate of encounter of each enzyme molecule with its substrate will depend on the concentration of the substrate molecule. For example, an abundant substrate may have a concentration of 0.5 mM and since water is 55 M, there is only about one such substrate molecule in the cell for every 10^5 water molecules. Nevertheless, an enzyme that could bind this substrate would collide with it about $500,000$ times a second.

The biological properties or function of a protein is determined by its physical interaction with other molecules. The substance that is bound by a protein is referred to as a

ligand

ligand for that protein. In cell signalling, ligands binding to membrane bound receptors provide the stimulus for intracellular reactions that transmit information to the genome of the cell. Antibodies, or immunoglobulins, are proteins produced by the immune sys-

antibody

tem in response to foreign molecules. A specific antibody binds tightly to its particular target (called an antigen), and thereby inactivates it. Antibodies can therefore be used in experiments to select and quantitate proteins. For example, considering a population of antibody molecules which suddenly encounter a population of ligands, diffusing in the fluid surrounding them. The frequent encounters of ligands and antibody will increase the formation (association) of antibody-ligand complexes. The population of such complexes will initially increase but eventually complexes will also break apart (dissociate). Eventually, a chemical equilibrium is reached in which the number of association events per second is equal to the number of dissociation events. From the concentrations of the ligand, antibody and the complex at equilibrium, one can calculate the equilibrium constant K_{eq} of the strength of binding. The same principle described here for antibodies, applies to any binding of molecules.

We are going to use capital letters to denote molecular species, e.g., A, B, ERK, MEK. A complex formed from proteins A and B is denoted either AB, $A-B$ or A/B. If the molecules are not referring to particular names, like A and B, we usually write AB for the complex. For known proteins, e.g., Ras* and Raf we write Ras*/Raf. In some cases the protein complex gets a separate name, e.g., for the MAPK/ERK complex we write MEK. Considering a reversible reaction $A + B \leftrightarrow (AB)$, for dissociation the reaction diagram is

$$(AB) \xrightarrow{k_d} A + B \ ,$$

dissociation rate

where the *dissociation rate* equals the product of k_d and the complex concentration (AB). Note that in the reaction diagrams the symbols denote molecular species while in mathematical equations we use square brackets to distinguish concentrations from counts of molecules. For the association of molecules,

$$A + B \xrightarrow{k_a} (AB) \ ,$$

association rate

the *association rate* is the product of k_a, A and B. At equilibrium,

$$k_a A \cdot B = k_d (AB)$$

equilibrium constant

and which leads us to the definition of the *equilibrium constant*

$$K_{\text{eq}} = \frac{(AB)}{A \cdot B} = \frac{k_a}{k_d} \ . \tag{1.28}$$

The equilibrium constant of this particular reaction has a unit of liters per mole. The larger the equilibrium constant, the stronger the binding between A and B. Taking an example which considers molecule numbers rather than concentrations [AJL+02], considering 1000 molecules of A and 1000 molecules of B, with concentration 10^{-9} M. For $K_{\text{eq}} = 10^{-10}$ of the reversible reaction

$$A + B \rightleftharpoons (AB) \ ,$$

there will be 270 A molecules, 270 B molecules and 730 (AB) molecules. For a reduction in binding energy of 2.8 kcal/M, reducing the equilibrium constant to $K_{\text{eq}} = 10^{-8}$, there will be 915 A molecules, 915 B molecules, and 85 AB molecules. For every, 1.4 kcal/M of free energy drop, the equilibrium constant increases by a factor of ten. Note that for the system to be in equilibrium, there is no flow of mass or material. In later sections we introduce the concept of *steady state*, for which changes in concentrations are zero. In dynamic systems theory, a steady state is sometimes also referred to as an equilibrium so that there is a risk for confusion. For the biochemist a biological system in equilibrium is dead.

1.5 Cell Signalling

For cells to combine into networks that realise higher levels of organisation, including for example tissue and organs, it is necessary for them to communicate, exchange information. The basis for this *inter*cellular signalling are the receptors in the cell membrane. The transmission of extracellular information to the genome is referred to *intra*cellular signalling. Inter- and intra-cellular information effects the transcription of information from the genome and the synthesis of proteins. For more comprehensive discussions of cell signalling see [Dow01, Han97]. The glossary on page 395 provides help with unfamiliar terminology.

The transmission of information is realised by chemical reaction networks, called pathways. Signals, passing these networks, are realised through changes in concentrations. The cell membrane and the nucleus in eucaryotic cells form physical barriers. There are principally two ways to pass these barriers - through active transport of molecules passing through the cell surface (e.g. via pores or gap junctions) or nucleus or via signal transduction, i.e., receptor stimulation and phosphorylation as a means to transmit information without the movement of molecules. We may refer to these two modes of signalling as the "radio" versus "courier" mode of signal transmission. The location of a signalling molecule within the cell affects the interaction with other proteins and hence the movement of molecules to different cellular locations, called *translocation*, influences the dynamics of a signalling pathway.

translocation

We are here going to focus on receptor-ligand signalling where extracellular molecules that bind to receptors in the cell membrane are referred to as *ligands*. Extracellular

ligand binding

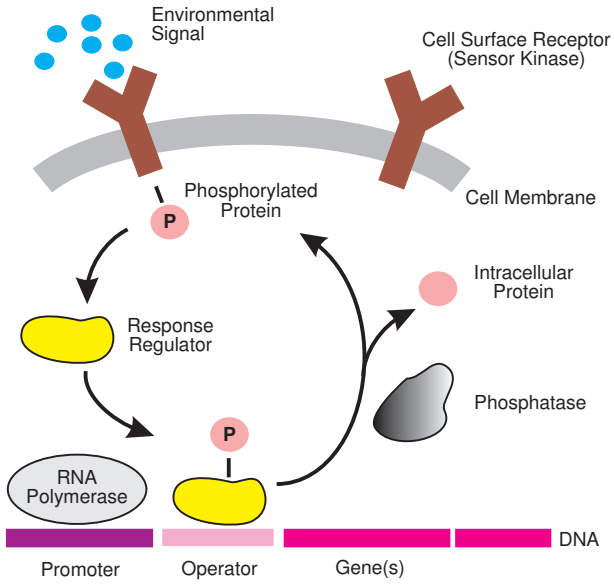


Figure 1.37: A drastic simplification of intra-cellular signalling. Extracellular stimulation of the receptors is transduced into the cytoplasm. A series of biochemical reactions transmits the signal towards the genome, where the transcription of genes can be affected as a consequence of receptor stimulation.

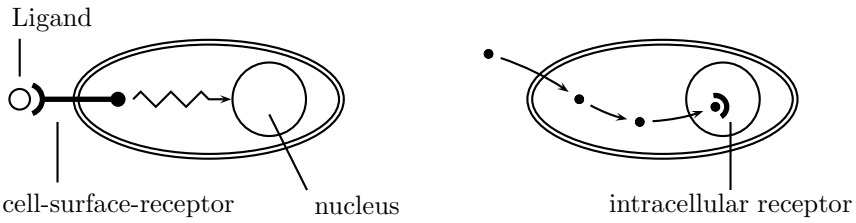


Figure 1.38: Left: Most commonly receptors are bound to the transmembrane, where they bind an extracellular signal molecule (Ligand). Right: Small signaling molecules can enter the cell where they activate receptors inside the cell.

signalling molecules include hormones, cytokines and growth factors. Usually extracellular signals are found at very low concentrations, in the order of 10^{-8} mol/L [Han97]. Binding to receptors is highly specific to particular ligands. Not all ligands that bind to receptors result in the activation of that receptor. Ligands that bind to receptors and thereby prevent activation, are called *antagonists*. Denoting the ligand with a 'L' and the receptor with 'R', ligand binding to monovalent receptors with only one binding site can be described as a reversible reaction

$$L + R \underset{k_2}{\overset{k_1}{\rightleftharpoons}} (LR) \ .$$

dissociation constant

The ratio of the concentrations, where 50% of the ligands are bound to receptors, is defined as via the *dissociation constant*

$$K_d = \frac{R \cdot L}{(LR)} = \frac{k_2}{k_1} \ .$$

The lower the K_d value, the higher the affinity of the receptor for its ligand. Generalising[34] the principle of *mass action* a mathematical model of monovalent receptor binding is given by the equation

$$\frac{d(LR)(t)}{dt} = k_1 L(t) R(t) - k_2 (LR)(t) \ ,$$

where k_1 ($\mathrm{M}^{-1}\mathrm{sec}^{-1}$) describes the rate constant of the receptor-ligand interaction and k_2 (sec^{-1}) describes the rate constant of the breakdown of the ligand/receptor complex LR. Solving this differential equation provides us with an equation for the temporal evolution of the ligand/receptor complex $(LR)(t)$. We got a bit ahead of ourselves here by jumping to this differential equation. In subsequent sections we are going to discuss ways to establish such differential equation models and compare it with alternative formulations[35]. We return to receptor modelling in Section 4.7.

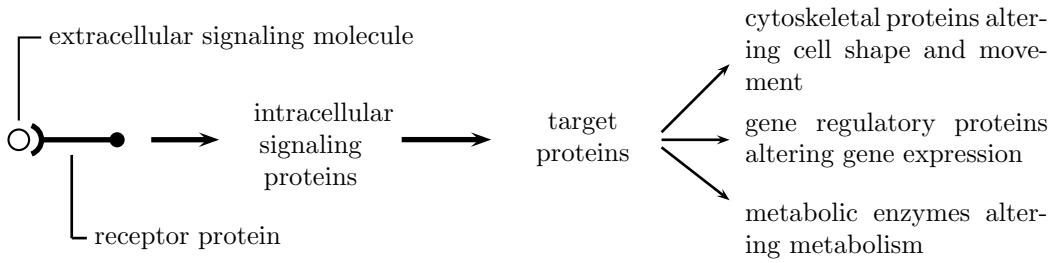


Figure 1.39: Basic molecular components involved in intracellular signaling.

The consequence of signalling through ligand-binding is in most cases a modification of the activity of intracellular enzymes or activation factors (e.g. transcription factors that determine the reading or *transcription* of information encoded in the genome). A change in enzyme activity is achieved through a change in its conformation (three-dimensional structure). The altered spatial arrangement of the active site (amino acids) reduces or increases the protein's catalytic action and binding to substrate.

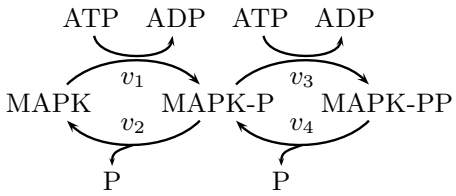
One of the most common ways to alter the spatial arrangement and hence the properties of a protein is by adding of one or more phosphate groups, a process known as *phosphorylation*. The enzymes that catalyze protein phosphorylation are known as *protein kinases* or kinases for short. The reverse process of dephosphorylation is catalyzed by phosphatases. Protein kinases and phosphatases are signalling molecules which catalyze the transfer of a phosphate group supplied by adenosine triphosphate (ATP) to

(de)phosphorylation

kinases, phosphatases

[34] The term 'generalised principle of mass action' indicates the fact that we are not strictly using this balance principle. For various reasons, including the indirect measurements in cell signalling, the differential equation models described here are in most cases *phenomenological* models rather than exact representations of physical interactions among molecules.

[35] Specifically for mathematical models of receptor binding the work of Lauffenburger and colleagues is notable [LL93].



$$\frac{d}{dt}\text{MAPK-P} = v_1 - v_3 - v_2 + v_4$$

Figure 1.40: Phosphorylation and dephosphorylation steps in the MAPK pathway.

and from target proteins respectively. The transfer ends with the release of adenosine diphosphate (ADP). The transfer of the phosphate group occurs only at specific *binding sites*, i.e., specific locations or amino acids of the target protein. The amino acids in the primary sequence of the polypeptide at which phosphorylation takes place are *serine*, *threonine*, and *tyrosine*. The kinases are grouped according to which amino acid they are specific for. Tyrosine protein kinases catalyze the phosphorylation of tyrosine residues, while serine/threosine protein kinases catalyze the phosphorylation of serine or threonine residues. Some protein kinases (such as MAPK) can act as both, tyrosine and serine/threosine kinases. Phosphorylated residues in a protein can act as binding sites for specific recognition domains in other proteins. A domain in a protein is a self-folding unit with a particular sequence and conformation. Certain domains allow proteins to recognise each other. Phosphorylation is thus a mechanism by which protein complexes can assemble. This results in a change of the localisation or activity of enzymes. Phosphorylation/dephosphorylation is a good regulatory mechanism since it can occur in under one second [Han97]. Furthermore, the activation of a single kinase molecule results in the phosphorylation of many enzymes and therefore result in an *amplification* of the intracellular signal.

binding sites

domains

amplification

While there are a vast number of proteins involved in signalling, many of the proteins are similar in the sense that they consist of components (domains, modules, motifs), some of which are found in many otherwise different protein molecules. At the amino acid sequence level, this similarity is expressed as *homology*. It is therefore in some sense not just the protein as such but particular aspects of it which determine its role in signalling. Because the properties of proteins and hence their ability to interact depends on whether they are in a phosphorylated or unphosphorylated state, in mathematical modelling we are going to introduce two variables for each of the states.

homology

In response to signals important cell functions are influenced. These include:

- cell death (apoptosis);
- cell growth (proliferation);
- specialisation (differentiation);
- stress response;

- cell cycle control.

Cell signalling is therefore of relevance to the development of an organism and the onset of disease. For example, cancer is a disease of uncontrolled cell proliferation.

1.6 Experimental Techniques

Molecular and cell biology has been to a large extend driven by the development of technologies. With the current speed at which new experimental techniques become available[36], one can expect major changes to our understanding of how the cell functions.

This section is intended to provide a few notes on methods for investigating cellular signal transduction experimentally. The motivation is that for mathematical modelling of signal transduction pathways it is important to appreciate the difficult and often indirect process by which biological information is obtained. Biological information relevant to mathematical modelling of signalling includes information about the interactions between proteins, information about the protein concentrations at a certain time point and/or in a certain cell component, and kinetic information about the dynamics of enzyme-catalysed reactions. Enzyme kinetics has been studied since the first half of the last century. Its theoretical aspects are thouroughly discussed in chapter 2 but in the following subsection 1.6.1 one experimental method of enzyme kinetic measurement is introduced. The subsequent subsection 1.6.2 outlines a method for protein identification and quantification, namely western blotting.

The given descriptions of experimental techniques are included to provide an appreciation of the difficulties in getting accurate and quantitative biological data. The books [AJL+02], [LC03], [VV04], and [KHK+05] provide basic introductions to these and various other experimental techniques[37].

1.6.1 Measurement of Enzyme Kinetics

Some proteins are enzymes, such as kinases and phosphatases, that catalyse the conversion of substrates into products. For example, for protein kinases, the substrates are proteins which are phosphorylated at specific phosphorylation sites. Metabolic enzymes catalyse the interconversion of smaller biochemical molecules, the metabolites. For the modelling of biochemical reaction kinetics as discussed in chapter 2, kinetic data is needed. Biochemical reactions and their properties can be observed and measured in an enzyme assay.

Exemplarily, the procedure for a continuous spectrophotometric enzyme assay is explained briefly. It exploits the property of NADPH (a currency metabolite such as ATP)

continuous assay

[36] We cannot do justice to the fullness of technologies involved and the reader is advised to consult the literature for further information. Major technological breakthroughs are reported in journals such as *Nature* and *Science*.

[37] For more comprehensive information on underlying principles and techniques of practical biochemistry we refer to [WW05].

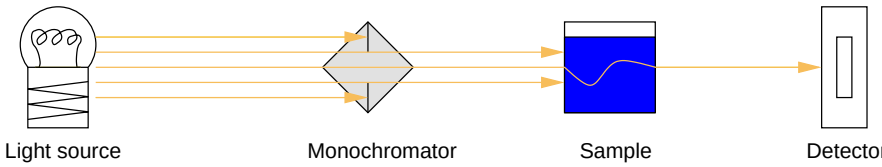


Figure 1.41: Scheme of the principle of photospectrometry. The light source emits a broad range of wavelengths but only light of a certain wavelength, e.g., $\lambda = 340$ nm, is selected by the monochromator for reaching the sample, behind which it is detected how much of the light was absorbed by the sample. Samples are typically placed in a transparent cell, known as a cuvette. Cuvettes are rectangular in shape, commonly with an internal width of 1 cm.

to absorbe light of the wave length 340 nm. The optical density or absorbance A is related to the concentration of NADPH as follows:

$$A = \alpha \cdot l \cdot [NADPH] \tag{1.29}$$

where α is the molar absorption coefficient of NADPH of 6220 $\text{cm}^{-1}M^{-1}$ and l is the path length, i.e., the distance that the light travels through the sample.

spectrophotometry

Therefore, changes of this metabolite in a sample solution can be monitored accurately with a spectrophotometer, a device for measuring light intensity as a function of the color, or more specifically, the wavelength of light (see Figure 1.41). Thus, the change of concentration of NADPH over time, for example in the reaction

$$\text{G6P} + \text{NADP}^+ \longrightarrow \text{glucono-1,5-lactone 6-phosphate} + \text{NADPH} + \text{H}^+$$

catalysed by glucose 6-phosphate dehydrogenase (G6PHD), can be studied directly. By adding different amounts of G6P to a solution containing the enzyme G6PHD a curve for enzyme activity as a function of substrate concentration can be measured. Other enzymes than G6PDH can be studied by coupling their reaction to the NADPH-producing reaction. For example, the reaction of phosphorylation of glucose catalysed by hexokinase

$$\text{Glucose} + \text{ATP} \longrightarrow \text{G6P} + \text{ADP}$$

can be assayed as follows. The assay cocktail contains hexokinase. The reaction is started by adding appropriate substrate solutions, glucose in this case. The assay cocktail also contains the enzyme glucose 6-phosphate dehydrogenase (G6PDH), and when G6P is formed via phosphorylation of glucose this leads to the reduction of NADP^+ to NADPH. The production of NADPH in the assay mixture can then be estimated with a spectrophotometer by monitoring the increase in absorbance at 340 nm as above. For this measurement, it is important to add G6PDH in abundance so that its activity is much higher than the activity of hexokinase, to ensure that the measured rate is of the hexokinase reaction. If the reaction one investigaets does not produce G6P directly but is only a few steps away from a G6P-producing reaction, then further coupling reactions can be introduced. An example is the formation of F6P in the fructose-phosphorylating reaction

$$\text{Fructose} + \text{ATP} \longrightarrow \text{F6P} + \text{ADP}$$

catalysed also by hexokinase. Adding phosphoglucoisomerase, an enzyme that interconverts G6P and F6P, to the assay cocktail channels the produced F6P to the G6PDH reaction which again can be monitored photospectrometrically. Again, it is necessary that the enzyme G6PDH and the enzymes that perform the coupling to the NAPDH-producing reaction are abundant such that the observed changes reflect the properties of the enzyme-catalysed reaction under investigation.

stopped assay

Unlike in the continuous enzyme assay, in the stopped enzyme assay, the studied reaction reaction is started in several cuvettes simultanously but then stopped (e.g., by adding a protein degrading agent) after different time intervals. Afterwards, measurements of the concentration of the product of the reaction are performed. This method provides a time series of concentrations.

Beside the spectrophotometric assays, there are other types of enzyme assays, e.g. fluorimetric, chromatographic, or radiometric assays. We again refer to the biochemistry textbooks for descriptions of these methods.

1.6.2 Protein Identification and Quantification by Western Blotting

The purpose of this section is to outline[38] a method of protein quantification and to indicate the difficulties in obtaining such data. The described method requires the sequential application of several techniques: protein separation by gel electrophoresis, transfer to a nitrocellulose membrane by electroblotting, immuno-probing, image processing, and, finally, image analysis.

1.6.2.1 Gel electrophoresis

gel electrophoresis

Macromolecules are commonly separated by *gel electrophoresis*. This method uses gels made of agarose for DNA or RNA, and polyacrylamide for proteins. The sample containing a mixture of macromolecules is loaded onto the gel by placing it into wells on one end of the gel, and then an electric field is applied that pulls the macromolecules through the gel via electromotive force. The gel acts like a sieve. Big proteins move slower than small proteins, and the proteins or nucleic acids are separated by size[39]. The resolution depends on the pore size of the gel and is only optimal for a certain size range. This is why agarose with wide pores is used for the large nucleic acids and polyacrylamide with smaller pores for the smaller proteins. Varying the concentration of polyacrylamide allows to adjust the size range for optimal resolution. Proteins of 200 Da - 70 kDa are well resolved by 7.5% polyacrylamide gels; proteins of 120 Da - 30 kDa are resolved on 10% polyacrylamide gels; and proteins of 50 Da - 10 kDa are resolved by 12.5% polyacrylamide gels.

Identification of proteins or assignment of a molecular weight is made possible by having one well in each gel reserved for a control solution, i.e., a solution containing the

[38] The text is based on notes kindly provided by Walter Kolch.

[39] In 2-D gel electrophoresis, another dimension for separation is added, e.g., by applying a pH gradient.

proteins in question or a solution containing proteins of known molecular weight. This provides marker bands for comparison (see Figure 1.45).

Gels can be used for protein purification or identification, e.g. by cutting out the area of the gel that holds the protein relevant for further investigation. For the purpose of quantification, however, the separated macromolecules need be immobilised by transferring them to a filter membrane.

1.6.2.2 Western blotting

Blotting

Blotting is the transfer of macromolecules from the gel onto a membrane, in order to make the separated macromolecules accessible, for instance, when other macromolecules are used as probes to specifically detect one of the separated macromolecules. Blotting can be done by various means. Traditionally, nucleic acids are blotted by capillary action as shown in Figure 1.42, where a stack of dry papertowels is used to draw the buffer from a tray at the bottom through the gel.

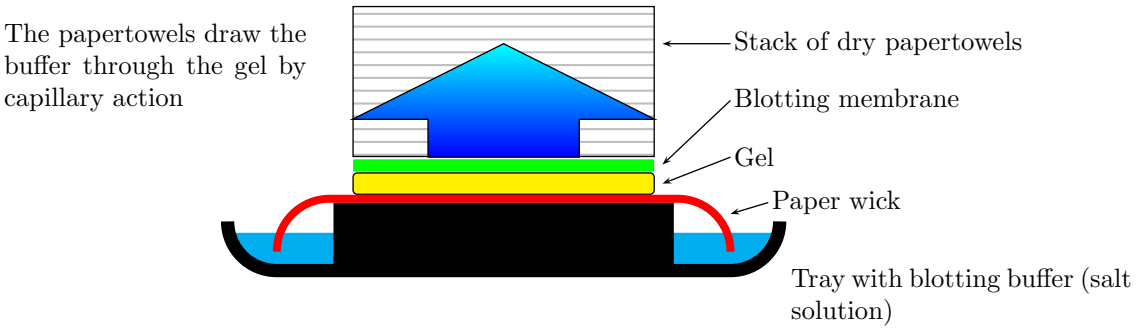


Figure 1.42: Blotting by capillary action.

Proteins are usually transferred onto a nitrocellulose membrane by electroblotting (see Figure 1.43). The protein-binding capacity of nitrocellulose membranes results from hydrophobic interactions, hydrogen bonding, and electrostatic interactions.

The membrane with DNA on it is called Southern blot[40]. If RNA is blotted, it is called a northern blot. If proteins are blotted, it is called a western blot.

1.6.2.3 Quantification of western blots

Proteins on the obtained blot are detected by immuno-probing. This is a two step process. First, the membrane is incubated with antibodies against the protein of interest (e.g. Ras), and washed several times. The antibody against the protein of interest

primary antibody

is called the *primary antibody*. The Ras antibody will bind to the immobilised Ras proteins on the membrane but will be washed off everywhere else. Next, the blot is

secondary antibody

incubated with a *secondary antibody*, and washed again several times afterwards. The

[40]The procedure is called Southern blotting as it was invented by Ed Southern, Oxford.

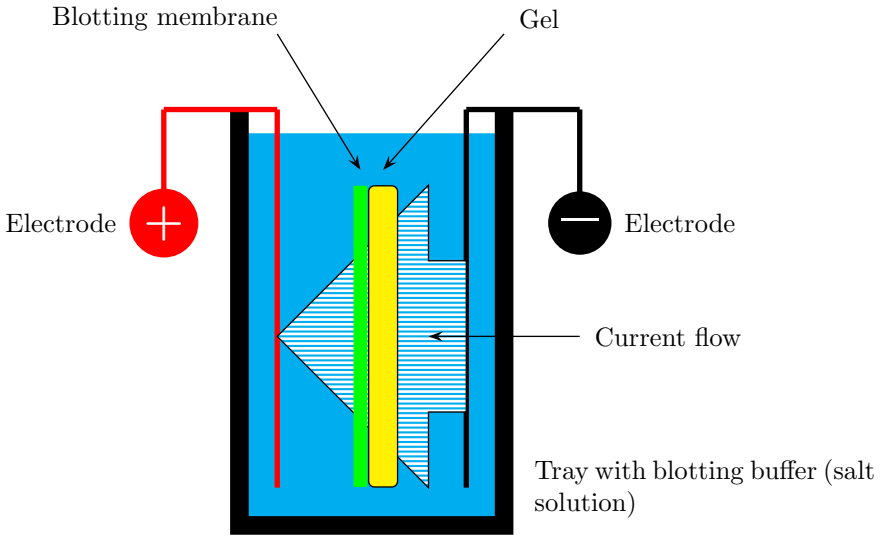


Figure 1.43: Electroblotting is blotting through an electrical field.

secondary antibody specifically binds to the primary antibody. For example, if the primary antibody was raised in rabbit, the secondary antibody is an anti-rabbit antibody, if the primary antibody was made in mice, the secondary antibody is an anti-mouse antibody, etc.

The secondary antibody is conjugated, i.e, covalently coupled, with the enzyme horseradish peroxidase, HRP. After the final wash the blot is overlaid with a thin layer of buffer containing hydrogen peroxide (H_2O_2), iodophenol and a chemiluminescent substrate, luminol, which will emit light when oxidised. The HRP conjugated to the secondary antibody uses hydrogen peroxide to oxidise luminol (see Figure 1.44). This reaction is enhanced and prolonged by iodophenol. Thus, light is produced at the place where the secondary antibody is bound. The light emission lasts for about one hour. The buffer is wiped off, and the blot is quickly exposed to photographic film to detect the light emission. The emitted light gives a *band* on the film. This detection procedure is called ECL (Enhanced Chemiluminescence).

band

Alternatively, instead of HRP, the secondary antibody can be conjugated with a radioactive label, with a fluorescent label, or with a reporter enzyme that converts its substrate resulting in a stain on the membrane. These techniques then require different types of camera and film.

The band on the film is scanned and then quantified by laser densitometry before image processing takes place. During image processing, care must be taken that the blots are not overexposed as one quickly is out of the linear range of light detection and goes into saturation. The linear range of film is not more than approximately 20, i.e., bands which differ by signal intensity of 20-fold or less are reliably quantified. This linear

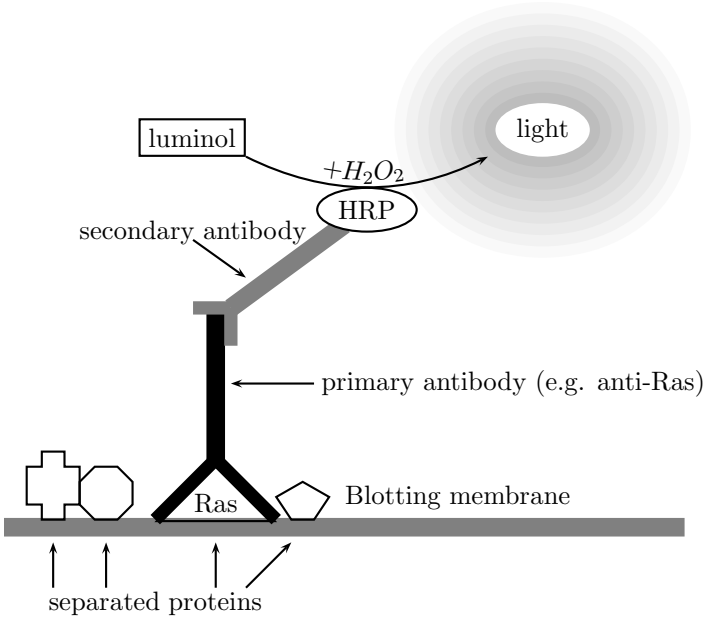


Figure 1.44: Illustration of immuno-probing a western blot with antibodies. The presence of proteins is made visible by coupling with a detectable and preferably measurable reaction.

range is most of the time sufficient unless one has very strong and very weak bands on the same blot. Then one has to scan different exposures of the blot and extrapolate. This is rather accurate, but is also labor intensive. Another option is to repeat the whole western blot (starting with the gel electrophoresis) with a diluted sample solution. This should result in weaker bands but it is possible only if enough sample material, i.e. cell lysate, is left from the original biological experiment.

An example of a western blot is shown in Figure 1.45.

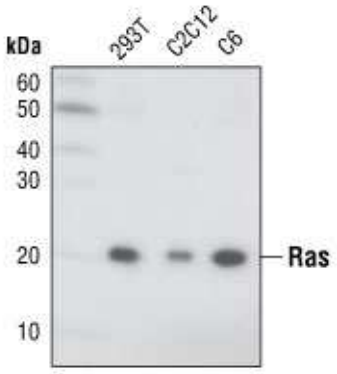


Figure 1.45: Western blot of Ras. This image is from the product page for Ras antibody #3339 (www.cellsignal.com/products/3339.html), SUBSTITUTE by one (with permission) from our collaborators! Julio?

As mentioned in section 1.5, phosphorylation of proteins plays an essnetial role in signal transduction. In order to distinguish between the phosphorylated and the unphosphorylated state of a protein, two specific antibodies against this protein are necessary, e.g. one antibody for total protein and one for phosphorylated protein only. The discovery of new antibodies and the production of these antibodies is difficult and expensive.

1.6.2.4 Considerations for modelling

When utilising the results of western blotting experiments in modelling, one has to be aware of its limitations.

1. A western blot shows *relative* changes between the samples on the blot!
2. Different western blots cannot be directly compared to each other!
3. A western blot does not yield information about the concentration of a protein!

The reasons for these are many: The intensity of the bands, or signal intensity, depends amongst others on how much protein was transferred from the gel to the membrane, on how long the ECL solution was incubated, as well as on how fast and how long the blot was exposed to film. For practical reasons these parameters are impossible to standardise. For instance, if the first exposure is under- or overexposed, then you have to put on another film - and everything has changed. Another reason is the *affinity* of the antibodies. Each antibody has a different affinity for its protein antigen. Typical K_d values are between 10^6 and 10^9. Thus, a good antibody will give a strong signal even with little protein present, and a poor antibody will give a weak signal even when lots of protein is present. Antibodies are proteins and therefore perishable molecules. There is batch to batch variation and also storage conditions can affect the affinity. These considerations apply to the primary and secondary antibody. Therefore, the observed signal intensity is a composite of the concentration of the protein antigen, the antibody affinities, the ECL conditions, and the exposure time of the film.

affinity

Thus, the only way to determine protein concentrations in a cell via western blotting, is to compare the *band intensity* obtained from a cell lysate to that of a purified protein of known concentration. For instance, to determine the concentration of the ERK protein in different cell lines, lysates of these cells were run alongside a serial dilution of a purified, recombinant ERK. The densitometric scan values of this series is used to make a standard curve. “Standard” in biochemical terms means known values for comparison. The standard curve should be a straight line, otherwise one is outside the linear range of the scanning. This curve relates protein concentration to densitometric scan units. These are the numbers from the scanner; they are arbitrary units. The concentrations of ERK in the different cell lines can be determined by mapping the scan units of the cell lysate to the standard curve. For the reasons mentioned above it is essential that the samples used to make the standard curve are highly pure, as the measured protein concentration reflects the quantity of the protein of interest plus any contaminating proteins. The purification of a protein to near homogeneity is extremely tedious. This

band intensity

absolute protein concentrations

makes the determination of *absolute protein concentrations* in cells or cell compartments so difficult and slow.

Fortunately, technologies are rapidly improving with a wide range of supplies, which differ in the way they quantify measurements. Improvements in the generation of quantitative and accurate data as well as standard operating procedures are an important aspect of systems biology. While in the early stages of molecular biology it was satisfactory to know whether a molecule is present, up- or downregulated, without much concern about measurement errors, this is a different story in systems biology. If experimental data are used for modelling it is paramount to have truly quantitative (rather than relative) and accurate long time series data.

1.7 The Dynamic Proteome

Before we continue with more technical chapters, we here briefly reflect upon the context of dynamic pathway modelling in which we are going to describe biochemical reactions networks using systems-theoretic concepts.

The area of cellular signalling investigates intracellular communication. The transmission of information within cells from receptors to gene activation by means of biochemical reaction networks (pathways) impinges on the development and disease of organisms. Our aim is to establish a mathematical/computational framework in which to investigate dynamic interactions within and between cells. In other words, we are concerned with *dynamic* pathway modelling since we do not simply map or list proteins in a pathway. Spatial-temporal sequences of reaction events in a biochemical network form a basis for *signals* that we observe as concentration changes in an experiment. We are not going to describe the physical basis of the reaction in every detail but choose a level of scale that is sufficiently high to allow a simplified description, which is however predictive.

dynamic pathway modelling

Mathematical modelling and simulation of molecular or cellular biological systems is challenging. We consider such systems as 'complex' for the following reasons. A collection of cells, but also an individual cell consist of many interacting subsystems. For example, choosing any particular pathway there will be other pathways that "cross talk". Due to the complexity of experiments to generate data and the sometimes complicated maths involved, we usually consider one pathway, or particular aspect of one pathway at a time. Since these systems (pathways) are interacting at different levels and in hierarchies, modelling is bound to be an *art* rather than an objective science. Although spatial aspects of the location of molecules in the cell, related diffusion or the transport of molecules, can in principle be encoded, for example, by partial differential equations, the available mathematical tools are often not easy to apply. Mathematical convenience is therefore one reason to make assumptions. Whether the underlying process is inherently random or deterministic may introduce further questions to how we represent this. For the kinetics of biochemical reactions, nonlinear ordinary differential equations are most commonly used for modelling while stochastic simulation is a popular avenue to avoid the complicated formal analysis of stochastic models.

In molecular biology experiments are typically expensive, time consuming undertak-

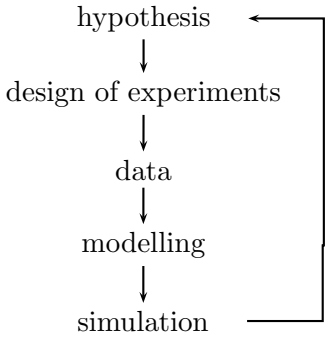


Figure 1.46: Systems biology requires an iteration of the modelling loop illustrated here. The diagram shows the role of mathematical modelling and simulation in testing hypotheses but also in generating hypotheses through prediction. The purpose of modelling is to support experimental design, helping to identify which variables to measure and why.

ings, which in most cases deliver data sets which fall short of the expectations of statisticians or mathematicians. In contrast to the engineering sciences, the observation of molecular or cellular dynamics are indirect, i.e., it is as yet not possible to obtain a continuous stream of accurate, quantitative measurements of an intact living cell. Experiments are usually destructive with regard to the components of the cell, or in order to visualise effects it is difficult not to alter the state of the system, even if only slightly. Although there is a trend towards single cell measurements, to this day we are studying the processes within *a single* cell by using thousands or millions of cells in a biochemical experiment. While statisticians would usually argue the context in which the data are generated should be irrelevant, for the analysis of molecular- or cell biological data the context in which the data were generated is crucial information to allow any sensible conclusion. It can therefore not be avoided that our models are representations of observations that help us to speculate about the true nature of the physical processes which give rise to signals and communication within and between cells. For example, when we observe steady changes of protein concentrations in a signal transduction pathway, we may want to model this phenomena with differential equations, although the underlying reactions, due to collisions of molecules, may in fact be a random process. If we insist on a stochastic model, we immediately need to consider the question of how to validate the model, i.e., how to estimate parameter values given only six to twelve time points of a nonlinear, non-stationary process.

Modelling implies a process of abstraction and is often also a form of generalisation. In this process we make numerous assumptions about the natural system under consideration and in order to simplify the mathematical approach, without loosing the ability to make predictions. It is therefore possible to build predictive models without them being precise. The Lotka-Volterra predator-prey equations of two competing populations are an example of an unrealistic model that has nevertheless value in that it helps asking the right questions[41]. Mathematical modelling and simulation should in this sense comple-

[41] Murray, [Mur02] provides a discussion of the standard Lotka-Volterra system and how more realistic scenarios can be dealt with.

ment the biologists reasoning, help him to generate and test hypotheses in conjunction with the design of experiments and experimental data.

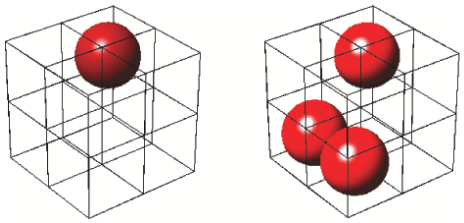
In subsequent sections we investigate the mathematical foundations of the most commonly employed mathematical concepts in modelling pathways, discuss their properties, question the assumptions involved and compare their application with examples.

Differential equations describe rates of changes and thus appear to be a *natural* framework to describe the observations we make in experiments. Differences and changes are what we can observe and what provides us with information about a system. The state-transition map describes the changes of states. Causal entailment is the principal aim of scientific modelling and that causation is the principle of explanation of *change* in the realm of matter. However, in modelling natural systems, causation is a *relationship*, not between things, but between changes of states of systems. This view that changes in space and time are the essence of causal entailment has been well explained by the philosopher Arthur Schopenhauer who argued that the subjective correlative of matter or causality, for the two are one and the same, is the *understanding*:

> "To know causality is the sole function of the understanding and its only power. Conversely, all causality, hence all matter, and consequently the whole of reality, is only for the understanding, through the understanding, in the understanding." [Mag97].

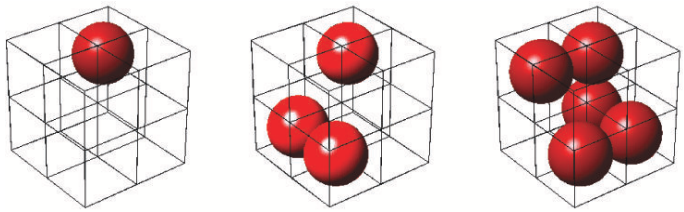
In experiments we usually look for *differences* but in our context of a dynamic systems perspective we are particularly interested in *change* over time. We can distinguish between a difference and a change, providing an example that also illustrates a difference between bioinformatics and systems biology. Let us consider the following picture as a toy model for two genomes:

bioinformatics



Comparing the two, there are $2^8 = 256$ pattern to discern. For example, from comparative genomics we know that genome sequences can be very similar, while the organisms, their physiology, behavior and appearance are very different. One then wonders how this difference in complexity is possible if the genetic material appears to be so similar. Another example is the total metamorphosis of the butterfly, there is one genome but essentially two proteomes. An explanation is that the genome may provide the basic information for an organism to develop and function, but that it is the dynamic interactions of molecules and components in the cell that give rise to biological functions. If we therefore consider again our toy model, allowing the eight genes to be switched on or off in a temporal sequence, for only three time points, there are already 256^3, i.e., more than 16 million information paths the system can describe:

genomics



time →

We may conclude from this crude illustration that it is *system dynamics* that determines biological function. In bacterial systems it is known that the speed of transcription can influence the folding, structure and thus to an extend also the function of proteins. Another illustration that knowledge of components or symbols is not enough is the following set:

systems dynamics

$$\{h, t, u, r, t, e, h, t, l, l, e, t, t, b, u, o, d, n, i, f, i\} \ .$$

Although not randomly arranged, looking at this list does not suffice to understand the meaning. Ordering them further does appear to make sense:

$$\{i, f, i, n, d, o, u, b, t, t, e, l, l, t, h, e, t, r, u, t, h\} \ ,$$

but a real understanding comes only if we read the symbols with a particular speed from left to right:

> "If in doubt, tell the truth." (Mark Twain)

Here again the temporal evolution is important: If we were given a letter a day, we would usually not be able to make sense of we are told, nor would make a foreign language and dialect make this easier. In line with the two quotations of Pauling and Poincare on page 3 we would argue that molecular and cell biology are built up from facts, as a cell is built from molecules. But a collection of facts is no more a science than a soup of molecules is a cell. Organisms and organs are complex structures of interdependent and subordinate components whose relationships and properties are largely determined by their function in the whole.

1.8 Outlook

There are various approaches to arrive at a mathematical model of intra- and intercellular dynamics. We are going to restrict ourselves to stochastic modelling and the use of differential equations. For differential equations there are again various perspectives one can take to motivate the set-up of the equations. To begin with, we are considering a reaction network or pathway involving N molecular species S_i. A network, which may include reversible reactions, is decomposed into M unidirectional basic reaction channels R_μ

$$R_\mu \colon l_{\mu 1} S_{p(\mu 1)} + l_{\mu 2} S_{p(\mu 2)} + \cdots + l_{\mu L_\mu} S_{p(\mu L_\mu)} \xrightarrow{k_\mu} \cdots$$

where L_μ is the number of reactant species in channel R_μ, $l_{\mu j}$ is the stoichiometric coefficient of reactant species $S_{p(\mu j)}$, $K_\mu = \sum_{j=1}^{L_\mu} l_{\mu j}$ denotes the molecularity of reaction

channel R_μ, and the index $p(\mu j)$ selects those S_i participating in R_μ. Assuming a constant temperature and that diffusion in the cell is fast, such that we can assume a homogenously distributed mixture in a fixed volume V and with a constant temperature and volume, one rate-equation model is the following

$$\frac{d}{dt}[S_i] = \sum_{\mu=1}^{M} \nu_{\mu i} k_\mu \prod_{j=1}^{L_\mu} [S_{p(\mu j)}]^{l_{\mu j}} \qquad i = 1, 2, \ldots, N \tag{1.30}$$

where the k_μ's are rate constants and ν_μ denotes the change in molecules of S_i resulting from a single R_μ reaction. Representation (1.30) is a system of ordinary differential equations:

$$\frac{dx_i(t)}{dt} = V_i\big(x_1(t), \ldots, x_N(t), \theta\big) \ , \tag{1.31}$$

where x_i denotes the n variables in question, θ denotes a parameter vector and V is a nonlinear function. The reader who is less accustomed to mathematical equations should not worry, we are going to approach the general formulations (1.30) and (3.46) below, with numerous examples of increasing complexity and generality.

Rate equation models have been widely used in describing biochemical reactions (e.g. [HS96, Fel97]). For modelling processes in living cells we are often not able to provide an exact description of the physical interactions of molecules and instead describe with our differential equations a mean of a population of interacting molecules, where the $[S_i]$ are most probable values. The fact that a differential equation is 'deterministic' does therefore not mean that it cannot describe a process that is inherently stochastic.

In a stochastic framework, we are looking at populations of molecules and wish to determine for each molecular species S_i the probability $\text{Prob}\{\#S_i(t) = n_i\}$ that at time t there are n_i molecules. For N molecular species, let n denote the N-dimensional state-vector, whose values are positive integers, $n \in \mathbb{Z}_+^N$. $\nu_\mu \in \mathbb{Z}^N$ are the step-changes occurring for elementary reaction indexed by μ. If S is a N-dimensional variable, we write $\text{Prob}\{\#S = n\} = P_n(t)$. Describing the changes in random variable S, we consider the following two state-transitions: First, from other states to state n, denoted

$$n - \nu_\mu \xrightarrow{a_\mu(n-\nu_\mu)} n \ ,$$

propensity

where $a_\mu(n - \nu_\mu)$ is the *propensity*, that is the probability per unit time, of a change ν_μ occurring, given that we are in state $n - \nu_\mu$. Secondly, moving away from state n is given as

$$n \xrightarrow{a_\mu(n)} n + \nu_\mu \ .$$

chemical master equation (CME)

From these definitions we arrive at an expression referred to as *master equation* or *chemical master equation* (CME)

$$\frac{\partial P_n(t)}{\partial t} = \sum_{\mu=1}^{M} \left[a_\mu(n - \nu_\mu) P_{(n-\nu_\mu)}(t) - a_\mu(n) P_n(t) \right] \ . \tag{1.32}$$

The first term on the right-hand side describes the change to state n, while the second term describes the changes away from state n. M denotes the number of (unidirectional) reaction channels. The product of the propensity with the probability should be read as an "and". The multiplication of a propensity and probability makes sense in light of the derivative against time on the left, in that a propensity, multiplied with dt gives a probability. In the above setting $a_\mu(n-\nu_\mu)$ and $a_\mu(n)$ are *transition probabilities*, while $P_n(t)$ and $P_{(n-\nu_\mu)}(t)$ are the probabilities that the system is at time t in state n, respectively $n-\nu_\mu$. A solution of the master equation describes the probability of there being n molecules at time t. Chapter 3 is going to provide a series of examples for stochastic processes for which the methodologies become increasingly general until we arrive at (3.46).

transition probability

A major difficulty with the CME is that the dimension of these sets of equations depends not only on the number of chemical species N but for any possible number of molecules of any species we have n differential-difference equations. Gillespie [Gil76, Gil77, Gil01, GB00] developed an algorithm to simulate or realise a CME model efficiently. The Gillespie approach to stochastic *simulation* has in recent years become popular in modelling intra-cellular dynamic processes [Kie02, RWA02, vGK01, MA97]. Some authors have unfortunately confused the simulation of a stochastic model with a stochastic model. The Gillespie algorithm does not provide a solution to the master equation but generates realisations. While a formal analysis of (3.46) is very difficult, it is possible to approximate the CME by a truncated Taylor expansion, leading to the Fokker-Planck equation, for which there exist some results [ELS01, Gar85, vK92]. Comparing (1.30) and (3.46), we note that while rate equations are deterministic in the sense that they employ differential equations, they are based on a probabilistic description of molecular kinetics. On the other hand, the CME is a stochastic formulation, but based on differential equations, with probabilities as variables. Although we are going to look at various stochastic models and their derivation, we eventually settle for model structures (1.31), to describe molecular principle from what we can observe in experiments.

The motto of this book is nicely captured in the following quotation by Ludwig von Bertalanffy, a founder of general systems theory and someone who laid the foundations for systems biology in the 1960s:

> "Considering the inconceivable complexity of processes even in a simple cell, it is little sort of a miracle that the simplest possible model – namely, a linear equation between two variables – actually applies in quite a number of cases.
>
> Thus even supposedly unadulterated facts of observation already are interfused with all sorts of conceptual pictures, model concepts, theories or whatever expression you choose. The choice is not whether to remain in the field of data or to theorise; the choice is only between models that are more or less abstract, generalised, near or more remote from direct observation, more or less suitable to represent observed phenomena.
>
> On the other hand, one should not take scientific models too seriously. Kroeber (1952), the great American anthropologist, once made a learned study of ladies' fashions. You know, sometimes skirts go down until they impede

> the lady in walking; again, up they go to the other possible extreme. Quantitative analysis revealed to Kroeber a secular trend as well as short-period fluctuations in the length of ladies' skirts. This is a perfectly little law of nature. I believe a certain amount of intellectual humility, lack of dogmatism, and good humor may go a long way to facilitate otherwise embittered debates about scientific theories and models."
>
> (From Bertalanffy's book *General Systems Theory*; the introductory part to the chapter 'Some Aspects of Systems Theory in Biology' [Ber68].)

The previous section took us from a relational world view to a general systems approach. The present section introduced the formal systems we are dealing with hereafter. The next section takes us on to the journey in which we apply the concept above to an understanding of the processes that make up a living system. Enjoy!

2 Biochemical Reaction Kinetics

For anyone interested in modelling and simulation of biochemical reaction networks or pathways, there are plenty of tools to choose from. These packages are roughly divided into whether they deal primarily with deterministic or stochastic models. Here we are discussing the close relationship between both ideas, in the context of biochemical reactions. We begin with simple, particular examples and increase the generality of the equations before we arrive at the general chemical master equation (3.46) and have available a set of more advanced mathematical tools to investigate the properties of such stochastic representations.

2.1 The Rate Equation Approach

We are now going to go through a short example of the modelling process. The aim is to describe changes in a population of molecules. We first consider only one kind or *species* of molecules, which we call S. As a conceptual framework in which we formulate our models we initially consider Ludwig Boltzmann's *Kinetic Theory of Gases* from 1877. It begins with the assumption that for constant pressure, temperature, and volume V, the number of collisions between any two molecules should be constant. Let $\#S \doteq n$ denote the average number of molecules of species S. If the probability of a reaction to occur is independent of the details of that collision, then the change Δn in the number of molecules is proportional to n as well as to a time interval Δt:

$$\boxed{\Delta n \propto n \cdot \Delta t}$$

There are several kinds of biochemical reactions, which in turn can be combined into networks or pathways. We start with the simplest reaction which proceeds by itself, involving only one molecular species. The empirical rule we described above can now be turned into a mathematical equation

$$\Delta n = k \cdot n \cdot \Delta t \ , \tag{2.1}$$

where Δt is assumed to be a *relatively* small interval of time. Dividing both sides by Δt,

$$\frac{\Delta n}{\Delta t} = k \cdot n \ ,$$

we can now consider what happens as Δt becomes infinitesimal small, $\Delta t \to 0$, leading to the differential operator. For large n, changes in this very small time interval will be very small compared to the overall changes in the population. We could thus turn

differential equation

from discrete changes in the number of molecules n to continuous changes $S(t)$ in the population or concentration of S. This then leads us to the ordinary *differential equation*[1]

$$\frac{dS(t)}{dt} = k \cdot S(t) \ , \tag{2.2}$$

as a model of the simple monomolecular or autocatalytic reaction. A differential equation describes the *rate of change* of variable S and is a translation of the observations and assumptions we make in modelling a natural system. To simulate the system and to make predictions for values of $S(t)$ for a range of time points t, we need to find a *solution* to the differential equation. For simple cases we may be able to find *analytical* solutions through mathematical analysis, while for more complex cases we need to resort to a *numerical* solution or simulation. The advantage of an analytical solution is that it is more general, we usually do not need to know an exact value for the parameters, here k. For our simple reaction the solution to the differential equation (2.2) is

analytical vs. numerical solutions

$$S(t) = S(0) \cdot e^{kt} \ , \tag{2.3}$$

rate constant

where k is the *rate constant* by which the conversion of reactant A proceeds and $S(t_0) = S(0)$ defines the initial condition. We frequently write S_0 for $S(0)$. The solution is thus dependent on a parameter value for k *and* the initial conditions. Since we could get the differential equation (2.2) from (2.3) by differentiating the equation, another way round to a solution of the differential equations is by integration. A simulation of a system of differential equation is therefore a numerical integration of (2.2). We are going to discuss the process by which we take the limit $\Delta t \to 0$ and integrate the differential equation in more detail further below in this section. Before that, we look at the application of our model.

The model (2.2) describes an *irreversible* biochemical reaction, that is, the population or concentration of S can either only increases or only decreases as the result of changes to the molecules. As we shall see later, this irreversible reactions are somewhat unrealistic for living cells. The reaction is also *monomolecular* since it involves only one kind of species. Whatever happens to S, we could consider it to be a transformation of S_1 into S_2. In other words, we have two kinds or forms of S, the biochemical notation for this is

$$S_1 \to S_2 \ ,$$

reactant species
product species

where the chemical species on the left are referred to as the *substrate* or *reactant species* while those on the right are called *product species*. As above, for a small time interval Δt, the changes ΔS_1, ΔS_2 satisfy the proportionality relation

$$\Delta S_1 \propto -S_1 \Delta t \ ,$$

and if the number of molecules is conserved

$$\Delta S_2 = -\Delta S_1 \ .$$

[1] An *ordinary* differential equation, as opposed to a *partial* differential equation, does not consider spatial distribution of components or diffusion.

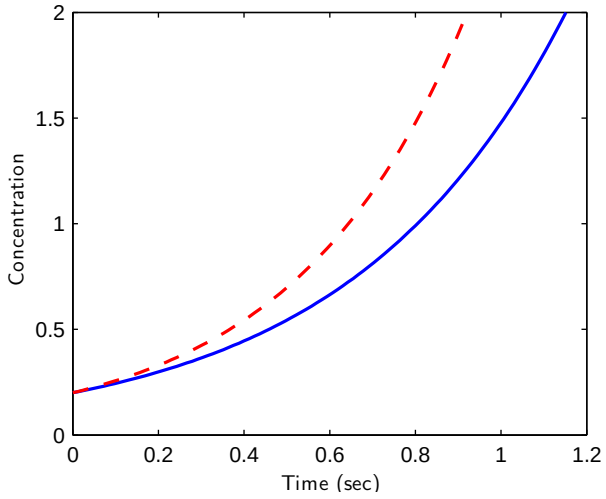


Figure 2.1: Simple exponential growth of a population, $S(t) = S(0) \cdot e^{kt}$. The dashed line is for parameter $k = 2.5\text{sec}^{-1}$ and the solid line for $k = 2\text{sec}^{-1}$. The solution of the differential equation $dS(t)/dt = kS(t)$, depends on the initial condition (here $S(0) = 0.2$), that is, for each initial condition there is a different solution. The fact that this curve growths unbounded suggests that it is not a realistic growth model.

For $\Delta t \to 0$ these relations turn into differential equations

$$\frac{d}{dt}S_1 = -kS_1 \qquad \frac{d}{dt}S_2 = kS_1 \ .$$

We are now at a point where we need to discuss the units of the elements of our model. The molar concentration $[S]$ is given as mole per liter:

$$1\frac{\text{mol}}{\text{L}} \equiv 1\text{M}$$

In general, in equations we denote the volume by V. Since a mole contains $6 \cdot 10^{23}$ molecules, in order to get a count of the actual number of molecules $\#S$ we would have to multiply the molar mass by the Avogadro constant

$$N_A = 6.02205 \cdot 10^{23}\text{mol}^{-1} \ .$$

With a slight abuse of terminology, we hereafter occasionally use S to denote a "count" of molecules in moles. Note, if it is clear from the context and explicitly stated, we may also use S to denote a concentration, leaving away the square brackets $[S]$ to simplify the notation. We are thus assuming that concentrations are taking real values and change smoothly. Using square brackets $[S]$ to denote molar concentrations (M) for our example, the unit of $d[S]/dt$ is M per second and we have

$$[S] = \frac{S}{V} \quad \text{and} \quad \#S = S \cdot N_A \ .$$

Although here the rate constant k is measured as 'per second', denoted $\sec^{-1}$, in general the units of the rate constant will depend on the type of the reaction and whether we are dealing with concentrations or counts.

Most if not all biochemical reactions in the living cell are reversible and so the next step in making our approach to modelling biochemical reactions more comprehensive, is to consider a reversible reaction in which compound S_1 is transformed into compound S_2, and *vice versa*:

reversible reaction

$$S_1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} S_2$$

where k_1 is called the *forward rate constant* and k_2 *reverse rate constant*. If the reversible reaction is in an *equilibrium*, the average concentrations remain constant[2] and the rates of changes are zero

equilibrium

$$\frac{d}{dt}[S_1] = \frac{d}{dt}[S_2] = 0 \ ,$$

which is the same as to say

$$\frac{k_1}{k_2} = \frac{[S_2]}{[S_1]} \ .$$

This relation is what C.M. Guldberg and P. Waage in 1864 described as the *law of mass action* (LMA). The key to the differential equation model was the assumed proportionality in (2.1). From our initial discussion, the following set of differential equations serves as a model for the reversible reaction above:

law of mass action

$$\frac{d}{dt}[S_1] = -k_1[S_1] + k_2[S_2] \ , \qquad \frac{d}{dt}[S_2] = k_1[S_1] - k_2[S_2] \ .$$

If we consider the total concentration of S_1 and S_2 together as constant, $[S_1]+[S_2] = S_T$ and substitute $[S_2] = S_T - [S_1]$ for $[S_2]$ in the equation for S_1, we obtain

$$\frac{d[S_1]}{dt} = -(k_1 + k_2)[S_1] + k_2 S_T \ .$$

Simplifying the notation, let us denote $S_1 \doteq x$, $\alpha \doteq k_1 + k_2$ and $\beta \doteq k_2 S_T$ so that the differential equation looks tidier:

$$\frac{dx}{dt} = \beta - \alpha x \ . \tag{2.4}$$

This model is then an example of the system considered in the previous chapter:

$$\dot{x} = \phi(x, \theta) \ ,$$

where $\theta = (\alpha, \beta)$. This is a basic differential equation, for which there are various ways to solve it. Here we simply state the result in order to compare it with (2.3) as a model for population growth:

$$x(t) = \frac{\beta}{\alpha} + \left(x(0) - \frac{\beta}{\alpha} \right) e^{-\alpha t} \ .$$

[2] Note that this does not mean nothing happens. There is still a flux of material, although the reactions are balanced such that macroscopically we do not observe changes.

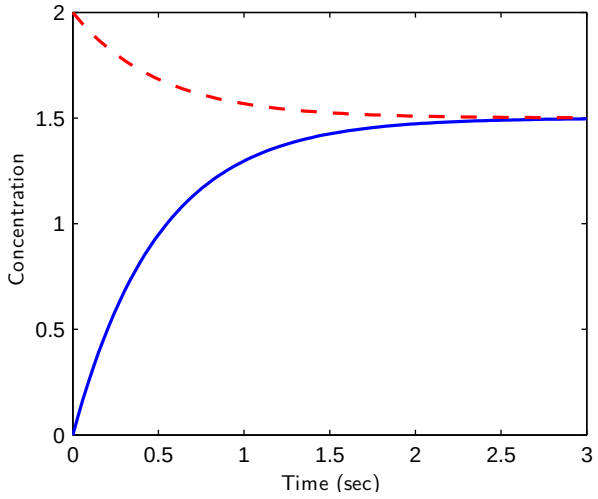


Figure 2.2: Numerical solution for the differential equation (2.4), which was derived as a model for a reversible reaction. The solid line is for $x(0) = 0$, while the dashed line is for $x(0) = 2$ M. $\alpha = 2\text{sec}^{-1}$, $\beta = 3$ M/sec.

Figure 2.2, shows simulation results for two different initial conditions. We observe that although the growth depends heavily on the initial conditions, the growth is limited to β/α.

2.1.1 Differential vs. difference equations

A justified criticism of the ODE model (2.2), as a representation of a biochemical reaction could be that we did not aim for *physical realism*, modelling the collisions of molecules. Instead we modelled what we observed: that at any time the change is proportional to Δt and the current number of molecules. A reasonable interpretation of (2.2) is then that $S(t)$ represents the average population level at time t. In this case we can view k as the difference between the formation rate k^+ of S and the decay rate k^- such that

$$\frac{dS(t)}{dt} = (k^+ - k^-)S(t) \quad \text{with solution} \quad S(t) = S(0) \cdot e^{(k^+ - k^-)t} \ . \tag{2.5}$$

In the derivation of the differential equation we made another assumption of a large population $\#S \doteq n$ such that discrete changes to the population are small enough to assume overall changes to the population are continuous (see Figure 2.3 for an illustration). This assumption in effect describes a discrete process with a continuous model. We can discuss this mathematically by approximating a difference equation by a differential equation. A simple intuitive example for this is the interest we receive for savings in a bank account. Say a bank offers on savings a rate of return r, in percent, compounded annually. If $S(0)$ is the initial saving put in the bank, and $S(t)$ its future value after t years, then the following *difference equation* models the growth of our money: difference equation

$$S(t+1) = (1+r)S(t) \ ,$$

with initial condition $S(t_0) = S(0)$. If we are interested in growth on a monthly basis,

$$S\left(t + \frac{1}{12}\right) = \left(1 + \frac{r}{12}\right) S(t) \ .$$

Here r is divided by 12 because it is by definition compounded annually (not monthly). In general, if a year is divided into m equal intervals, the difference equation becomes

$$S\left(t + \frac{1}{m}\right) = \left(1 + \frac{r}{m}\right) S(t) \ ,$$

which can be written more conveniently

$$\frac{S\left(t + \frac{1}{m}\right) - S(t)}{\frac{1}{m}} = rS(t) \ .$$

As m goes to infinity, denoted $m \to \infty$, the above difference equation becomes a differential equation

$$\frac{dS(t)}{dt} = rS(t) \ , \tag{2.6}$$

where t is now a continuous time variable. One reason for choosing a continuous time representation with continuous changes is the rich set of analytical tools that is available to investigate the properties of such equations.

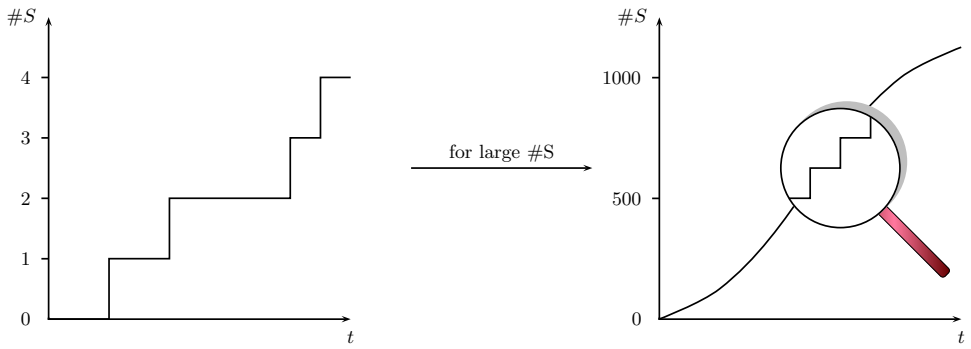


Figure 2.3: Approximation of discrete steps as continuous changes. For large numbers of molecules changes to the population appear smooth and may be represented by a continuous model.

2.1.2 Numerical simulation

Here we look at another assumption made in our model (2.2), that of a small interval of time Δt. This question is closely related to finding solutions to differential equations. For simple linear ordinary differential equations like the one above we can find exact or analytical solutions. For more complicated cases, in particular nonlinear equations, which we consider later, we can use numerical integration to obtain an approximation

[AS72, PTVF93]. The simplest approach to obtain a numerical solution of a general ordinary differential is called the *forward Euler method*:

uler method

$$\frac{dS}{dt} \approx \frac{\Delta S}{\Delta t} = \frac{S(t+\Delta t) - S(t)}{\Delta t} \ , \tag{2.7}$$

where ΔS and Δt are small, but not infinitesimal. If this approximation to the derivative is substituted into (2.5) and the equation is rearranged we get

$$\begin{aligned} \frac{S(t+\Delta t) - S(t)}{\Delta t} &\approx k^+ S(t) - k^- S(t) \\ S(t+\Delta t) &= S(t) + S(t)(k^+ - k^-)\Delta t \ . \end{aligned}$$

The justification for the differential d/dt may be considered a mathematical explanation. A physical argument is that in order to avoid surface effects, influencing the interactions among molecules, we consider an infinitely large system $\lim V \to \infty$. To avoid that the concentration goes to zero, the number of molecules must become very large in order to move from a discrete sum to a continuous integral.

2.2 Biochemical Reaction Modelling

Following the above introduction to differential equation modelling, the present section is to provide a more comprehensive survey of biochemical reaction modelling. The theoretical and experimental description of chemical reactions is related to the field of *chemical kinetics*. A primary objective in this area is to determine the rate of a chemical reaction, i.e., describing the velocity of conversion of reactants to products. Another task is the investigation of the influence of external factors, like temperature, pressure, and other chemical species on the chemical reaction under consideration. The determination of the reaction mechanism, the way the products are formed, which intermediates are created, is a further field of chemical kinetics.

2.3 Fundamental Quantities and Definitions

The pivotal quantity in the description of chemical reactions is the *reaction rate*. The general chemical reaction

$$|l_1|\, S_1 + |l_2|\, S_2 + \ldots + |l_i|\, S_i \ \rightarrow \ |l_{i+1}|\, S_{i+1} + |l_{i+2}|\, S_{i+2} + \ldots \tag{2.8}$$

can be summarised by

$$0 = \sum_i l_i S_i \ ,$$

where l_i denotes the *stoichiometric coefficient* for the i-th component of the chemical reaction defined in (2.8). The reaction rate of this reaction is defined as

stoichiometric coefficient
reaction rate

$$r(t) = \frac{1}{l_i} \frac{d\#S_i(t)}{dt} \ , \tag{2.9}$$

where $\#S_i$ is the number of molecules of species S_i in the considered volume. The stoichiometric coefficients l_i are negative for reactants and positive for products. From the definition (2.9) it follows that the reaction rate is a positive definite quantity identical for all participating species S_i.

In wet-lab experiments and in modeling reactions in a cell we usually assume a constant volume, although in the cell this is surely not always a realistic assumption. For a system with constant volume we can transform (2.9) into an equation for the concentration $[S_i]$ of species S_i

$$r(t) = \frac{1}{l_i}\frac{1}{V}\frac{d\#S_i(t)}{dt} = \frac{1}{l_i}\frac{d\#S_i/V}{dt} = \frac{1}{l_i}\frac{d[S_i](t)}{dt} \ . \tag{2.10}$$

In general one can use any quantity which is proportional to the number of molecules, for instance the particle density or the partial pressure of gases but for practical considerations the measurement of concentrations is often easier than the count of molecules. But there is a more important difference between these quantities. The particle number is an *extensive property*, i.e., it depends on the size of the system. The concentration, particle density, ... are *intensive quantities* independent from the particular system under consideration. We hereafter assume a system with constant volume, in which case we get for (2.8) a reaction rate

$$\begin{aligned} v(t) &= -\frac{1}{|l_1|}\frac{dS_1(t)}{dt} = -\frac{1}{|l_2|}\frac{dS_2(t)}{dt} = \ldots = -\frac{1}{|l_i|}\frac{dS_i(t)}{dt} \\ &= \frac{1}{|l_{i+1}|}\frac{dS_{i+1}(t)}{dt} = \frac{1}{|l_{i+2}|}\frac{dS_{i+2}(t)}{dt} = \ldots \ , \end{aligned} \tag{2.11}$$

where $S_i(t)$ represents either a particle number or a proportional quantity. Degrading species are characterised by a minus sign while an increase is indicated by a plus sign. According to (2.11) the reaction rate is proportional to the change of concentration. Another possibility to investigate the reaction rate is given by the advancement or *extent of an reaction* $\epsilon(t)$. This quantity is a measure of the progress of the chemical reaction under consideration. It is defined as

$$\epsilon(t) = \begin{cases} \frac{1}{l_i}\left(S_i(t) - S_i(0)\right) & \text{for products} \ , \\ \\ \frac{1}{l_i}\left(S_i(0) - S_i(t)\right) & \text{for reactants} \ . \end{cases} \tag{2.12}$$

The extent $\epsilon(t)$ relates the initial conditions $S_i(0)$ to the time dependent variables $S_i(t)$ and has the same value for all species. We are using it for the analytical integration of time laws of higher orders later in the text. With the extent the reaction rate is

$$v(t) = \frac{d\epsilon(t)}{dt} = \frac{1}{l_i}\frac{dS_i(t)}{dt} \tag{2.13}$$

and is interpreted as rate of change of the advancement of a reaction.

It is possible to formulate a conservation law which relates the reactants and the products in a closed system:

$$S_0 = S_1(t) + S_2(t) \ , \tag{2.14}$$

where S_0 is the initial concentration of the reactant S_1, where $S_1(t)$, $S_2(t)$ are time dependent concentrations. Here, we assume that the initial concentration of the product S_2 is zero. Relation (2.14) describes the conservation of the number of atoms in a *closed system* under the influence of the chemical process. One distinguishes between three types of systems according to their possibilities to exchange energy and matter with the environment. The most restricted system is an isolated or fully closed system, where no transfer of energy and matter is possible. If we allow the exchange of energy, but no exchange of matter, the system is called closed. Such a system is often used for chemical and biochemical experiments. In an *open system* also a transfer of matter is possible. Examples are flow reactors in the chemical industry and of course the cell in its natural environment. The creation and the disappearance of atoms is not a chemical process. Since the concentration is a function of the molecule number, we obtain a conservation of concentrations. Often, this law is also called *mass conservation*. From a physicist's perspective this is wrong. The mass is not a conserved quantity in chemical systems. As mentioned above, we can use any quantity proportional to the particle number. For each of them one can formulate a specific conservation law of the form (2.14). To avoid a restriction of systems in our treatment we will simply call it conservation law for this reason.

closed system

closed/open system

conservation law

The advantage of conservation laws is, that they simplify the description of the system of interest and give conditions to narrow relevant solutions down. The simplification arises from the fact, that each conservation law eliminates a variable and reduces the order of a system of coupled differential equations. Further famous examples of conservation laws are energy conservation, momentum conservation and the angular momentum conservation.

2.4 Basic Principles and Assumptions

In the previous section we defined the reaction rate as a differential equation that depends on the change of participating species over time. In order to obtain the temporal behavior of molecular species, we have to specify the functional relation of change. The class of differential equations is not automatically restricted to ordinary differential equations. In ODEs the rate of change is described by a continuous function. The chemical conversion is however not a continuous process as it is postulated for the use of functions. If one wishes to consider the discreteness of the process without a change of the general framework, this is possible but requires the introduction of what are called the Dirac-δ-function and the Heaviside step-function θ. In a strict mathematical sense these are not functions but distributions and hence the differential equations are strictly speaking no longer 'ordinary' differential equations. The fact that it is possible to describe discrete changes in reactions with differential equations is worth mentioning. We are going to return to this question when we introduce stochastic models. In the literature stochastic models are frequently justified by stating that differential equations are not able to capture discrete changes. A formal theoretical description of discrete changes in reactions in terms of distributions is possible although non-trivial. One has to know the time of

each reaction within the considered system. But since a prediction of the exact time is not possible, only statistical properties are known for chemical reactions. In particular the probability of a reaction can be calculated. Stochastic simulations work in this way. The simulations calculate the time of a reaction from probability distribution functions. The result of a simulation run is one possible realisation of the temporal evolution of systems and corresponds to the formal treatment described above. The calculation of probabilities of a reaction requires a detailed information about the process of a chemical reaction. Collision theory is a successful approach to justify experimental rate laws. Two particles can only react if they interact which each other. In a classical view both particles must collide, like two balls on the billiard table.

The description of a molecule as a hard sphere is motivated by the subsequently feasible assumption of an interaction with a short contact time between the molecules. This means the interaction range is much smaller than the average distance between molecules. It is then possible to assume collisions as independent, an assumption useful in the context of stochastic Markov models. Because of the finite interaction range we also have a finite interaction time. This is the time, a molecule needs to move through the potential of its collision partner. Within this time old bonds break and new ones are established. Analogues to our assumption on the interaction length the interaction time is small and negligible in comparison to the time between two collisions. It follows from this that the number of reactions per time is related to the number of collisions within the considered time interval. Statistical physics is one possible tool for the solution of this problem. The main assumption made in statistical physics is, that the properties of the investigated system are well described by the 'expected value' or mean value. Therefore, all results obtained in due course have to be interpreted as averages. There are fluctuations around the expected value, but these are assumed small in comparison to the average. Because of the use of averages we change from a discrete description to a continuous formulation. Furthermore we assume an ideal gas in thermodynamic equilibrium. Then, the velocity of the molecules is given by the Maxwell-Boltzmann distribution function. In order to avoid surface effects[3] one assumes an infinitely expanded system. Because of the infinite system volume V one has to increase the particle number $\#S$ to infinity to keep the right particle density. All further treatments are in this thermodynamic limit, requiring that in the limit of $V, \#S \to \infty$ the ratio $\#S/V$ is constant and finite. Last but not least we have to make an assumption each chemical reaction is independent from the others. In addition to these assumptions we also restrict our treatment to a special class of systems called *isothermal* and *isochore* reaction systems. This means that during the reaction no temperature changes and no volume changes occur.

What follows is a short summary of chemical kinetics. From collision theory the reaction rate is equal to the number of molecules participating in the chemical reaction and the *rate coefficient* k. The rate coefficient[4] k summaries details of the reaction and

rate coefficient

[3] The presence of a reactive surface can dramatically change the properties of a (bio)chemical system. The catalyst of cars works in this way.

[4] Often k is called rate constant because of its time independence. But k is dependent on system parameters like the temperature or the pH-value and hence the term 'coefficient'.

is only fully known for very simple systems. This requires information about particle motion and is therefore temperature dependent. Furthermore, not all collisions lead to a chemical reaction. The collision energy can be too small to initiate the reaction. The molecules have a false direction, for instance, the molecules do not hit their reactive sites and cannot react. The molecules are too fast, it is not enough time to break the old and establish the new chemical bindings. These are only few of the possible reasons for a non-reactive collision. In summary, the theoretical calculation of the rate coefficient is complicated. For this reason we have to resort to experimental or estimated values. If identical species react, the rate coefficient also contains a symmetry factor avoiding a double counting[5].

A further common assumption for pathway modeling is, that we can decompose more complicated chemical reactions into a sequence of elementary reactions, which we can describe using chemicals kinetics. The most common classification is to distinguish the reaction by the number of participating molecules. The simplest reaction is the monomolecular reaction with only one molecule. In a bimolecular reaction two molecules or substrates form a product. This is the most frequently occurring reaction in nature. Trimolecular reactions are rare, because the probability for a collision of three particles within a tiny time interval is rather small. We will deal with these elementary reactions and more complicated reactions in the following sections.

2.5 Elementary Reactions

Pathways are networks of biochemical reactions. To describe the behavior of this reaction networks it would then seem plausible to use chemical kinetics to derive the mathematical equations. We shall later see that for many intracellular processes we will in practice not be able to measure interactions in all detail, nor will we know all the properties of the proteins involved. This will in effect mean that we aggregate information about mechanistic detail. As a consequence, the model is not derived from "first principles" (of physical mechanics) but of a "phenomenological nature". The expression "phenomenological" has negative associations, suggesting some arbitrariness in the construction but this is not true. In fact, the very definition of a pathway implies a reduced model. Virtually all pathways that are investigated in the context of biomedical research are considering only a reduced set of proteins. In practice it is impossible to generate data for all proteins of interest and one is forced to a selection. It is for this reason that a thorough understanding of the assumptions in modelling is so important.

For the present section, we assume that a reaction network can be decomposed into elementary reactions. For simple reactions it is possible to derive rate laws and determine solutions to differential equations analytically.

[5] For instance, for a reaction of two identical molecules one have to introduce a symmetry factor 1/2 otherwise one counts each collision twice.

2.5.1 Monomolecular reactions

The monomolecular reaction

$$S_1 \xrightarrow{k} S_2 \tag{2.15}$$

monomolecular reaction

is the simplest elementary reaction, converting species S_1 into S_2. The variable k above the arrow is the corresponding rate coefficient. For the *monomolecular reaction* the rate coefficient has the unit time^{-1}, independent from the units used for the species. The quantity $k\,dt$ is the probability that a reaction occurs within the time interval dt.

reaction rate

According to the common chemical kinetics the *reaction rate* is

$$r(t) = -\frac{dS_1(t)}{dt} = k\,S_1(t) \tag{2.16}$$

which can integrated by separation of variables

$$\frac{dS_1(t)}{S_1(t)} = -k\,dt\ . \tag{2.17}$$

The integration within the limits $S_1(0)$ to $S_1(t)$ and from 0 to t in the time domain

$$\int_{S_1(0)}^{S_1(t)} \frac{dS_1(t)}{S_1(t)} = -k \int_0^t dt \tag{2.18}$$

leads to

$$\ln \frac{S_1(t)}{S_1(0)} = -kt\ . \tag{2.19}$$

Solving this for $S_1(t)$ we obtain the familiar exponential law

$$S_1(t) = S_1(0)\exp\{-kt\}\ , \tag{2.20}$$

for the temporal evolution of S_1. $S_1(0)$ is the initial condition at $t = 0$. The solution for the product S_2 is obtained from the conservation law

$$S_2(t) = S_1(0) - S_1(t) + S_2(0) \tag{2.21}$$

as

$$S_2(t) = S_1(0)\,[1 - \exp\{-kt\}] + S_2(0)\ , \tag{2.22}$$

with the initial value $S_2(0)$.

2.5.1.1 Characteristic times

The t_y-time is the time where the normalised quantity $S_1(t)/S_1(0)$ has the value y. Hence, possible values for y lie in the interval $[0, 1]$. From (2.20), it follows that

$$t_y = -\frac{\ln y}{k}\ . \tag{2.23}$$

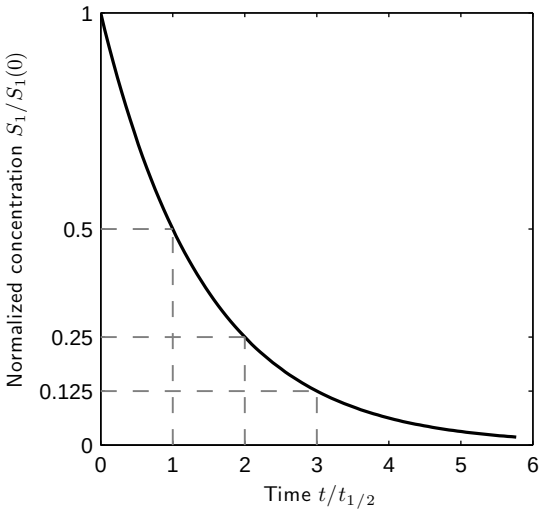


Figure 2.4: Monomolecular reaction (2.15) as function of the half life $t_{1/2}$. The normalised concentration $S_1(t)/S_1(0)$ has the value 2^{-n} at the n-th multiple of the half life.

half life

The most commonly known t_y-time is the *half life* $t_{1/2}$. At this time point half of the initial amount $S_1(0)$ is transformed into product S_2. It follows from the general definition (2.23)

$$t_{1/2} = \frac{\ln 2}{k} \approx \frac{0.69}{k}\,, \tag{2.24}$$

which is independent from S_1. This means that it takes always $t_{1/2}$ to halve the amount of S_1. This is illustrated in Figure 2.4, where $S_1(t)$, normalised by $S_1(0)$, is drawn as function of the half life $t_{1/2}$. At every multiple n of $t_{1/2}$ the ratio $S_1(t)/S_1(0)$ has the value $1/2^n$ as mentioned before.

2.5.1.2 Measurement of the rate coefficient

In case of a monomolecular reaction the rate coefficient can be measured in a simple way. From (2.19) the equation

$$\ln S_1(t) = \ln S_1(0) - k\,t \tag{2.25}$$

defines a straight line with slope $-k$ and an intersection with the y-axes $\ln S_1(0)$. If one applies the logarithm of measured data, as function of time, then the rate coefficient k is obtained from a linear fit. This is sketched in Figure 2.5.

2.5.2 Bimolecular reactions

Probably the most common reaction that occurs in cells is the bimolecular reaction. There are two different ways by which two reactants combine to form one or more products. In the reaction

$$S_1 + S_2 \xrightarrow{k} \text{ products}$$

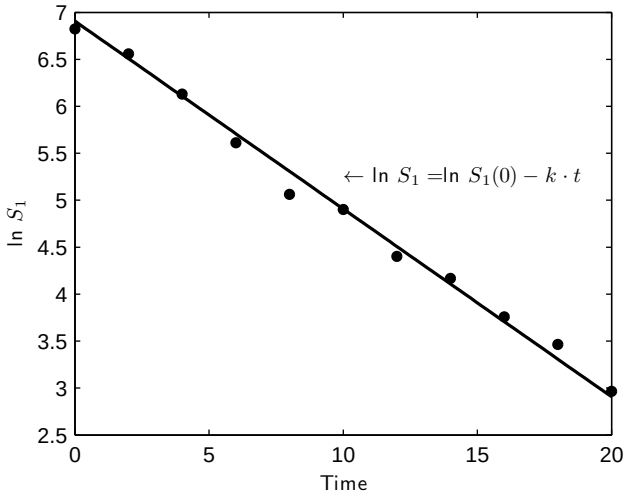


Figure 2.5: Measurement of the rate coefficient k for a monomolecular reaction. The logarithm of the concentration S_1 as function of time t is a straight line with slope $-k$. The initial concentration $S_1(0)$ is given by the intersection with the y-axis. The parameters are obtained by a linear fit to measured data.

two different species are transformed into products, but the reaction

$$2\,S_1 \xrightarrow{k} \text{ products}$$

is also possible. We will discuss both types of bimolecular reactions in the following two sections.

2.5.2.1 Bimolecular reactions of two different species

bimolecular reaction

The *bimolecular reaction* of two different molecular species S_1 and S_2 is defined by

$$S_1 + S_2 \xrightarrow{k} \text{ products}\ , \tag{2.26}$$

where k is the rate coefficient for the bimolecular reaction. The units are now dependent on the units of S_1 and S_2. If S_1 and S_2 are concentrations, the rate coefficient has the unit concentration per time. From *collision theory* we have for the reaction rate

collision theory

$$r(t) = -\frac{dS_1(t)}{dt} = -\frac{dS_2(t)}{dt} = k\,S_1(t)\,S_2(t)\ . \tag{2.27}$$

With the help of the extent variable $\epsilon(t)$, (2.12), we can transform this equation to

$$r(t) = \frac{d\epsilon(t)}{dt} = k\left[S_1(0) - \epsilon(t)\right]\left[S_2(0) - \epsilon(t)\right]\ , \tag{2.28}$$

with initial concentrations $S_1(0)$ and $S_2(0)$. The differential equation is now solvable by separation of variables. The analytical solution of the integral

$$kt = \int_0^{\epsilon(t)} \frac{d\epsilon(t)}{\left[S_1(0) - \epsilon(t)\right]\left[S_2(0) - \epsilon(t)\right]} \tag{2.29}$$

is obtained, if one uses the following expansion into partial fractions

$$\frac{1}{\left[S_1(0) - \epsilon(t)\right]\left[S_2(0) - \epsilon(t)\right]} = \frac{1}{S_1(0) - S_2(0)} \left[\frac{1}{S_2(0) - \epsilon(t)} - \frac{1}{S_1(0) - \epsilon(t)}\right] .$$

The result

$$\begin{aligned} kt &= \frac{1}{S_1(0) - S_2(0)} \ln \frac{S_2(0)}{S_1(0)} \frac{S_1(0) - \epsilon(t)}{S_2(0) - \epsilon(t)} \\ &= \frac{1}{S_1(0) - S_2(0)} \ln \frac{S_2(0)}{S_1(0)} \frac{S_1(t)}{S_2(t)} \end{aligned} \tag{2.30}$$

is transposable to $S_1(t)$ and $S_2(t)$ using the relation $\epsilon = S_1(0) - S_1(t) = S_2(0) - S_2(t)$. One obtains

$$S_1(t) = \frac{S_2(0) - S_1(0)}{S_2(0)/S_1(0)\, \exp\{[S_2(0) - S_1(0)]kt\} - 1} \tag{2.31}$$

and

$$S_2(t) = \frac{S_1(0) - S_2(0)}{S_1(0)/S_2(0)\, \exp\{[S_1(0) - S_2(0)]kt\} - 1} \tag{2.32}$$

as the time law for bimolecular reaction (2.26). The reactants decrease exponentially. The component abounding at the beginning is left over at the end of the reaction. If one plots the logarithmic term in (2.30) versus the reaction time t, one obtain a straight line with slope $-k\,(S_1(0) - S_2(0))$ allowing the measurement of the rate coefficient k from experiment and a linear regression.

Equations (2.31) and (2.32) can be simplified in the case of stoichiometric concentrations $S_1(0) = S_2(0)$. Under this conditions the relation $S_1(t) = S_2(t)$ holds at every time. The reaction rate (2.27) can rewritten as

$$r(t) = -\frac{dS_1(t)}{dt} = -\frac{dS_2(t)}{dt} = k\left[S_1(t)\right]^2 = k\left[S_2(t)\right]^2 \ , \tag{2.33}$$

which is again solvable by separation of variables. In this special case the time law is

$$S_1(t) = \frac{S_1(0)}{S_1(0)kt + 1} = \frac{S_2(0)}{S_2(0)kt + 1} = S_2(t) \ . \tag{2.34}$$

If one of the reacting species is in great excess, for instance $S_2(0) \gg S_1(0)$, the time law can be further simplified. The extent is controlled by the second component S_1. For the abounding species S_2 holds

$$S_2(t) = S_2(0) - \epsilon(t) \approx S_2(0) \tag{2.35}$$

while (2.12) is valid for $S_1(t)$. With these approximations we obtain the differential equation

$$\begin{aligned} r(t) &= -\frac{dS_1(t)}{dt} = k\, S_1(t)\, S_2(t) \approx k\, S_1(t)\, S_2(0) \\ &\approx k'\, S_1(t)\ , \end{aligned} \tag{2.36}$$

where the new effective coefficient k' is the product of the original rate coefficient k and the concentration $S_2(0)$. It follows, that for this case the time law can be reduced to the monomolecular case treated in the previous section. Such reactions are referred to as *pseudo-monomolecular* or *kryptobimolecular*.

2.5.3 Bimolecular reaction of identical species

If both reactants are from the same species

$$2\, S_1 \xrightarrow{k} \text{ products}\ , \tag{2.37}$$

we have a bimolecular reaction for two identical molecules. We write

$$r(t) = -\frac{1}{2}\frac{dS_1(t)}{dt} = k\,[S_1(t)]^2 \tag{2.38}$$

for the reaction rate. The prefactor $1/2$ ensures that one gets the same rate for the reactants and products. If one uses the change of concentration, this condition is, in general, not fulfilled.

The differential equation can be solved by separation of variables. For the dynamic concentration $S_1(t)$ one obtains

$$S_1(t) = \frac{S_1(0)}{2S_1(0)kt + 1}\ , \tag{2.39}$$

which is similar to (2.34). Both equations differ in a factor two in the denominator.

2.5.4 Trimolecular reactions

The reaction of three molecules to products is rare because the probability that three independent molecules collide at the same time or within a small time interval is very small. There are three possible ways for such a reaction. The first one is the reaction of three species

trimolecular reaction

$$S_1 + S_2 + S_3 \xrightarrow{k} \text{ products}$$

forming the products. In the second possibility

$$2S_1 + S_2 \xrightarrow{k} \text{ products}$$

two molecules of species S_1 react with a third particle S_2. Last but not least, three identical particles

$$3S_1 \xrightarrow{k} \text{ products}$$

can be transformed into the products.

2.5.4.1 Trimolecular reactions of different species

For the trimolecular reaction

$$S_1 + S_2 + S_3 \xrightarrow{k} \text{ products} \tag{2.40}$$

of three different species the reaction rate is

$$r(t) = -\frac{dS_1(t)}{dt} = -\frac{dS_2(t)}{dt} = -\frac{dS_3(t)}{dt} = k\, S_1(t)\, S_2(t)\, S_3(t)\,. \tag{2.41}$$

If one introduces the extent variable ϵ into the last equation

$$r(t) = \frac{d\epsilon(t)}{dt} = k\left[S_1(0) - \epsilon(t)\right]\left[S_2(0) - \epsilon(t)\right]\left[S_3(0) - \epsilon(t)\right]\ , \tag{2.42}$$

with initial concentrations $S_1(0)$, $S_2(0)$ and $S_3(0)$, it is possible to solve the differential equation by separation of variables. One obtains

$$\begin{aligned} kt \;=\; & \frac{1}{\left[S_1(0) - S_2(0)\right]\left[S_3(0) - S_1(0)\right]} \ln\frac{S_1(t)}{S_1(0)} \\ & + \frac{1}{\left[S_1(0) - S_2(0)\right]\left[S_2(0) - S_3(0)\right]} \ln\frac{S_2(t)}{S_2(0)} \\ & + \frac{1}{\left[S_2(0) - S_3(0)\right]\left[S_3(0) - S_1(0)\right]} \ln\frac{S_3(t)}{S_3(0)}\ . \end{aligned} \tag{2.43}$$

If all participating species have stoichiometric concentrations[6] we can simplify the approach ('ansatz'). Another known example is the so called 'product ansatz', where we assume that the solution of two parameter-dependent problem is separable into a product of two terms depending on one parameter only. In contrast, collision theory is an approach to the description of the temporal change of reacting species. For the reaction rate (2.41) with the relation $S_1(t) = S_2(t) = S_3(t)$ holding for all time. The resulting differential equation

$$r(t) = -\frac{dS_1(t)}{dt} = k\left[S_1(t)\right]^3 \tag{2.44}$$

is easy to solve and one obtains for the concentrations

$$S_1(t) = \sqrt{\frac{S_1(0)^2}{1 + 2S_1(0)^2 kt}} \tag{2.45}$$

as function of time. The results for $S_2(t)$ and $S_3(t)$ follow, if we exchange S_1 for S_2 and S_3 in the same way.

[6] The ratio of concentrations is equal to the ratio of stoichiometric coefficients.

2.5.4.2 Trimolecular reactions of two different species

The trimolecular reaction

$$2S_1 + S_2 \xrightarrow{k} \text{ products} \tag{2.46}$$

describes the reaction of two molecules of species S_1 with one particle of species S_2 into products. For the reaction rate

$$r(t) = -\frac{1}{2}\frac{dS_1(t)}{dt} = -\frac{dS_2(t)}{dt} = k\left[S_1(t)\right]^2 S_2(t) \ . \tag{2.47}$$

Again, one substitutes the time dependent variables $S_1(t)$ and $S_2(t)$ by the corresponding relation between the initial values and the extent $\epsilon(t)$. The resulting differential equation

$$r(t) = \frac{d\epsilon(t)}{dt} = k\left[S_1(0) - \epsilon(t)\right]^2\left[S_2(0) - \epsilon(t)\right] \tag{2.48}$$

can be solved by separation of variables. The result

$$kt = \frac{1}{2S_2(0) - S_1(0)}\left[\frac{S_1(0) - S_1(t)}{S_1(0)S_1(t)} + \frac{1}{2S_2(0) - S_1(0)}\right] \tag{2.49}$$

can be simplified, if one assumes stoichiometric concentrations. Then the relation $S_1(t) = 2S_2(t)$ has to be satisfied for all time. The differential equation is now

$$r(t) = -\frac{1}{2}\frac{dS_1(t)}{dt} = \frac{k\left[S_1(t)\right]^3}{2} \ . \tag{2.50}$$

Thus, the temporal evolution of the concentration of species S_1 obeys the same time law (2.45) as the species in the trimolecular reaction of three different particle for stoichiometric conditions. With the aim of the above relation between S_1 and S_2 one obtains

$$S_2(t) = \frac{1}{2}\sqrt{\frac{S_2(0)^2}{1 + S_2(0)^2 kt/2}} \tag{2.51}$$

for the second species S_2.

2.5.4.3 Trimolecular reactions of three identical molecules

The reaction of three molecules of the same species

$$3S_1 \xrightarrow{k} \text{ products}, \tag{2.52}$$

is the third possibility for a realisation of a trimolecular reaction. The reaction rate is given as

$$r(t) = -\frac{1}{3}\frac{dS_1(t)}{dt} = k\left[S_1(t)\right]^3 \ . \tag{2.53}$$

It follows

$$S_1(t) = \sqrt{\frac{S_1(0)^2}{1 + 6S_1(0)^2 kt}} \tag{2.54}$$

for the concentration of species S_1 as a function of time.

2.5.5 Higher and rational reaction orders

Up to now, we treated elementary reactions with one, two or three participating molecular species. The developed formalism shall expand to non-elementary reactions. For this purpose we revert to the general chemical reaction

$$|l_1|S_1 + |l_2|S_2 + + \ldots + |l_i|S_i \xrightarrow{k} |l_{i+1}|S_{i+1} + |l_{i+2}|S_{i+2} + \ldots \tag{2.55}$$

and introduce the reaction rate as

$$\begin{aligned} r(t) &= k\,[S_1(t)]^{n_1}\,[S_2(t)]^{n_2}\ldots[S_i(t)]^{n_i}\,[S_{i+1}(t)]^{n_{i+1}}\,[S_{i+2}(t)]^{n_{i+2}}\ldots \\ &= k\prod_i [S_i(t)]^{n_i}\ . \end{aligned} \tag{2.56}$$

The total reaction order then is defined as sum over partial orders

$$n = \sum_i n_i\ , \tag{2.57}$$

which can take values greater than three, as well as rational values. Therefore, the so described reaction is in general not an elementary reaction. The partial reaction orders n_i of each species, in general, do not coincide with the stoichiometric coefficients. The reaction rate obtained is an approximation and does not reflect the true reaction mechanism, but often it gives a first indication to the mechanism.

If one assumes stoichiometric concentrations, one is able to obtain some important quantities. The reaction rate is then

$$\frac{1}{l_1}\frac{dS_1(t)}{dt} = \frac{1}{l_2}\frac{dS_2(t)}{dt} = \ldots = \frac{d\epsilon(t)}{dt} = k\left[S_1(0) - \epsilon(t)\right]^n\ , \tag{2.58}$$

where $\epsilon(t)$ is the extent variable. By separation of variables one obtains

$$\frac{S_1(t)}{S_1(0)} = \left(\frac{1}{1 + S_1(0)^{n-1}\,(n-1)\,kt}\right)^{1/(n-1)}\ , \tag{2.59}$$

for the normalised or dimensionless concentration $S_1(t)/S_1(0)$. The half life, defined as the time, where the normalised concentration is $S_1(t)/S_1(0) = 1/2$, can be determined from (2.59). After some algebraic transformations one obtains

$$t_{1/2} = \frac{2^{n-1}-1}{(n-1)k}\,S_1(0)^{-n+1}\,. \tag{2.60}$$

The dependence from the initial concentration $S_1(0)$ in (2.60) allows the determination of the total reaction order n. In an experiment the half life can be measured as function of the initial concentration. A double logarithmic representation of the data gives a straight line with a slope $m = -n+1$, which can be used to determine the total reaction order n. For such practical considerations and experimental techniques the reader is referred to [AdP02, Seg93, CB04].

reaction order	half life $t_{1/2}$
1	$t_{1/2} = \text{const.}$
2	$t_{1/2} \sim S_1(0)^{-1}$
3	$t_{1/2} \sim S_1(0)^{-2}$
1/2	$t_{1/2} \sim \sqrt{S_1(0)}$
3/2	$t_{1/2} \sim 1/\sqrt{S_1(0)}$

Table 2.1: The dependency of the half life $t_{1/2}$ from the initial concentration for the general chemical reaction (2.55) with stoichiometric conditions. $S_1(0)$ is the initial concentration of reactant S_1.

2.5.5.1 Reactions of zeroth order

A special case of the generalised reaction is the reaction of "zeroth order". For such a reaction the reaction rate is independent from the concentration. Choosing a zeroth order degradation of species S_1 as an example, the differential equation

$$-\frac{dS_1(t)}{dt} = k \tag{2.61}$$

is easy to integrate. The result is the linear function

$$S_1(t) = S_1(0) - kt\,, \tag{2.62}$$

where the slope is given by the rate coefficient k. Reactions of zeroth order appear, if the rate is governed by a temporal constant non-chemical process. Examples include reactions on a surface, where the concentration of the reactant is constant by adsorption or a constant external flow of matter. The saturation of the enzyme complex in the enzyme kinetic reaction is a further example for such a behavior. Often, this approximation is used to simplify kinetic equations. The transient time, the reaction takes to reach this state and at the end of the reaction, is assumed as small compared to the saturated state.

At the end of this section we have to made an important comment on this treatment of (bio)chemical reactions. Within this framework one describes the dependency between the reactants and the products without a proper model of the reaction mechanism. Hence, the total order n cannot be interpreted as number of participating molecules in an elementary reaction. But it is an easy and fast way to find a functional relation between the reactants and the products.

2.6 Complex Reactions

In the previous section we introduced the concept of elementary reactions and demonstrated their properties. As mentioned before, we assume that chemical reactions consist of a set of elementary reactions. In the this section we want to classify some basic complex reactions and describe their properties. Before we go into details, we distinguish between three basic kinds of "complex reactions":

1. Reversible reactions,
$$S_1 \rightleftharpoons S_2\,,$$
2. Parallel reactions,
$$\begin{array}{rcl} & \nearrow & S_2 \\ S_1 & \rightarrow & S_3 \\ & \searrow & S_4 \end{array}$$
3. Consecutive reactions,
$$S_1 \rightarrow S_2 \rightarrow S_3\ .$$

More complex reactions can be composed of these three basic classes. Some possible reaction schemes are:

- Consecutive reactions with reversible parts
  - pre-equilibrium
$$S_1 \rightleftharpoons S_2 \rightarrow S_3 \rightarrow S_4$$
  - downstream equilibrium
$$S_1 \rightarrow S_2 \rightleftharpoons S_3$$
- Consecutive reactions combined with parallel reactions
$$\begin{array}{rcl} & \nearrow & S_3 \\ S_1 \rightarrow S_2 & \rightarrow & S_4 \\ & \searrow & S_5 \rightarrow S_6 \end{array}$$
- Competitive consecutive reactions
$$\begin{aligned} S_1 + S_2 &\rightarrow S_3 + S_4 \\ S_1 + S_3 &\rightarrow S_5 + S_6 \end{aligned}$$
- Closed consecutive reactions (chain reactions)
$$\begin{aligned} S_1 &\rightarrow 2X \\ S_1 + X &\rightarrow S_2 \rightarrow S_3 + X \end{aligned}$$

We can continue this list with more complicated complex reactions, but in order to convey an idea of the complexity of reaction mechanisms this short overview is sufficient.

Each step in a complex reaction scheme is represented by an elementary reaction, which can be described by a differential equation. Because steps are not independent of each other, we now obtain a system of coupled differential equations. In general, there is no analytical solution for such systems. One has to use numerical methods to solve, simulate, the differential equations. We provide some analytically solvable simple examples to demonstrate the properties of more complex reactions.

2.6.1 Reversible reactions

reversible reaction

A *reversible reaction* consists of two elementary reactions, the forward reaction and the reverse reaction. Both are characterised by rate coefficients. The simplest example for such a reaction is the monomolecular reversible reaction

$$S_1 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} S_2 \ , \tag{2.63}$$

where the forward and the backward reaction are first-order reactions. The rate coefficient of the forward reaction is k_1 and the coefficient of the backward reaction is k_{-1}. If we assume the initial concentration $S_1(0) \doteq S_0$ and $S_2(0) = 0$ the conservation law

$$S_1(t) + S_2(t) = S_1(0) \tag{2.64}$$

must hold for all times. The corresponding reaction rate is obtained from the difference

$$r(t) = r_1(t) - r_{-1}(t) \tag{2.65}$$

of the forward and the backward reaction. With the use of the representation of elementary reactions we obtain the differential equations

$$\begin{aligned} r &= -\frac{dS_1(t)}{dt} = k_1\, S_1(t) - k_{-1}\, S_2(t) \\ &= \frac{dS_2(t)}{dt} = k_1\, S_1(t) - k_{-1}\, S_2(t) = -\frac{dS_1(t)}{dt} \\ &= \frac{d\epsilon(t)}{dt} = k_1 \left[S_1(0) - \epsilon(t)\right] - k_{-1} \left[S_2(0) + \epsilon(t)\right] \end{aligned} \tag{2.66}$$

for the reaction rate. The first term on the ride-hand-side corresponds to the production of S_2 in a monomolecular reaction from S_1 and the second term to the analogue reverse reaction. From (2.66) it follows, that there is a point, where the reaction rate is zero. The forward and the backward reaction are balanced at this point, the same amount of S_2 is produced the same amount of S_2 as is transformed back to S_1. From this we see that a reaction rate of zero does not mean, that nothing happens[7]. A net change of concentration is not measurable from a macroscopic point of view, but in a microscopic view the reactions are still going on. This special state is called *chemical equilibrium*. For our example of a reversible reaction, this state is defined as

chemical equilibrium

$$0 = k_1 S_1(t) - k_{-1} S_2(t)\,. \tag{2.67}$$

More general, the right-hand-side must be zero. This mathematical condition is a necessary but not a sufficient constraint, especially for complex systems. It has to fulfill some more physical conditions, by which we decide if it is a stable or an unstable state. Only

[7] Except at time $t = 0$, where the reaction was started. At this point, we postulate that nothing happened before.

the stable state is referred to as equilibrium state (see also Chapter 4.9). The solution of equation (2.67) is

$$K_{\mathrm{eq}} = \frac{k_1}{k_{-1}} = \frac{S_{2,\mathrm{eq}}}{S_{1,\mathrm{eq}}} \tag{2.68}$$

defines the equilibrium constant K_{eq} as the ratio of the rate coefficient of the forward and the rate coefficient of the backward reaction. This quantity measures the affinity of S_1 to transform to S_2. The equation (2.68) is the famous "law of mass action" for an monomolecular reaction. Equations (2.66) are kinetic rate equations. The *law of mass action* is a result of these equations and the assumptions behind them. To call this representations 'law of mass action'" can however be misleading and one might speak of a 'generalised law of mass action' is sometimes used. For S_2 we define the dissociation constant K_{d} describing the process of the backward transformation into the reactants. It is the reciprocal of the equilibrium constant K_{eq}. From (2.68) it follows that in equilibrium the ratio of the concentrations match the equilibrium constant. The corresponding concentrations are called equilibrium concentrations $S_{1,\mathrm{eq}}$ and $S_{2,\mathrm{eq}}$. The species with a higher production rate has the higher equilibrium concentration.

law of mass action

The temporal evolution of the reversible reaction (2.63) can be solved analytically using the third differential equation in (2.66). The integration over the extent variable $\epsilon(t)$ can be carried out by separation of variables. We obtain the integrals

$$\int_0^{\epsilon(t)} d\epsilon(t) \left[(K_{\mathrm{eq}} + 1) \left(\frac{K_{\mathrm{eq}} S_1(0) - S_2(0)}{K_{\mathrm{eq}} + 1} - \epsilon \right) \right]^{-1} = \int_0^t k_{-1}\, dt \ . \tag{2.69}$$

After integration and some manipulations we get the result

$$\begin{aligned} \epsilon(t) &= \frac{k_1 S_1(0) - k_{-1} S_2(0)}{k_1 + k_{-1}} \left[1 - \exp\left\{ -(k_1 + k_{-1})t \right\} \right] \\ &= \epsilon_{\mathrm{eq}} \left[1 - \exp\left\{ -(k_1 + k_{-1})t \right\} \right] \end{aligned} \tag{2.70}$$

for the extent variable. The extent increases exponentially to its equilibrium value. The relaxation time $\tau = (k_1 + k_{-1})^{-1}$ is a measure how fast the reaction tends to the equilibrium. The time evolution for the reactant S_1 follows as

$$S_1(t) = S_{1,\mathrm{eq}} + (S_1(0) - S_{1,\mathrm{eq}}) \exp\left\{ -(k_1 + k_{-1})t \right\} \ , \tag{2.71}$$

with initial concentration $S_1(0)$ and equilibrium concentration $S_{1,\mathrm{eq}}$. From the conservation law

$$S_1(t) + S_2(t) = S_1(0) \ ,$$

where we assume that $S_2(0) = 0$, we obtain

$$S_2(t) = (S_1(0) - S_{1,\mathrm{eq}}) \left[1 - \exp\left\{ -(k_1 + k_{-1})t \right\} \right] \tag{2.72}$$

for the product. Both, reactant and products, reach their equilibrium concentration, exponentially. This behavior is shown in Figure 2.6. Additionally, the equilibrium values

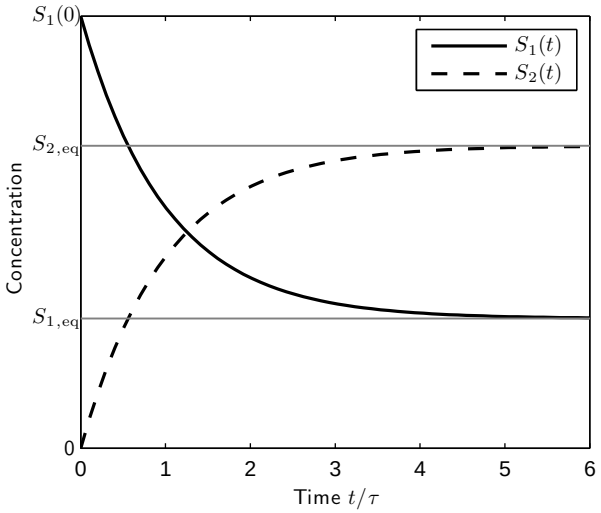


Figure 2.6: The monomolecular reversible reaction as function of the relaxation time τ. The solid line represent the dynamic change of the reactant and the dashed line the change of the product to their equilibrium states shown as horizontal lines. The ratio of the equilibrium states is given by the law of mass action (2.68).

are drawn as horizontal lines for comparison. We choose $k_1 > k_{-1}$ for this example, hence the equilibrium state of S_2 is higher than the equilibrium state of the reactant S_1.

The rate coefficients of a monomolecular reversible reaction can be determined from a logarithmic representation of $(S_1(0) - S_{1,\mathrm{eq}})/(S_1(t) - S_{1,\mathrm{eq}})$ and the equilibrium with (2.68). The slope of the logarithmic plot is proportional to the inverse relaxation time $(k_1 + k_{-1})$. The determination of the rate coefficients k_1 and k_{-1} requires the measurement of the dynamic change of concentrations and the measurement of equilibrium data.

The treatment of reversible reactions can be generalised to higher reaction orders. As an example we choose a bimolecular forward- and backward reaction

$$S_1 + S_2 \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} S_3 + S_4 \ . \tag{2.73}$$

We assume, that only the reactants are present in stoichiometric amounts at the beginning of the reaction. Furthermore we use the extent variable $\epsilon(t)$ to simplify the kinetic equation.

t	S_1	S_2	S_3	S_4
0	$S_1(0)$	$S_1(0)$	0	0
t	$S_1(0) - \epsilon(t)$	$S_1(0) - \epsilon(t)$	$\epsilon(t)$	$\epsilon(t)$
∞	$S_1(0) - \epsilon_{\mathrm{eq}}$	$S_1(0) - \epsilon_{\mathrm{eq}}$	ϵ_{eq}	ϵ_{eq}

Table 2.2: Relations between the reactants and products to the extent variable ϵ for the reversible reaction (2.73).

From this follows the conservation law

conservation law

$$S_1(0) + S_2(0) = 2S_1(0) = 2\left(S_{1,\mathrm{eq}} + \epsilon_{\mathrm{eq}}\right) \ . \tag{2.74}$$

The time law is

$$\begin{aligned} r(t) &= -\frac{dS_1(t)}{dt} = -\frac{dS_2(t)}{dt} = \frac{d\epsilon(t)}{dt} \\ &= k_2\, S_1(t)\, S_2(t) - k_{-2}\, S_3(t)\, S_4(t) \\ &= k_2\left[S_1(0) - \epsilon(t)\right]^2 - k_{-2}\,\left[\epsilon(t)\right]^2 \end{aligned} \tag{2.75}$$

is analytically solvable. The integration by expansion into partial fractions gives

$$\ln \frac{\epsilon(t)\left[S_1(0) - 2\epsilon_{\mathrm{eq}}\right] + S_1(0)\epsilon_{\mathrm{eq}}}{S_1(0)\left[\epsilon_{\mathrm{eq}} - \epsilon(t)\right]} = \frac{2S_1(0)\left[S_1(0) - \epsilon_{\mathrm{eq}}\right]}{\epsilon_{\mathrm{eq}}}\, k_2 t \tag{2.76}$$

after numerous rearrangements. We can use it as an instruction for the determination of the rate coefficient k_2. The second coefficient is obtained again from the equilibrium solution

$$K_{\mathrm{eq}} = \frac{S_{3,\mathrm{eq}} S_{4,\mathrm{eq}}}{S_{1,\mathrm{eq}} S_{2,\mathrm{eq}}} = \frac{\epsilon_{\mathrm{eq}}^2}{\left[S_1(0) - \epsilon_{\mathrm{eq}}\right]^2} = \frac{k_2}{k_{-2}}\,, \tag{2.77}$$

as a generalisation of the law of mass action for the bimolecular reversible reaction (2.73).

2.6.2 Parallel reactions

In a *parallel reaction*, several reactions of the same reactants proceed side by side but produce different end products. Each reaction mechanism can consist of an elementary reaction, a reversible reaction, or a more complex mechanism. The reaction of one species with several partners in the reaction volume is a competitive reaction and no parallel reaction. We will discuss this towards the end of this section.

parallel reaction

In order to illustrate some elementary properties of side reactions we consider the first-order reaction

$$\begin{array}{rcl} & \xrightarrow{k_1} & S_2 \\ S_1 & \xrightarrow{k_2} & S_3 \\ & \xrightarrow{k_3} & S_4 \end{array} \tag{2.78}$$

composed of three irreversible monomolecular reactions. The species S_1 is converted in a monomolecular reaction into products S_2, S_3, and S_4. The term $k_i\, dt$ is the probability that the i-th reaction occurs in the time interval dt. The resulting kinetic equation is

$$\begin{aligned} r(t) &= -\frac{dS_1(t)}{dt} = k_1\, S_1(t) + k_2\, S_1(t) + k_3\, S_1(t) \\ &= k\, S_1(t)\ , \end{aligned} \tag{2.79}$$

where k is the total rate coefficient

$$k = \sum_i k_i \tag{2.80}$$

and $k\,dt$ the probability, that one of the reactions takes place. The solution of such a differential equation is known from the discussion of monomolecular reaction. In analogy, we obtain for the temporal evolution an exponential expression

$$S_1(t) = S_1(0)\,\exp\{-kt\}\,, \tag{2.81}$$

where $S_1(0)$ is the initial concentration. The temporal evolution of the reactant is indistinguishable from the evolution of a monomolecular reaction with the same rate coefficient. More generally, the dynamic change of the reactants is for the same coefficient k independent of the number of products.

The creation of products S_2, S_3, and S_4 is governed by the kinetic equations

$$\frac{dS_2(t)}{dt} = k_1\,S_1(t)\;, \qquad \frac{dS_3(t)}{dt} = k_2\,S_1(t)\;, \qquad \frac{dS_4(t)}{dt} = k_3\,S_1(t)\;, \tag{2.82}$$

which can be transformed with (2.81) into

$$\frac{dS_2(t)}{dt} = k_1\,S_1(0)\,\exp\{-kt\}\;, \qquad \ldots \tag{2.83}$$

Integration leads the exponentials

$$S_2(t) = \frac{k_1}{k}\left[1-\exp\{-kt\}\right]S_1(0) + S_2(0)\;, \tag{2.84}$$

$$S_3(t) = \frac{k_2}{k}\left[1-\exp\{-kt\}\right]S_1(0) + S_3(0)\;, \tag{2.85}$$

$$S_4(t) = \frac{k_3}{k}\left[1-\exp\{-kt\}\right]S_1(0) + S_4(0)\;. \tag{2.86}$$

The comparison of the transient concentrations[8] shows a further important property of parallel reactions. If we calculate the ratios of the dynamic concentrations

$$[S_2(t)-S_2(0)] : [S_3(t)-S_3(0)] : [S_4(t)-S_4(0)] = k_1 : k_2 : k_3 \tag{2.87}$$

we obtain the 'Principle of Wegscheider'. It says, that the ratio is equal to the ratio of the rate coefficients k_i and constant.

The time evolution for the parallel reaction (2.78) is shown in Figure 2.7, for the special case of $S_2(0) = S_3(0) = S_4(0) = 0$. For these initial conditions we can expand the conservation law to

$$S_1(0) = S_1(t) + S_2(t) + S_3(t) + S_4(t)\;. \tag{2.88}$$

[8] The prefix transient distinguishes between the produced concentration in the temporal process of the reaction and the initial concentrations. In this sense the full concentration is given by a transient and an initial part.

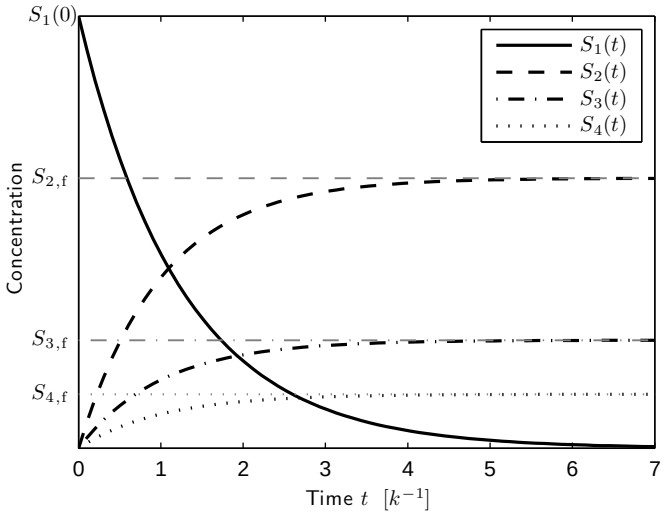


Figure 2.7: The temporal evolution of the reactant S_1 and the products S_2, S_3, S_4 of the parallel reaction (2.78) in units of the initial concentration $S_1(0)$. The horizontal lines denote the final concentration $S_{2,\mathrm{f}}$, $S_{3,\mathrm{f}}$, and $S_{4,\mathrm{f}}$ of the products.

At the end of the reaction, species S_1 is completely consumed and (2.88) is determined by the final concentration $S_{2,\mathrm{f}}$, $S_{3,\mathrm{f}}$, and $S_{4,\mathrm{f}}$. We obtain with (2.84)-(2.86) for these concentrations

$$S_{2,\mathrm{f}} = \frac{k_1}{k} S_1(0) \ ,$$
$$S_{3,\mathrm{f}} = \frac{k_2}{k} S_1(0) \ ,$$
$$S_{4,\mathrm{f}} = \frac{k_3}{k} S_1(0) \ .$$

The concentrations are determined by the ratio of the individual rate coefficient k_i and the total rate coefficient k (2.80). The exponential decrease of S_1 and the exponential increase of the product is clearly visible in the figure. For comparison, the product concentrations at the end of the reaction are also shown. The reaction $S_1 \to S_2$ is the fastest reaction, hence it is the main product of the side reaction (2.78).

These results are valid for parallel reactions in general. Recapitulating, parallel reactions have the following behavior

- The time evolution of the reactants is independent from the number of products and has the same order as each elementary reaction. It is determined by the total rate coefficient (2.80), only.
- The fastest side reaction is strongest participant on the extent. It controls the main product.
- The products are build in the ratio of their rate coefficient (Principle of Wegscheider).

2.6.2.1 Differentiation of parallel reactions and competitive reactions

In our discussion of parallel reactions we tried to show how a parallel reaction can be separated from a competitive reaction in an experiment. In a parallel reaction the same reactants can be transformed in different products. In a competitive reaction, for instance, the reactant S_1 participates in two independent reactions with different partners. For instance, let us consider two bimolecular reactions

$$\begin{aligned} S_1 + S_2 &\xrightarrow{k_1} S_4 + \dots \\ S_1 + S_3 &\xrightarrow{k_2} S_5 + \dots \end{aligned} \tag{2.89}$$

with the common reactant S_1. The corresponding system of coupled differential equations is

$$\begin{aligned} \frac{dS_1(t)}{dt} &= -\left(k_1\, S_2(t) + k_2\, S_3(t)\right) S_1(t)\,, \\ \frac{dS_2(t)}{dt} &= -k_1\, S_1(t)\, S_2(t)\,, \\ \frac{dS_3(t)}{dt} &= -k_2\, S_1(t)\, S_3(t)\,, \\ \frac{dS_4(t)}{dt} &= k_1\, S_1(t)\, S_2(t)\,, \\ \frac{dS_5(t)}{dt} &= k_2\, S_1(t)\, S_3(t)\,. \end{aligned} \tag{2.90}$$

If we use Wegscheider's principle in differential form, we obtain with

$$\frac{dS_4(t)}{dS_5(t)} = \frac{k_1\, S_1(t)\, S_2(t)}{k_2\, S_1(t)\, S_3(t)} = \frac{k_1}{k_2}\,\frac{S_2(t)}{S_3(t)} \tag{2.91}$$

an expression that depends on the present concentration of S_2 and S_3. For a parallel reaction this ratio has to be a constant.

2.6.3 Consecutive reactions

consecutive reaction

An important class of complex reactions are *consecutive reactions*. These reactions include one or more unstable intermediates. Some partial reactions can be reversible and there can be arborisation as a result of parallel reaction.

To simplify matters we consider an unidirectional and monomolecular sequence of reactions

$$S_1 \xrightarrow{k_1} S_2 \xrightarrow{k_2} S_3\,, \tag{2.92}$$

which is like a radioactive decay sequence, with initial condition $S_2(0) = S_3(0) = 0$. The decay of S_1 into the intermediate S_2 is governed by the differential equation

$$\frac{dS_1(t)}{dt} = -k_1\, S_1(t)\ , \tag{2.93}$$

with the known solution

$$S_1(t) = S_1(0)\, \exp\{-k_1\, t\} \ . \tag{2.94}$$

The differential equation for S_2 is more complicated. It consists of two parts

$$\frac{dS_2(t)}{dt} = k_1\, S_1(t) - k_2\, S_2(t) \ , \tag{2.95}$$

where the first term describes the production of S_2 from S_1 and the second term the decay into the final product S_3. With (2.94) we transform this equation into a first-order linear differential equation that can be solved analytically:

$$\frac{dS_2(t)}{dt} = k_1\, S_1(0)\, \exp\{-k_1\, t\} - k_2\, S_2(t) \ . \tag{2.96}$$

We obtain for S_2

$$S_2(t) = \frac{k_1}{k_2 - k_1}\, S_1(0)\, \left[\exp\{-k_1 t\} - \exp\{-k_2 t\}\right] \ . \tag{2.97}$$

The temporal evolution has a maximum at

$$t_{\max} = \frac{\ln(k_1/k_2)}{k_1 - k_2} \ . \tag{2.98}$$

This is a typical property for an unstable intermediate. In the beginning S_1 decays faster due to its higher concentration, whereas S_2 decays slowly. The result is an increase of the intermediate. In the course of the reaction the rate of decay of S_1 decreases. At $t_{\max}$ the production rate and the decay rate of the intermediate are equal. After this time the decay rate is faster and S_2 decreases, too. The corresponding concentration is

$$S_{2,\max} = S_2(t_{\max}) = \left(\frac{k_1}{k_2}\right)^{\frac{k_2}{k_2 - k_1}} \ . \tag{2.99}$$

If we use (2.97) the differential equation for the final product is

$$\begin{aligned} \frac{dS_3(t)}{dt} &= k_2\, S_2(t) \\ &= \frac{k_1\, k_2}{k_2 - k_1}\, S_1(0)\, \left[\exp\{-k_1 t\} - \exp\{-k_2 t\}\right] \ . \end{aligned} \tag{2.100}$$

It is proportional to the time law of S_2. Hence, the reaction rate of S_3 has a maximum at $t_{\max}$. Remember we assumed $S_2(0) = S_3(0) = 0$, the conservation law

$$S_1(0) = S_1(t) + S_2(t) + S_3(t) \ , \tag{2.101}$$

permits an easier way to solve the temporal evolution. The result

$$S_3(t) = S_1(0) \left[1 - \frac{k_2\, \exp\{-k_1\, t\} - k_1 \exp\{-k_2\, t\}}{k_2 - k_1}\right] \tag{2.102}$$

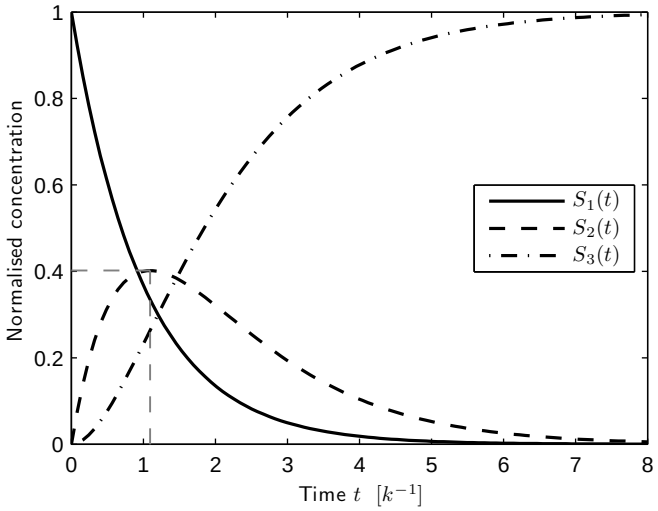


Figure 2.8: The temporal evolution of consecutive reaction (2.92) as a function of the dimensionless time $k_1\,t$. On the y-axis the normalised concentrations $S_1(t)/S_1(0)$, $S_2(t)/S_1(0)$, and $S_3(t)/S_1(0)$ are plotted. Additionally the time $t_{\max}$ (2.98) and the maximum value of the intermediate $S_{2,\max}$ (2.99) are shown.

changes the sign of its second derivative at $t_{\max}$. This behavior results in a typical sigmoidal shape of the time evolution of the final product, as shown in Figure 2.8. There the time evolution for all three participating species is compared. The starting substance S_1 decreases with the exponential decay law (2.94). As mentioned before, intermediate S_2 first increases, goes through a maximum and finally decreases. The time at which the maximum occurs and its value are plotted by thin dashed lines. Both are dependent on the ratio of the rate coefficient k_1 and k_2. The time evolution of the final product S_3 is a monotonously increasing function with the predicted inflection point at $t_{\max}$.

The radioactive decay sequence (2.92) is a simple example for a consecutive reaction. Reactions of higher order or/and higher complexity lead to more complicated kinetic equations and are often only numerically or approximately solvable.

2.6.3.1 Rate-determining steps

rate-determining step

If one reaction is much slower than subsequent ones, it can determine the overall rate of the reaction. In some cases we can then simplify the formalism for a consecutive reaction. We consider again the radioactive decay sequence (2.92). First, let us assume the case where $k_1 \gg k_2$, i.e., the second reaction is slower than the first. Now we expand

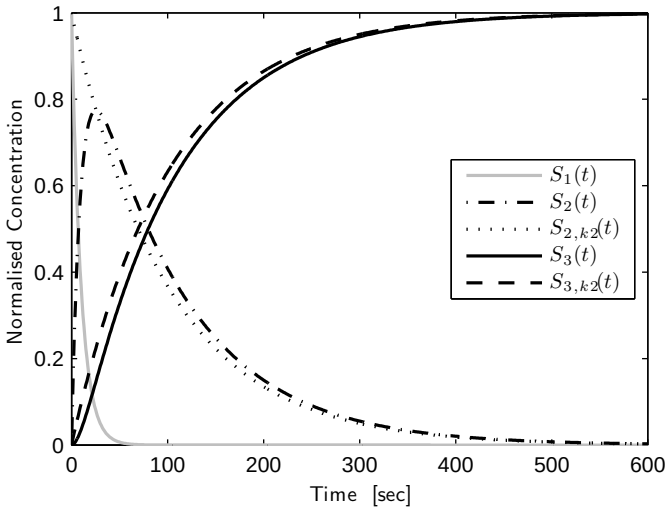


Figure 2.9: Comparison of the full solution for the consecutive reaction (2.92) and the approximation of the rate-determining step for a slow second partial reaction. After a short starting time, the approximation shows good agreement with the full solution. We choose $k_1 = 0.1\,\mathrm{s}^{-1}$ and $k_2 = 0.01\,\mathrm{s}^{-1}$ for this example. The concentrations are normalised to the initial concentration $S_1(0)$.

the prefactor[9] in equation (2.97)

$$\frac{k_1}{k_2 - k_1} = -\frac{1}{1 - k_2/k_1} \approx \left(-1 - \frac{k_2}{k_1} - \mathcal{O}\left(\frac{k_2^2}{k_1^2}\right)\right) \approx -1\,, \tag{2.103}$$

where the symbol $\mathcal{O}(\ldots)$ denotes the order in respect of the expansion parameter of first neglected term of the expansion. The sign corresponds to the trend of contribution. Furthermore we compare the exponential functions within the brackets. If $k_1\,t$ is always much greater then $k_2\,t$, we obtain

$$\exp\{-k_2\,t\} \gg \exp\{-k_1\,t\} \tag{2.104}$$

and neglect the k_1 dependent exponential function against the k_2 dependent exponential. With these approximations we obtain for the intermediate the new time law

$$S_{2,k_2}(t) = S_1(0)\exp\{-k_2\,t\} \tag{2.105}$$

and for the final product

$$S_{3,k_2}(t) = S_1(0)\left[1 - \exp\{-k_2\,t\}\right]\,. \tag{2.106}$$

[9]
$$\frac{1}{1-x} \approx 1 + x + x^2 + \ldots \qquad \text{for} \qquad x \ll 1$$

Within this approximation, the temporal evolution is determined by the slow decay of the intermediate S_2 into the product S_3. The subscript k_2 shall denote this property and distinguish between the full and the approximative solution. We compare both solutions in Figure 2.9, where our main focus is on the final product. The decay of S_1 is kept unchanged. It is almost completely transformed into the intermediate before the intermediate decays into S_3. After a short starting time, the principle of the rate-limiting step is a good approximation.

In a second part we now assume, that the decay of S_1 is the slowest step of sequence reaction (2.92). This means $k_1 \ll k_2$. We again expand the prefactor

$$\frac{k_1}{k_2 - k_1} = \frac{k_1}{k_2} \frac{1}{1 - k_1/k_2} \approx \frac{k_1}{k_2} \left[1 + \frac{k_1}{k_2} + \mathcal{O}\left(\frac{k_1^2}{k_2^2} \right) \right] \approx \frac{k_1}{k_2} \tag{2.107}$$

and the comparison of the exponentials gets

$$\exp\{-k_1\, t\} \gg \exp\{-k_2\, t\}\ . \tag{2.108}$$

We can now simplify equations (2.97) and (2.102), whereas the equation for S_1 (2.94) remains unchanged. Within this approximation the evolution of the intermediate is governed by

$$S_{2,k_1}(t) = S_1(0)\, \frac{k_1}{k_2}\, \exp\{-k_1\, t\} \tag{2.109}$$

and of the final product by

$$S_{3,k_1}(t) = S_1(0) \left[1 - \exp\{-k_1\, t\} \right] = S_1(0) - S_1(t)\,. \tag{2.110}$$

Again, the subscript denotes the rate-determining step. The comparison of the approximation and the full solution is shown in Figure 2.10. The production of the intermediate is much slower than its decay. A molecule from species S_2 is transformed practically immediately into the final product. The concentration of the intermediate is always small in comparison to S_1 and S_3. Also in this case, the approximation gives a good description of the reaction except the very first time.

We now give a more general discussion of the rate limiting step. If one step of a consecutive reaction is much slower than the others, it determines the total rate of reaction or more precisely, the rate of all following steps. This reaction is called the rate-determining step. The total reaction rate is the rate of the production of the final product of the sequence. This statement includes a further property of consecutive reactions. The reaction rate is not equal for each step. Each partial reaction before the rate-determining step has its own rate depending on the specific reaction and the steps before. The approximation of the rate-determining step defines a limiting total reaction rate for the considered sequence. The reaction rate for the described system cannot be faster than this limit. This is a direct consequence from the principle of cause and effect. The effect cannot occur before the cause, you cannot create a new species without its components. If there are faster steps before the rate-determining step we obtain a congestion of its participating molecules. If the following steps are faster they

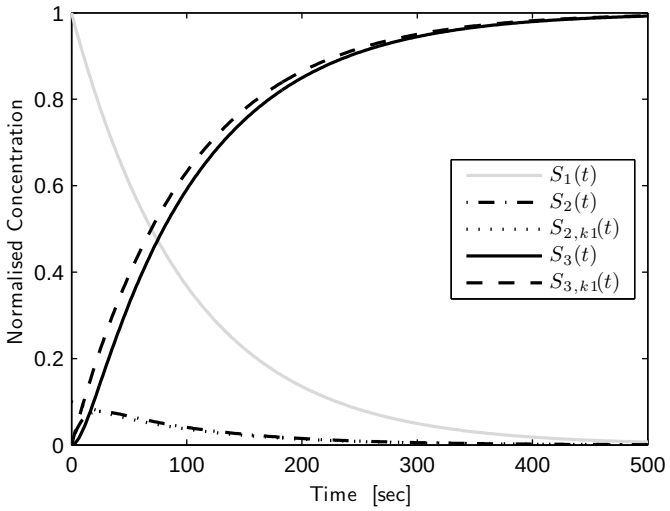


Figure 2.10: Comparison of the full solution and the approximation for a fast second step. The rate coefficients are $k_1 = 0.01\,\mathrm{s}^{-1}$ and $k_2 = 0.1\,\mathrm{s}^{-1}$. The concentration of the intermediate is always small in comparison to other species. It is transformed practically immediately into S_3.

have to wait on the slower reaction. However, the rate-determining step is not just the slowest step: it must be slow and be a crucial gateway for the formation of products. If a faster step also lead to products, the slowest step is irrelevant because the slow reaction can be sidestepped. To finish our discussion we have to give a criteria which step in a consecutive reaction is the slowest.

For this purpose we discuss a more sophisticated model for an monomolecular reaction. It was introduced by Lindemann and experimentally verified by Hinshelwood [AdP02]. Starting point for the model is the idea that a single molecule is excited by a collision. The excited molecule can loose its energy by a new collision or decays in a monomolecular step into the product.

$$S_1 + M \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} S_1^* + M$$
$$S_1^* \xrightarrow{k_1} S_2$$

The molecule M can be from species S_1, S_2 or an inert-gas. Here, an inert-gas is a gas of other species, which do not react with the considered molecules S_1,S_2, and S_1^*. Collisions between two excited molecules S_1^* are negligible in comparison to the others. But a collision is a clear bimolecular process, how we can get a first-order kinetics from this mechanism? The rates for the components follow the system of coupled differential

equations

$$\frac{dS_1(t)}{dt} = -k_2\, S_1(t)\, M(t) + k_{-2}\, S_1^*(t)\, M(t)\,, \tag{2.111}$$

$$\frac{dS_1^*(t)}{dt} = k_2\, S_1(t)\, M(t) - k_{-2}\, S_1^*(t)\, M(t) - k_1\, S_1^*(t) \tag{2.112}$$

$$\frac{dS_2(t)}{dt} = k_1\, S_1^*(t)\ . \tag{2.113}$$

A simplification arises, if we use a steady state approximation for the excited molecule S_1^*. From (2.112), it follows

$$S_1^*(t) = \frac{k_2\, S_1(t)\, M(t)}{k_{-2}\, M(t) + k_1}\ , \tag{2.114}$$

which we can insert into (2.111) or (2.113). The result for the reaction rate is

$$-\frac{dS_1(t)}{dt} = \frac{dS_2(t)}{dt} = \frac{k_1\, k_2\, M(t)\, S_1(t)}{k_{-2}\, M(t) + k_1} \tag{2.115}$$

for which no reaction order can be defined. For a closed system is $M(t) \approx$ const. and the rate equation transforms to a pseudo-first order law

$$-\frac{dS_1(t)}{dt} = \frac{dS_2(t)}{dt} = k_{\text{eff}}\, S_1(t)\ , \tag{2.116}$$

where the effective rate coefficient is

$$k_{\text{eff}} = \frac{k_1\, k_2\, M}{k_{-2}\, M + k_1}\,. \tag{2.117}$$

The effective constant depends on the collision partners M.

Now, we go back to our original question and determine the rate-determining step. For that purpose we consider different amounts of M. First we assume $k_{-2}\, M \ll k_1$. The rate law simplifies to

$$-\frac{dS_1(t)}{dt} = \frac{dS_2(t)}{dt} = k_2\, M\, S_1(t)\,, \tag{2.118}$$

which obeys second-order kinetics. The rate-determining step is the bimolecular collision of S_1 and M. On the other hand, if $k_{-2}\, M \gg k_1$ we obtain a rate of first order

$$-\frac{dS_1(t)}{dt} = \frac{dS_2(t)}{dt} = \frac{k_1\, k_2}{k_{-2}}\, S_1(t)\ . \tag{2.119}$$

The rate-determining step is now the monomolecular decay of excited molecules into products.

From this example we are able to define the rate-determining step. The rate-determining step is the step with the smallest probability to occur within the time interval dt or the with the smallest rate, respectively. The rate coefficient alone is not a proper criteria.

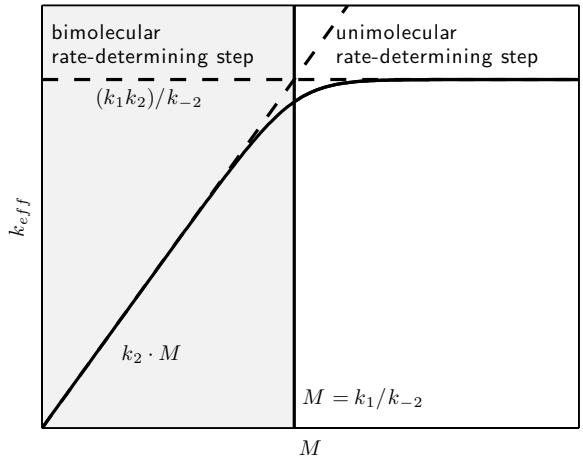


Figure 2.11: Log-log plot for the effective rate constant k_{eff} as function of the collision partner M. For $k_2\,M \ll k_1$, see (2.117), the reaction follows second-order kinetics determined by the bimolecular excitation of S_1 and for $k_2\,M \gg k_1$ first-order kinetics with the monomolecular decay into the product S_2 as product. We use the intersection of the asymptotes of k_{eff} to separate the M-k_{eff}-plain into a bimolecular and monomolecular limited region.

Only in special cases it is the process with the smallest rate coefficient. Remember, the rate for an elementary reaction is the product of the rate coefficient and the participating species. As we demonstrated the rate-determining step depends on the current conditions in the reaction volume. For complex reactions often one cannot define an unique rate-determining step. Instead one has to consider different regimes as we have done it for the sophisticated model of monomolecular reactions. Furthermore, the rate-determining step can be time dependent. Complex reactions with (auto)catalytic reactions are a typical example for such a behavior.

2.6.3.2 The (quasi-)steady state approximation

The second case of a preliminary rate-determining step can be expanded into a more rigorous approximation. Let S_2 an unstable intermediate with a short lifetime and small concentration in comparison to the other participants on the sequence. Its consumption more or less simultaneously with its production. In this case we can assume for its rate

$$\frac{dS_2(t)}{dt} \approx 0 \ . \tag{2.120}$$

It follows that the balance equation is given as

$$\frac{dS_2(t)}{dt} = k_1\,S_1(t) - k_2\,S_2(t) \approx 0 \ , \tag{2.121}$$

leading to the relation

$$k_1\,S_1(t) = k_2\,S_2(t) \ . \tag{2.122}$$

From this we obtain

$$S_2(t) = \frac{k_1}{k_2} S_1(t) \tag{2.123}$$

and finally from (2.100) the simplified time law

$$\frac{dS_3(t)}{dt} = k_1 S_1(t) \tag{2.124}$$

for the final product. An integration leads again to 2.110.

At the end of this section we generalise the steady state approximation. Within this approximation we assume, that during the major part of the reaction the rate of change of intermediates are negligible small[10]. Thereby we neglect an initial time period, where the intermediates rise from zero. In our previous examples, see Figure 2.9 and 2.10, the first 25 seconds cannot be described with the approximation of the rate-determining step. For the next minutes it is a good description. But a ratio of 10(0.1) is not a great difference between the rate coefficient. We chose this ratio for demonstration purposes only.

The amount of the intermediates does not need to be negligible in comparison to the reactants and products as we assume in our example. This more restrictive approximation is often called a *quasi-steady state*.

quasi-steady state

The steady state approximation is a very powerful tool in the analytic treatment of complex reactions. Because of the increasing mathematical complexity reaction schemes involving many steps is nearly always analytically unsolvable. One approach is a numerical solution of the differential equation. An alternative approach is to make an approximation. On the other hand, an approximation restricts the range of validity of the model. For instance, the steady state approximation in (2.112) assumes that the formation of the excited molecule and its decay back into the deactivated form are much faster than the formation of the product. This is only possible if $k_{-2} \gg k_1$, but not when $k_{-2} \ll k_1$.

Hence, the approximations of steady state and rate-determining step have to be used carefully. Their validity is limited and has to recontrol for each specific reaction system and its parameters. An usage far away from the validity region leads to wrong results.

2.6.4 Autocatalytic reactions

catalyst

A particular class of reactions are catalytic reactions. A *catalyst* accelerates the reaction and is released unmodified as product. It occurs as reactant and as product. Note, a catalyst can affect only reactions happening from alone. In other words, the reactions have to be possible from thermodynamic reasons. It also does not change the equilibrium properties. In this section we want to discuss a special kind of catalytic reactions, the autocatalytic reaction. In these reactions a product accelerates its own production. In contrast to the other catalytic reactions we have not to add an additional substance. Known examples for such reactions are chain reactions[11]. Thereby, one distinguishes

[10] Here negligible small means small in comparison to the rates of the other participating species.

[11] In a chain reaction substances are involved recycling the reactants and starting a new reaction cycle.

between two types of (auto)catalytic reactions[12]. The catalyst can increase the reaction rate, unfortunately, this behavior is called (auto)catalysis, too. The inverse effect of a decrease of reaction rate is called (auto)inhibition. The corresponding substance is an (auto)inhibitor.

For an introduction to autocatalysis we choose the simplest possible model, an unidirectional monomolecular reaction. The stoichiometric formula is

autocatalysis

$$S_1 \xrightarrow[S_2]{k} S_2\,, \tag{2.125}$$

where the subscript S_2 denotes, that the product S_2 acts as a autocatalyst. In a more detailed representation one often uses the formula

$$S_1 + S_2 \xrightarrow{k} 2S_2 \tag{2.126}$$

for an autocatalytic reaction. According to the second chemical formula the kinetic equation is

$$\begin{aligned} r(t) &= -\frac{dS_1(t)}{dt} = \frac{dS_2(t)}{dt} = \frac{d\epsilon(t)}{dt} \\ &= k\,S_1(t)\,S_2(t) \\ &= k\,\big[S_1(0) - \epsilon(t)\big]\big[S_2(0) + \epsilon(t)\big]\ . \end{aligned} \tag{2.127}$$

The autocatalytic effect of the product S_2 formally increase the reaction order. Be aware, this is only an 'Ansatz', not a full description of the reaction mechanism. The mechanism of catalytic reaction is often complicated, hence we forbear to specify it. Instead, we discuss some basic properties in our model.

The autocatalytic reaction fulfills the conservation law

$$S_1(0) = S_1(t) + S_2(t) - S_2(0)\ , \tag{2.128}$$

where we assume an initial concentration of the autocatalyst. Kinetic equations (2.127) are solvable by separation of variables. The integral

$$kt = \int_0^{\epsilon(t)} \frac{d\epsilon(t)}{\big[S_1(0) - \epsilon(t)\big]\big[S_2(0) + \epsilon(t)\big]} \tag{2.129}$$

is solved by an expansion into partial fractions. We obtain

$$kt = \frac{1}{S_1(0) + S_2(0)}\,\ln\frac{S_1(0)}{S_2(0)}\,\frac{S_2(t)}{S_1(t)} \tag{2.130}$$

from which we get

$$S_1(t) = \frac{S_1(0) + S_2(0)}{S_2(0)/S_1(0)\,\exp\Big\{\big[S_1(0) + S_2(0)\big]\,kt\Big\} + 1} \tag{2.131}$$

[12] An acceleration is a change of velocity, not necessarily an increase.

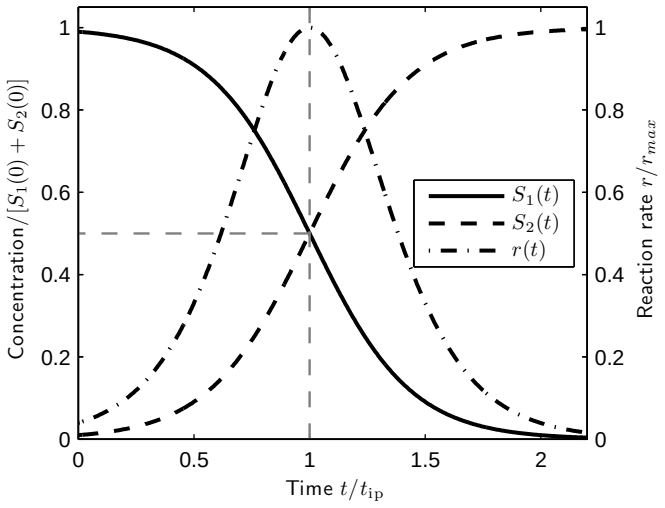


Figure 2.12: The time evolution and the reaction as function of time for the autocatalytic reaction (2.125). We use normalised units for the concentrations and the time dependent rate. The time is plotted in units of the time of the inflection point (2.133). In contrast to uncatalyzed monomolecular reaction the reaction rate is an increasing function for $t < t_{\mathrm{ip}}$.

and

$$S_2(t) = \frac{S_1(0) + S_2(0)}{S_1(0)/S_2(0)\exp\Big\{-\big[S_1(0)+S_2(0)\big]\,kt\Big\} + 1} \tag{2.132}$$

for the reactant and the product as function of time. Both functions have an inflection point typically for autocatalytic reactions at

$$t_{\mathrm{ip}} = \frac{1}{\big[S_1(0)+S_2(0)\big]\;k}\,\ln\frac{S_1(0)}{S_2(0)}\;. \tag{2.133}$$

The associated concentrations are

$$S_{1,\mathrm{ip}} = \frac{S_1(0)+S_2(0)}{2} = S_{2,\mathrm{ip}}\;. \tag{2.134}$$

Also, the reaction rate shows some interesting and typical properties. At the beginning the rate is small, because of the small amount of the catalyst. With increasing concentration of S_2 the rate increases and reaches a maximum

$$r_{\max} = \frac{k}{4}(S_1(0)+S_2(0))^2 \tag{2.135}$$

at the same time point (2.133), where the concentrations have their inflection point. After this point the reaction gets slower because of the decreasing amount of the reactant S_1.

In Figure 2.12 the time-dependent normalised concentrations $S_1(t)$, $S_2(t)$ and the normalised reaction rate are plotted. The time is in units of the inflection point. As

mentioned before the concentrations have an inflection point where the reaction rate reaches its maximum.

The reaction starts with a small but finite initial concentration of the autocatalyst S_2. Without this condition the ansatz for the reaction rate does not work. We avoid this non-physical behavior by invoking an extra uncatalyzed reaction converting S_1 directly to S_2. How we mentioned before such a reaction must exist, but it can be very slow. The rate equation is now

$$r = -\frac{dS_1(t)}{dt} = \frac{dS_2(t)}{dt} = \left[k_0 + k\, S_2(t)\right] S_1(t)\,, \tag{2.136}$$

where k_0 is the rate coefficient of the uncatalyzed monomolecular reaction. After an initial time the second term within the brackets is dominant resulting in (2.127).

2.7 The enzyme kinetic reaction

The vast majority of reactions in the cell is facilitated by a catalyst. These type of reactions have been studied for a long time in the context of enzyme kinetics and have found many applications in modelling metabolic networks. The idea of a reaction that is facilitated by another molecular species will also prove useful when we come to signal transduction pathways, where the enzyme is called a kinase. The present section derives the well known Michaelis-Menten representation in detail and considers several distinct kinds of enzymatic reaction types, including competitive binding, inhibitors, reversible Michaelis-Menten equation. We discuss the consequences of these different formulations on the dynamics the models can capture.

2.7.1 The Michaelis-Menten equation

enzyme kinetics

The starting point of our discussion of enzyme kinetics is the well-known Michaelis-Menten equation [BH25]. Within this framework we assume that the conversion of the substrate S into the product P is catalyzed by the enzyme E. The enzyme and the substrate form an intermediate enzyme-substrate complex C which can degrade into the reactants or into product and enzyme. The kinetic mechanism can be presented in the usual graphical form

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_2} E + P \tag{2.137}$$

where the k_i denote the rate coefficients for each elementary reaction. The kinetic rate equations for that are now a system of four coupled differential equations

$$\begin{aligned}
\dot{S} &= -k_1\, E(t)\, S(t) + k_{-1}\, C(t)\,, && (2.138)\\
\dot{E} &= -k_1\, E(t)\, S(t) + [k_{-1} + k_2] C(t)\,, && (2.139)\\
\dot{C} &= k_1\, E(t)\, S(t) - [k_{-1} + k_2] C(t) = -\dot{E}\,, && (2.140)\\
\dot{P} &= k_2\, C(t)\,. && (2.141)
\end{aligned}$$

There is a conservation law

$$E(0) = E(t) + C(t) \tag{2.142}$$

for the enzyme, where $E(0)$ is the initial concentration. A similar law can be found for the substrate

$$S(0) = S(t) + C(t) + P(t)\,. \tag{2.143}$$

Considering these conservation laws the system of coupled differential equations (2.138)-(2.141) reduces from a fourth to a second-order system. Usually, the equations for the substrate (2.138) and for the enzyme-substrate complex C (2.140) are kept. In the further treatment, one obtains an expression for the rate of change of the substrate, namely is the Michaelis-Menten equation. This result is identified as the reaction rate. Formally, this is not correct, because of for a complex reaction scheme the reaction rate is the rate of change for the product only. As we show for consecutive reactions, the rate of change of the intermediate steps is different. Nevertheless, within the approach of Michaelis and Menten, the rates of substrate and product are equal. Therefore, the reaction rate can be derived from the substrate change.

To prove this we use only the conservation law (2.142) for the enzyme and obtain the system of differential equations (2.138), (2.140) and (2.141). This system is not solvable in terms of simple analytic functions [SM97], therefore we follow Michaelis and Menten [MM13] and Briggs and Haldane [BH25] and introduce a steady state assumption[13] for the complex

$$\dot{C} \approx 0 = k_1\, E(t)\, S(t) - [k_{-1} + k_2] C(t)\,. \tag{2.144}$$

From this equation we obtain a constant concentration for the enzyme-substrate complex

$$C(t) = \frac{k_1}{k_{-1} + k_2}\, E(t)\, S(t) = \frac{E(t)\, S(t)}{K_\mathrm{M}} \tag{2.145}$$

or for the enzyme

$$E(t)\, S(t) = K_\mathrm{M}\, C(t)\,. \tag{2.146}$$

Insertion into equation (2.138) leads to

$$\dot{S} = -\dot{P} = -k_2\, C(t) = -k_2\, \frac{E(t)\, S(t)}{K_\mathrm{M}}\ . \tag{2.147}$$

Within the steady state approximation the mechanism (2.137) formally reduces to a bimolecular reaction of the substrate and the enzyme. Furthermore, we introduce the Michaelis constant

$$K_\mathrm{M} = \frac{k_{-1} + k_2}{k_1}\,. \tag{2.148}$$

Because the rate coefficients k_i depend on temperature, pH-value, etc. the Michaelis constant and the limiting rate are not constant quantities. Their values change with the environmental conditions of the considered system.

[13] In their original work Michaelis and Menten consider only the reverse reaction into the complex and its decay into substrate and enzyme and neglect the conversion into the product. Briggs and Haldane generalised the steady state assumption into the given expression. The resulting relation for the reaction rate has in both cases the same form.

The reduction of the apparent reaction mechanism and the resulting mathematical simplification is typical for the steady state assumption. It makes this approximation a powerful tool in the analytical treatment of the complex biochemical reactions in metabolism and inter- and intracellular communications.

Combining the conservation law (2.142) and the steady state complex concentration (2.145), one gets

$$C(t) = \frac{E(0)\,S(t)}{K_\mathrm{M} + S(t)} \tag{2.149}$$

for the enzyme. This result we insert into (2.147) and obtain

$$\begin{aligned} V = -\dot{S} = \dot{P} &= \frac{k_2\,E(0)\,S(t)}{K_\mathrm{M} + S(t)} \\ &= \frac{V_\mathrm{max}\,S(t)}{K_\mathrm{M} + S(t)} \end{aligned} \tag{2.150}$$

the well-known Michaelis-Menten equation, where $V_\mathrm{max} = k_2\,E(0)$ is the limiting rate for the enzyme kinetic reaction.

For $K_\mathrm{M} \gg S(t)$ equation (2.150) provides a linear dependence of the reaction rate

$$-\dot{S} = \dot{P} \approx \frac{k_2\,E(0)}{K_\mathrm{M}}\,S(t) \tag{2.151}$$

from the substrate concentration. The rate-determining step for low substrate concentrations is the bimolecular formation of the enzyme-substrate complex. In this approximation the limiting rate is reached if the substrate concentration is equal to the Michaelis constant. For $S(t) \gg K_\mathrm{M}$ the Michaelis-Menten equation (2.150) reduces to

$$-\dot{S} = \dot{P} = V_\mathrm{max} \tag{2.152}$$

a constant rate. The pre-equilibria of enzyme, substrate and enzyme-substrate complex rests completely with the side of the complex. All enzyme molecules are bound to a substrate molecule, and according to the conservation law (2.142), the complex concentration is equal to the initial enzyme concentration. The reaction rate is determined by the decay of the complex into product and enzyme. This process is described by the rate coefficient k_2 and independent from the substrate concentration. Hence, the reaction rate cannot be increased further by an increase of the substrate concentration. According to IUPAC[14] it is called 'limiting rate' and not maximum rate as in older books, and the usual notation V_max is retained for this quantity.

limiting rate

The term introduced limiting regimes which allow an independent measurement of the Michaelis constant and the limiting rate. For a known enzyme concentration we can calculate the rate coefficient k_2. The coefficients k_1 and k_{-1} are not uniquely determined. Only if one assumes $k_{-1} \gg k_2$ the Michaelis constant reduces to the ratio

$$K_\mathrm{M} \approx \frac{k_{-1}}{k_1}\,, \tag{2.153}$$

[14] IUPAC - International Union of Pure and Applied Chemistry

as did the constant originally introduced by Michaelis and Menten.

While the present section introduced the basic treatment of enzyme kinetic reactions we will use the following sections to consider more complex conversion schemes.

2.7.1.1 Dimensionless representation – Activity plot

The dimensionless representation of the Michaelis-Menten equation is a very useful tool for comparison of results for more complex enzyme kinetic reactions. It gives a very clear insight into the properties of the catalytic conversion of a substrate into the needed product. It allows the comparison of more complex enzyme kinetic reactions, we will introduce in the next chapters, with the Michaelis-Menten model. Furthermore its general form allows an easy discussion of general properties of the enzymatic conversion of a substrate into a product.

Therefore, we rearrange the Michaelis-Menten equation (2.150) into

$$A = \frac{V}{V_{\max}} = \frac{S(t)/K_{\mathrm{M}}}{1 + S(t)/K_{\mathrm{M}}}\,, \tag{2.154}$$

activity

where the new defined ratio of the current reaction rate V and the limiting rate $V_{\max}$ is called *activity*. The substrate concentration is now given in units of the Michaelis constant K_{M}. The important activity $A_{0.5}$, where the half limiting rate is reached, is related to this constant. Because of the chosen dimensionless concentration it is now obtained for a value $S/K_{\mathrm{M}} = 1$. The activity has its limiting value of $A = 1$, if the reaction rate approaches to its limiting rate $V_{\max}$.

2.7.2 The enzyme kinetic reaction with second intermediate

Another possible reaction scheme for the catalytic conversion of a substrate is a mechanism with an additional intermediate complex C_2. The stoichiometric formula for such a reaction is

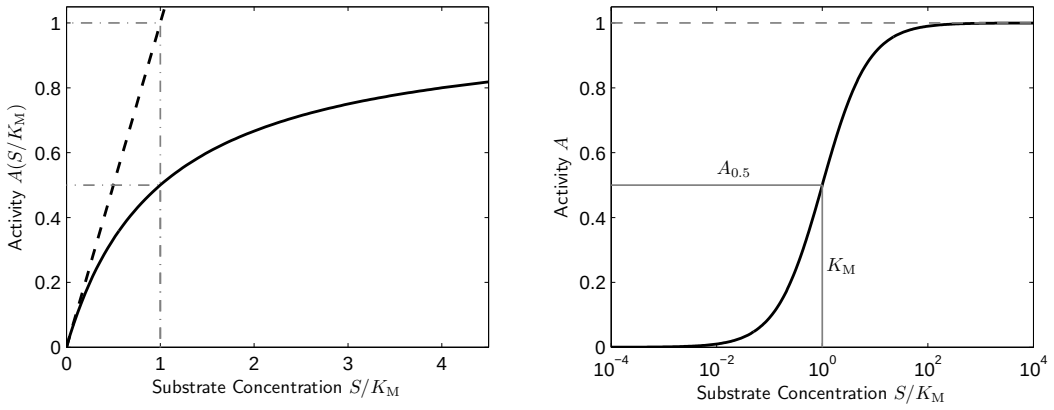
$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C_1 \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} C_2 \xrightarrow{k_3} E + P \tag{2.155}$$

where the transition between the intermediates is fully reversible. The corresponding system of coupled differential equations is

$$\begin{aligned}
\dot{S} &= -k_1\, E(t)\, S(t) + k_{-1}\, C_1(t)\,, && (2.156)\\
\dot{E} &= -k_1\, E(t)\, S(t) + k_{-1}\, C_1(t) + k_3\, C_2(t)\,, && (2.157)\\
\dot{C}_1 &= k_1\, E(t)\, S(t) - [k_{-1} + k_2]\, C_1(t) + k_{-2}\, C_2(t)\,, && (2.158)\\
\dot{C}_2 &= k_2\, C_1(t) - [k_{-2} + k_3]\, C_2(t)\,, && (2.159)\\
\dot{P} &= k_3\, C_2(t)\,. && (2.160)
\end{aligned}$$

For this reaction scheme, all equations are linear independent, in contrast to the system of coupled differential equations (2.138)-(2.141) for the Michaelis-Menten model. For the enzyme we find now the conservation law

$$E(0) = E(t) + C_1(t) + C_2(t)\,. \tag{2.161}$$



(a) (b)

Figure 2.13: Activity (solid line) as function of the substrate concentration in a linear (left) and a logarithmic (right) representation. In the linear representation, the small-concentration-approximation is drawn as a dashed line. Additionally, the activity value for the concentration $S/K_M = 1$ is indicated. In the logarithmic representation, the high-concentration-limit is drawn with a dashed line. The thin solid lines corresponds to the Michaelis constant and the half activity, respectively. The sigmoidal shape of the activity curve is a result of the logarithmic representation. It shows the change of the rate-determining step with increasing substrate concentration. For low concentrations the bimolecular formation of the enzyme-substrate complex, and for high concentrations the decomposition of the complex into product P and free enzyme E determines the reaction rate.

Again we assume the enzyme-substrate complexes C_1 and C_2 at steady states. Considering the conservation law (2.161) we obtain for the first complex C_1

$$C_1(t) = \frac{k_1\, E(0)\, S(t) - [k_1\, S - k_{-2}]\, C_2(t)}{k_{-1} + k_2 + k_1 S(t)}\,. \tag{2.162}$$

This expression we insert into the equation for the second complex C_2 resulting in the steady state concentration

$$C_2(t) = \frac{k_1 k_2\, E(0)\, S(t)}{k_{-1}k_{-2} + k_{-1}k_3 + k_2 k_3 + [k_1 k_2 + k_1 k_{-2} + k_1 k_3]\, S(t)}\,. \tag{2.163}$$

which we insert into the differential equation for the product (2.160). Finally, we obtain for the reaction rate of this type of enzyme kinetic reaction the expression

$$V(t) = \dot{P} = \frac{k_1 k_2 k_3\, E(0)\, S(t)}{k_{-1}k_{-2} + k_{-1}k_3 + k_2 k_3 + [k_1 k_2 + k_1 k_{-2} + k_1 k_3]\, S(t)}\,. \tag{2.164}$$

Introducing an apparent Michaelis constant

$$K^{\text{app}} = \frac{k_{-1}k_{-2} + k_{-1}k_3 + k_2 k_3}{k_1 k_2 + k_1 k_{-2} + k_1 k_3} \tag{2.165}$$

and an apparent limiting rate

$$V^{\text{app}} = \frac{k_2 k_3}{k_2 + k_{-2} + k_3}\, E(0) \tag{2.166}$$

the expression (2.164) simplifies to

$$V(t) = \frac{V^{\text{app}}\, S(t)}{K^{\text{app}} + S(t)}\,, \tag{2.167}$$

a Michaelis-Menten-like form. Whereas the Michaelis constant (2.148) is interpretable as a dissociation constant, the apparent Michaelis constant prohibits such a physical interpretation.

Analogous to the simpler Michaelis-Menten scheme the mechanism (2.155) formally reduces to a bimolecular reverse reaction. To prove this, we transform Eq. (2.159) at steady state into an expression for

$$k_2\, C_1(t) = [k_{-2} + k_3]\, C_2(t) \tag{2.168}$$

and insert it into (2.158). The result

$$k_{-1}\, C_1 = k_1\, S(t) E(t) - k_3 C_2(t) \tag{2.169}$$

we use to simplify the rate of change of the substrate

$$\dot{S} = -\dot{P} = -k_3\, C_2\,. \tag{2.170}$$

The reaction rate is

$$\begin{aligned} V(t) &= -\dot{S} = \dot{P} = k_3\, C_2(t) \\ &= \frac{k_1 k_2 k_3\, E(0)\, S(t)}{k_1 k_2 + k_2 k_{-2} + k_2 k_3 + (k_2 k_{-2} + k_1 k_3) S(t)}\,. \end{aligned} \tag{2.171}$$

2.7.3 Constant substrate concentration

If we assume a constant substrate concentration, for instance because of external flows or an external reservoir, we can further simplify the Michaelis-Menten equation (2.150). In this special case the right hand side is a constant. The reaction rate

$$V = -\dot{S} = \dot{P} = k_{\text{eff}} \tag{2.172}$$

follows zero-order kinetics. The introduced effective rate coefficient is defined as

$$k_{\text{eff}} = \frac{k_2\, E(0)\, S}{K_{\text{M}} + S}\,. \tag{2.173}$$

For high substrate concentrations $S \gg K_{\text{M}}$ the rate (2.172) reduces to

$$V \approx k_2\, E(0)\,,$$

the decay of the enzyme-substrate complex into product and enzyme is the rate-determining step. On the other hand, for small substrate concentrations $S \ll K_{\text{M}}$ the rate

$$V \approx k_1\, E(0)\, S(t)$$

is limited by the bimolecular formation of the enzyme-substrate complex C_1.

2.7.4 Interactions with other reaction mechanisms and reaction partners

Above we discussed the conversion of a substrate into a product by a catalyzing enzyme for a single reaction channel and without the presence of further substances. The next sections focus on the interaction of different reaction channels and the influences of other substances to a specified conversion. We show some simple examples of inhibitory mechanisms to the enzyme kinetic reaction (2.137). The term inhibition is misleading here, because after the definition in chemistry a catalyst[15] is an additional substance changing the reaction rate in a positive or negative way. But for some types of inhibition we have no additional substrates or enzymes.

2.7.4.1 Competitive substrates

A standard example for an inhibitory mechanism is the case of two competitive substrates. The two different substrates are catalyzed by the same enzyme E. In the stoichiometric representation

inhibition

$$\begin{aligned} S + E &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C_1 \xrightarrow{k_2} E + P_1\,, \\ I + E &\underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} C_2 \xrightarrow{k_4} E + P_2\,, \end{aligned} \tag{2.174}$$

[15] Every additional substance influencing the reaction rate is a catalyst. If it increases the reaction rate (positive catalysis) we call it catalyst. An inhibitor decreases the rate (negative catalysis).

we assume that both processes follow the model of Michaelis and Menten. The first reaction in (2.174) is our desired conversion reaction from S to the product P_1. It is suppressed by the reaction of an inhibitor I with the enzyme, where we allow the production of an unused product P_2. This system of chemical reactions is described by a system of seven coupled differential equations. The system

$$\begin{aligned}
\dot{S} &= -k_1\, S(t)\, E(t) + k_{-1}\, C_1(t) && (2.175)\\
\dot{I} &= -k_3\, I(t)\, E(t) + k_{-3}\, C_2(t) && (2.176)\\
\dot{E} &= -\left[k_1\, S(t) + k_3\, I(t)\right] E(t) \\
&\quad + (k_{-1} + k_2)\, C_1(t) + (k_{-3} + k_4)\, C_2(t) && (2.177)\\
\dot{C}_1 &= k_1\, S(t)\, E(t) - (k_{-1} + k_2)\, C_1(t) && (2.178)\\
\dot{C}_2 &= k_3\, I(t)\, E(t) - (k_{-3} + k_4)\, C_2(t) && (2.179)\\
\dot{P}_1 &= k_2\, C_1(t) && (2.180)\\
\dot{P}_2 &= k_4\, C_2(t) && (2.181)
\end{aligned}$$

can be reduced in the same manner as for the single enzyme kinetic reaction. The conservation law for the enzyme

$$E(0) = E(t) + C_1(t) + C_2(t) \tag{2.182}$$

and the steady state assumption for both complexes cancel three differential equations. Furthermore we can show again, that the rate of degradation of substrate is equal to the production rate of product P_1 in this approximation. The same can be shown for the inhibitor reaction. Only the differential equations for the substrate and the inhibitor remain. It follows from (2.178) that

$$C_1(t) = \frac{E(0)\, S(t)}{K_{\mathrm{MS}}\left[1 + I(t)/K_{\mathrm{MI}}\right] + S(t)} \tag{2.183}$$

for the enzyme-substrate complex and

$$C_2(t) = \frac{E(0)\, I(t)}{K_{\mathrm{MI}}\left[1 + S(t)/K_{\mathrm{MS}}\right] + I(t)} \tag{2.184}$$

for the inhibitor-enzyme complex. The number of bound substrate molecules is reduced in comparison to the single enzyme kinetic reaction, see equation (2.149). Because this concentration determines the conversion rate into the product, a reduction of this concentration leads to a decrease of the reaction rate. The enzyme kinetic reaction (2.137) is suppressed by the presence of a second molecule reacting with the same enzyme.

For the reaction rate we obtain with (2.183)

$$V^{\mathrm{S}}(t) = -\frac{V^{\mathrm{S}}_{\max}\, S(t)}{K_{\mathrm{MS}}\left[1 + I(t)/K_{\mathrm{MI}}\right] + S(t)} \tag{2.185}$$

and for the inhibitor

$$V^{\mathrm{I}}(t) = -\frac{V^{\mathrm{I}}_{\max}\, I(t)}{K_{\mathrm{MI}}\left[1 + S(t)/K_{\mathrm{MS}}\right] + I(t)}\,, \tag{2.186}$$

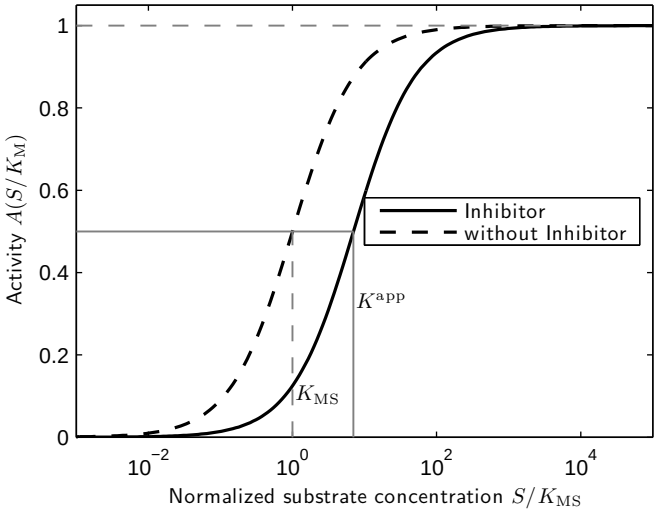


Figure 2.14: The reaction rate of the enzyme kinetic reaction in presence of an inhibitor I reacting with the same enzyme to a second unused product as function of the substrate concentration in logarithmic representation. The corresponding concentration of the Michaelis constant K_{MS} and the apparent Michaelis constant K^{app} is shown by thin solid lines.

where $V^{\mathrm{S}}_{\mathrm{max}} = k_2\, E(0)$ is the limiting rate for the substrate conversion and $V^{\mathrm{I}}_{\mathrm{max}}$ the corresponding quantity for the inhibitory reaction. In the limit of $S \gg K_{\mathrm{MS}}$ the reaction rate reaches its limiting value and the inhibitor is negligible. In the contrary case, the reaction rate is again proportional to the substrate concentration but it is reduced by a factor $1 + I/K_{\mathrm{MI}}$ in comparison to the result from our previous discussion. Finally we investigate the rate in dependence from substrate concentration and a given inhibitor concentration. Then we can treat the inhibitor amount as constant and combine

$$K^{\mathrm{app}} = K_{\mathrm{MS}}\,[1 + I/K_{\mathrm{MI}}]$$

into the apparent Michaelis constant K^{app}, which is greater than the Michaelis constant K_{MS} of the single enzyme kinetic reaction. The result is a Michaelis-Menten like differential equation

$$V^{\mathrm{S}}(t) = -\frac{V^{\mathrm{S}}_{\mathrm{max}}\, S(t)}{K^{\mathrm{app}} + S(t)}\,. \tag{2.187}$$

In Figure 2.14 the Michaelis-Menten equation (2.150) is compared with result in presence of an inhibitor I. Whereas for small substrate concentrations the inhibitor strongly affects the reactions rate, for high substrate concentrations the inhibitor is negligible, as mentioned before, and the reaction rate approaches to its limiting rate determined by the decay rate of the enzyme-substrate complex.

2.7.4.2 Non-productive binding

Another mechanism of suppression is the existence of other reaction channels. In a first example we want to discuss an enzyme kinetic reaction where the enzyme and the substrate can bind to a second enzyme-substrate complex that cannot convert into product and free enzyme. Because this second reaction channel does not form any product this scheme is called non-productive binding[16]. The corresponding system of stoichiometric equations is given as

$$\begin{aligned} S + E &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C_1 \xrightarrow{k_2} E + P_1 \,, \\ S + E &\underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} C_2 \,, \end{aligned} \tag{2.188}$$

where the second reaction is a reversible reaction between substrate and enzyme and the enzyme-substrate complex. The rates of change are now given by the system of differential equations

$$\dot{S} = -(k_1 + k_3)\, E(t)\, S(t) + k_{-1}\, C_1(t) + k_{-3}\, C_2(t) \,, \tag{2.189}$$

$$\dot{E} = -[k_1 + k_3]\, E(t)\, S(t) + [k_{-1} + k_2]\, C_1(t) + k_{-3}\, C_2(t) \,, \tag{2.190}$$

$$\dot{C}_1 = k_1\, E(t)\, S(t) - [k_{-1} + k_2]\, C_1(t) \,, \tag{2.191}$$

$$\dot{C}_2 = k_3\, E(t)\, S(t) - k_{-3}\, C_2(t) \,, \tag{2.192}$$

$$\dot{P} = k_2\, C_1(t) \tag{2.193}$$

with the conservation law

$$E(0) = E(t) + C_1(t) + C_2(t) \,. \tag{2.194}$$

In a steady state approximation for the complexes we obtain

$$C_1(t) = \frac{E(t)\, S(t)}{K_{\mathrm{M1}}} \tag{2.195}$$

$$C_2(t) = \frac{E(t)\, S(t)}{K_{\mathrm{D}}} \tag{2.196}$$

for the complex concentration, where

$$K_{\mathrm{M1}} = \frac{k_{-1} + k_2}{k_1} \tag{2.197}$$

is the usual Michaelis constant and

$$K_{\mathrm{D}} = \frac{k_{-3}}{k_3} \tag{2.198}$$

the dissociation constant of the non-productive complex. An analysis of the remaining differential equations of substrate and product concentration shows again, that they are

[16] The enzyme-substrate complex is an unstable intermediate (transition state), it cannot exist alone.

the same except for the sign. Taking into account the conservation law (2.194) we obtain for the enzyme concentration

$$E(t) = \frac{E(0)}{1 + S(t)/K_{\mathrm{M1}} + S(t)/K_{\mathrm{D}}}\,, \tag{2.199}$$

which leads to a reaction rate

$$\begin{aligned} V(t) &= \dot{P} = -\dot{S} \\ &= \frac{V_{\max}\, S(t)}{K_{\mathrm{M1}}\,[1 + S(t)/K_{\mathrm{M1}} + S(t)/K_{\mathrm{D}}]}\,, \end{aligned} \tag{2.200}$$

where $V_{\max} = k_2\, E(0)$ is the limiting rate known from previous sections. From (2.200) one can see, that the reaction rate is decreased by the second non-productive parallel reaction. The obtained expression for the rate we can transform into a Michaelis-Menten like relation if we introduce effective constants. For the non-productive binding scheme (2.188) we have the apparent limiting rate

$$V_{\max}^{\mathrm{app}} = \frac{V_{\max}}{1 + K_{\mathrm{M1}}/K_{\mathrm{D}}} \tag{2.201}$$

and an apparent Michaelis constant

$$K^{\mathrm{app}} = \frac{K_{\mathrm{M1}}}{1 + K_{\mathrm{M1}}/K_{\mathrm{D}}}\,. \tag{2.202}$$

Both coefficients are smaller than their pendants in the Michaelis-Menten equation (2.150). Again, we get a Michaelis-Menten like equation

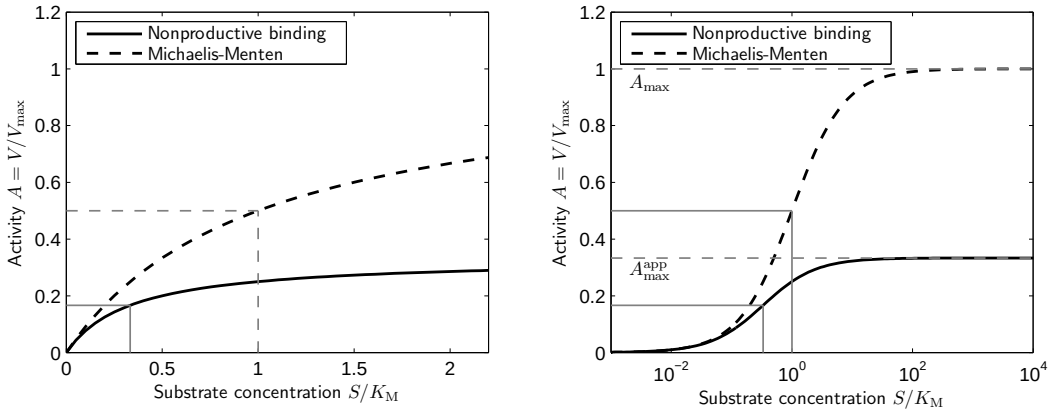
$$V = \frac{V_{\max}^{\mathrm{app}}\, S(t)}{K_{\mathrm{M}}^{\mathrm{app}} + S(t)}\,, \tag{2.203}$$

for the reaction rate. In comparison to a conversion according to (2.137) with the same rate coefficients the non-productive reaction channel decelerates the conversion into P, because a fraction of the enzyme is bound in non-productive complex C_2. This is shown in Figure 2.15, where we compare the result of the Michaelis-Menten equation (2.150) and the Michaelis-Menten like equation (2.203). In both cases we use the same rate coefficients for the successful conversion reaction of the substrate.

An enzymatic reaction with a non-productive channel cannot be distinguished from a reaction without such a channel by the measurement of substrate or product concentration, because the apparent constants depends only on rate coefficient. For this purpose one has to identify the complexes.

2.7.4.3 Non-productive binding with an additional product

In this section we expand our previous treatment and introduce the second reaction as a parallel reaction, where the produced product P_2 is an unused by-product. From this



(a) (b)

Figure 2.15: Comparison of non-productive reaction scheme (2.188) with the Michaelis-Menten mechanism (2.137) in linear (left) and logarithmic (right) representation. In the presence of a second complex which cannot lead to a product the reaction rate is reduced. Note, that the apparent Michaelis constant and the apparent limiting rate are always smaller than the corresponding parameter of the original Michaelis-Menten equation (2.150).

we can expect that the second reaction channel decelerates down the reaction rate and decrease production of the desired product P_1. The stoichiometric representation of such a biochemical system is

$$\begin{aligned} S + E &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C_1 \xrightarrow{k_2} E + P_1\,, \\ S + E &\underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} C_2 \xrightarrow{k_4} E + P_2\,, \end{aligned} \tag{2.204}$$

from which follows the system of six coupled differential equations

$$\dot{S} = -(k_1 + k_3)\, E(t)\, S(t) + k_{-1}\, C_1(t) + k_{-3}\, C_2(t)\,, \tag{2.205}$$

$$\dot{E} = -[k_1 + k_3]\, E(t)\, S(t) + [k_{-1} + k_2]\, C_1(t) + [k_{-3} + k_4]\, C_2(t)\,, \tag{2.206}$$

$$\dot{C}_1 = k_1\, E(t)\, S(t) - [k_{-1} + k_2]\, C_1(t)\,, \tag{2.207}$$

$$\dot{C}_2 = k_3\, E(t)\, S(t) - [k_{-3} + k_4]\, C_2(t)\,, \tag{2.208}$$

$$\dot{P}_1 = k_2\, C_1(t)\,, \tag{2.209}$$

$$\dot{P}_2 = k_4\, C_2(t)\,. \tag{2.210}$$

Again, we have the conservation law

$$E(0) = E(t) + C_1(t) + C_2(t) \tag{2.211}$$

which cancels the differential equation for the enzyme. A further simplification arises from the steady state assumption for both complexes. Their steady state concentrations

are

$$C_1(t) = \frac{k_1}{k_{-1} + k_2} E(t)\, S(t) = \frac{E(t)\, S(t)}{K_{\mathrm{M1}}}\,, \tag{2.212}$$

$$C_2(t) = \frac{k_3}{k_{-3} + k_4} E(t)\, S(t) = \frac{E(t)\, S(t)}{K_{\mathrm{M2}}}\,, \tag{2.213}$$

where we introduced the Michaelis constants K_{M1} and K_{M2}. With the conservation law (2.211) and $V_{\mathrm{max1}} = k_2\, E(0)$ the reaction rate for P_1 is

$$V^{P_1}(t) = \dot{P}_1 = \frac{V_{\mathrm{max1}}\, S(t)}{K_{\mathrm{M1}} + S(t)\,(1 + K_{\mathrm{M1}}/K_{\mathrm{M2}})}\,. \tag{2.214}$$

With the apparent Michaelis constant

$$K_1^{\mathrm{app}} = \frac{K_{\mathrm{M1}}}{1 + K_{\mathrm{M1}}/K_{\mathrm{M2}}} \tag{2.215}$$

and the apparent limiting rate is

$$V_{\mathrm{max1}}^{\mathrm{app}} = \frac{V_{\mathrm{max1}}}{1 + K_{\mathrm{M1}}/K_{\mathrm{M2}}} \tag{2.216}$$

we transform (2.214) into the Michaelis-Menten like form

$$V^{P_1}(t) = \frac{V_{\mathrm{max1}}^{\mathrm{app}}\, S(t)}{K_1^{\mathrm{app}} + S(t)}\,. \tag{2.217}$$

In the same manner we obtain for by-product P_2 a similar relation, where the apparent Michaelis constant is given as

$$K_2^{\mathrm{app}} = \frac{K_{\mathrm{M2}}}{1 + K_{\mathrm{M2}}/K_{\mathrm{M1}}} \tag{2.218}$$

and the apparent limiting rate

$$V_{\mathrm{max2}}^{\mathrm{app}} = \frac{V_{\mathrm{max2}}}{1 + K_{\mathrm{M2}}/K_{\mathrm{M1}}}\,, \tag{2.219}$$

where the limiting rate is $V_{\mathrm{max2}} = k_4\, E(0)$. Last, but not least we have to calculate the rate of substrate depletion. From (2.212) follows

$$k_1\, E(t)\, S(t) = (k_{-1} + k_2)\, C_1(t)$$

and from (2.213)

$$k_3\, E(t)\, S(t) = (k_{-3} + k_4)\, C_1(t)\,.$$

If we insert both expressions into the rate of change of substrate S (2.205) the contributions of reverse reaction of complexes C_1 and C_2 are compensated and we obtain the rate

$$\dot{S} = -k_2\, C_1(t) - k_4\, C_2(t) \tag{2.220}$$

depending only on the decay into the products and free enzyme. On the other hand, the rate of production of P_1 and P_2 is determined by these two terms and we can rewrite the last equation into

$$\dot{S} = -\left[\dot{P}_1 + \dot{P}_2\right] . \tag{2.221}$$

The rate of change of the substrate is the sum of the rates of change of the products. In contrast to our previous reaction schemes, the rates are not equal, except for the sign, any more. This is typical for a reaction with more than one product. The consumption of the substrate is separated into two independent channels. Therefore, the concentrations of substrate and product[17] as a function of time should be measured in an experiment. A comparison of the sum of the reaction rates of the products and the rate of change of the substrate gives evidence if there are unknown reaction channels. Furthermore, the differences arising from unknown channels can be used to estimate the portion of these reactions on the consumption of the substrate. The insertion of the reaction rates for the products (2.217) and its analogue for P_2 leads to

$$\dot{S} = -\left(\frac{V_{\mathrm{max1}}^{\mathrm{app}}}{K_1^{\mathrm{app}} + S(t)} + \frac{V_{\mathrm{max2}}^{\mathrm{app}}}{K_2^{\mathrm{app}} + S(t)}\right) S(t) \tag{2.222}$$

a rate that is not presentable in a Michaelis-Menten-like form. Only in the special case $K_{\mathrm{M1}}^{\mathrm{app}} = K_{\mathrm{M2}}^{\mathrm{app}}$ the rate of substrate S can be transformed into such a form.

It is interesting to analyze some asymptotic properties of (2.222). First we consider the limit of low substrate concentrations $S \ll K_{\mathrm{M1}}^{\mathrm{app}}, K_{\mathrm{M2}}^{\mathrm{app}}$. Then the rate of degradation reduces to

$$\dot{S} = -\left(\frac{V_{\mathrm{max1}}}{K_{\mathrm{M1}}} + \frac{V_{\mathrm{max2}}}{K_{\mathrm{M2}}}\right) S(t) , \tag{2.223}$$

which is a linear relation similar to the result for the Michaelis-Menten equation (2.151). In the other limit $S \gg K_{\mathrm{M1}}^{\mathrm{app}}, K_{\mathrm{M2}}^{\mathrm{app}}$ one obtains with

$$\dot{S} = -\left(V_{\mathrm{max1}}^{\mathrm{app}} + V_{\mathrm{max2}}^{\mathrm{app}}\right) = V_{\mathrm{max}}^{\mathrm{T}} \tag{2.224}$$

a limiting rate, too. The formal differences appear only in an intermediate range of concentrations. In figure 2.16 we use these asymptotes to compute a Michaelis-Menten result for the substrate and compare it with (2.222).

2.7.4.4 Competitive Enzymes

Besides the inhibition by another substrate competing for the same enzyme, it is also possible that a second enzyme E_1 reacts with the substrate S. The reaction scheme then is

$$\begin{aligned} S + E_1 &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C_1 \xrightarrow{k_2} E_1 + P_1 , \\ S + E_2 &\underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} C_2 \xrightarrow{k_4} E_2 + P_2 , \end{aligned} \tag{2.225}$$

[17] The known or desired product has to be measured.

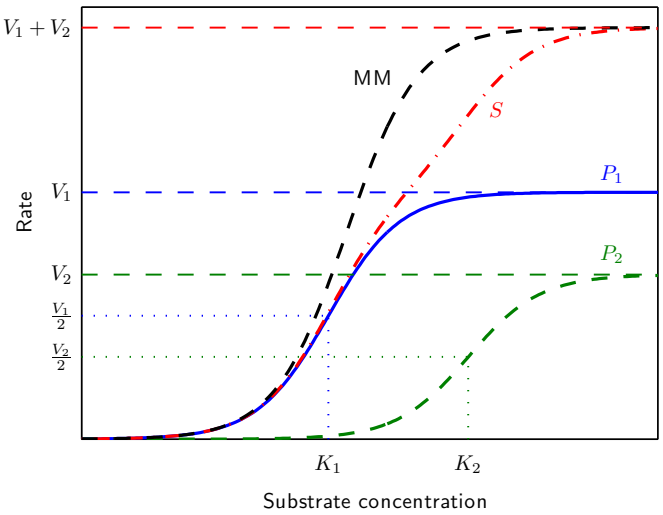


Figure 2.16: The reaction rate for substrate, product P_1, and unused product P_2 as function of the substrate concentration in comparison to a Michaelis-Menten curve with the same asymptotes. Whereas the products obey a Michaelis-Menten-like equation the substrate is not transformable in such a form. The apparent Michaelis constants of the products and the corresponding rate are mentioned by dotted lines in the same color.

where we assume, that the second reaction leads to a product P_2 used for instance in another pathway. The corresponding system of coupled differential equations consists of seven equations. The system

$$\begin{aligned}
\dot{S} &= -\left[k_1\, E_1(t) + k_3\, E_2(t)\right] S(t) + k_{-1}\, C_1(t) + k_{-3}\, C_2(t)\,, && (2.226)\\
\dot{E}_1 &= -k_1\, E_1(t)\, S(t) + \left[k_{-1} + k_2\right] C_1(t)\,, && (2.227)\\
\dot{E}_2 &= -k_3\, E_2(t)\, S(t) + \left[k_{-3} + k_4\right] C_2(t)\,, && (2.228)\\
\dot{C}_1 &= k_1\, E_1(t)\, S(t) - \left[k_{-1} + k_2\right] C_1(t)\,, && (2.229)\\
\dot{C}_2 &= k_3\, E_2(t)\, S(t) - \left[k_{-3} + k_4\right] C_2(t)\,, && (2.230)\\
\dot{P}_1 &= k_2\, C_1(t)\,, && (2.231)\\
\dot{P}_2 &= k_4\, C_2(t) && (2.232)
\end{aligned}$$

can be simplified if we use the conservation laws for the enzymes

$$\begin{aligned}
E_1(0) &= E_1(t) + C_1(t) && (2.233)\\
E_2(0) &= E_2(t) + C_2(t) && (2.234)
\end{aligned}$$

to cancel the differential equations for both enzymes (2.227) and (2.228). If the complexes reside in a steady state the system reduces further. Analogous to the treatment of the Michaelis-Menten model (2.137) we obtain for the products

$$\dot{P}_1 = \frac{V_{\max 1}\, S(t)}{K_{\mathrm{M}1} + S(t)} \qquad (2.235)$$

and

$$\dot{P}_2 = \frac{V_{\max 2}\, S(t)}{K_{\mathrm{M}2} + S(t)}\,, \tag{2.236}$$

where we define the limiting rate $V_{\max 1}$, $V_{\max 2}$, and the Michaelis constants $K_{\mathrm{M}1}$, $K_{\mathrm{M}2}$ in the usual way. The reaction rates of the products remain unaffected by the presence of the other reaction. In steady state we have

$$\begin{aligned} k_1\, E_1\, S &= (k_{-1} + k_2)\, C_1(t) \\ k_3\, E_2\, S &= (k_{-3} + k_4)\, C_2(t) \end{aligned}$$

from which follows for the substrate

$$\begin{aligned} \dot{S} &= -\left[k_2\, C_1(t) + k_4\, C_2(t)\right] \\ &= -\left(\dot{P}_1 + \dot{P}_2\right)\,. \end{aligned} \tag{2.237}$$

With (2.235) and (2.236) we obtain for the substrate

$$\dot{S} = -\left(\frac{V_{\max 1}}{K_{\mathrm{M}1} + S(t)} + \frac{V_{\max 2}}{K_{\mathrm{M}2} + S(t)}\right) S(t) \tag{2.238}$$

an equation which is no more Michaelis-Menten like. Only in the case of $K_{\mathrm{M}1} = K_{\mathrm{M}2}$ the equation (2.238) reduces to such a relation, but because we assume that E_1 and E_2 are different enzymes this is a hard restriction in the treatment of enzyme kinetic reactions.

Another possible reaction scheme is a competitive enzyme that cannot lead to a product P_2 and forms only a second complex C_2. The stoichiometric equations than simplify to

$$\begin{aligned} S + E_1 &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C_1 \xrightarrow{k_2} E + P \\ S + E_2 &\underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} C_2 \end{aligned} \tag{2.239}$$

and the corresponding system of differential equations is

$$\dot{S} = -k_1\, S(t)\, E_1(t) - k_3\, S(t)\, E_2 + k_{-1}\, C_1(t) + k_{-3}\, C_2(t)\,, \tag{2.240}$$

$$\dot{E}_1 = -k_1\, S(t)\, E_1(t) + (k_{-1} + k_2)\, C_1(t)\,, \tag{2.241}$$

$$\dot{E}_2 = -k_3\, S(t)\, E_2(t) + k_{-3}\, C_2(t)\,, \tag{2.242}$$

$$\dot{C}_1 = k_1\, S(t)\, E_1(t) - (k_{-1} + k_2)\, C_1(t)\,, \tag{2.243}$$

$$\dot{C}_2 = k_3\, S(t)\, E_2(t) - k_{-3}\, C_2(t)\,, \tag{2.244}$$

$$\dot{P} = k_2\, C_1(t)\,. \tag{2.245}$$

Taking into account the conservation laws

$$\begin{aligned} E_1(0) &= E_1(t) + C_1(t)\,, \\ E_2(0) &= E_2(t) + C_2(t)\,, \end{aligned}$$

and the steady state assumption for complexes we obtain again the result of a single enzyme kinetic reaction (2.150).

This result reflects some general properties in the discussed treatment of an enzyme kinetic reaction. The reaction rate of products obeys in this framework, conservation laws for enzymes and the steady state assumption for complexes, a Michaelis-Menten-like relation. The different mechanisms lead to new apparent Michaelis constants and limiting rates. Only if more than one product is converted, the absolute value of the rate of change of substrate is not equal to the reaction rate. The rate of change does not follow a Michaelis-Menten law in the case of more products.

2.7.5 Combination of enzyme kinetic reaction and an uncatalyzed conversion reaction

In our previous discussion of the conversion of a substrate into a product we considered only a catalyzed reaction. Because of thermodynamic reasons there is an uncatalyzed reaction too. This is a general property of catalysis. A catalyst cannot make possible a biochemical reaction, if the final state does not exist. But if such a state exists, there is a certain transition probability which might be very small but is not zero, between the initial state (substrate) and the final state (product). Furthermore, a catalyst only affects the rate of change not the equilibrium. The equilibrium is determined by thermodynamic quantities. For our investigation we choose an monomolecular reaction

$$\begin{aligned} S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_2} E + P\,, \\ S \xrightarrow{k_\mathrm{u}} P\,, \end{aligned} \tag{2.246}$$

as the uncatalyzed branch. In analogy to the treatment of Michaelis and Menten we neglect a possible reverse reaction into the substrate. With these assumptions we obtain the set of coupled differential equations

$$\dot{S} = -k_1\,E(t)\,S(t) + k_{-1}\,C(t) - k_\mathrm{u}\,S(t), \tag{2.247}$$

$$\dot{E} = -k_1\,E(t)\,S(t) + [k_{-1} + k_2]C(t)\,, \tag{2.248}$$

$$\dot{C} = k_1\,E(t)\,S(t) - [k_{-1} + k_2]C(t) = -\dot{E}\,, \tag{2.249}$$

$$\dot{P} = k_2\,C(t) + k_\mathrm{u}\,S(t), \tag{2.250}$$

where the rate of change of the substrate and product have an additional term and the equations for the enzyme remains unchanged in comparison to the model of Michaelis and Menten. Taking into account the steady-state assumption for the complex C and the conservation law for the enzyme we are able to reduce the system of differential equations. We obtain that the degradation rate of the substrate and the conversion rate of the product are equal,

$$\dot{P} = -\dot{S} \tag{2.251}$$

$$= \left(\frac{V_\mathrm{max}}{K_\mathrm{M} + S(t)} + k_\mathrm{u}\right) S(t)\,, \tag{2.252}$$

but now with an additional term within the reaction rate. There are two limiting cases. If

$$k_{\mathrm{u}} \ll \frac{V_{\max}}{K_{\mathrm{M}} + S(t)} \tag{2.253}$$

we can neglect the uncatalyzed reaction and the Michaelis-Menten equation is reproduced. The other asymptote we obtain in the limit

$$k_{\mathrm{u}} \gg \frac{V_{\max}}{K_{\mathrm{M}} + S(t)} \,. \tag{2.254}$$

In contrast to the treatment of Michaelis and Menten the reaction rate linearly increases with the substrate concentration. Because of the saturation of the enzyme kinetic reaction the uncatalyzed reaction is now the dominant reaction. The transition between both behaviors is characterised by the relation

$$k_{\mathrm{u}} = \frac{V_{\max}}{K_{\mathrm{M}} + S(t)} \,, \tag{2.255}$$

where both branches have a rate of the same magnitude. For biochemical reactions we can assume $S \gg K_{\mathrm{M}}$ leading to the relation

$$k_{\mathrm{u}}\, S = V_{\max} \,. \tag{2.256}$$

If the direct conversion rate is equal to the limiting rate $V_{\max}$ the direct reaction becomes more important.

In Figure 2.17 the reaction rate is shown in an activity plot as a function of the substrate concentration. For demonstration purposes we vary the rate coefficient of the uncatalyzed conversion by a multiple of the rate coefficient k_2 of the Michaelis-Menten model.

2.7.6 Reversible Michaelis-Menten equation

In the previous sections we neglected a possible reverse branch catalyzing the product into the substrate. Such an approximation assumes that the reverse reaction is very slow. Either the corresponding rate coefficient k_{-2} is nearly zero or the amount of the product is small in comparison to substrate. In the first case, the activation energy for the reverse way is very high. We can consider a small product concentration at the initial phase of the enzyme kinetic reaction and that the product is immediately transferred away. The simplest reaction mechanism of a reversible enzyme-kinetic reaction is

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E + P \,, \tag{2.257}$$

where we add an additional reverse reaction to mechanism of Michaelis and Menten (2.137). The corresponding system of coupled differential equations is

$$\begin{aligned}
\dot{S} &= -k_1\, E(t)\, S(t) + k_{-1}\, C(t) \,, && (2.258)\\
\dot{E} &= -[k_1\, S(t) + k_{-2}\, P(t)]\, E(t) + [k_{-1} + k_2] C(t) \,, && (2.259)\\
\dot{C} &= [k_1\, S(t) + k_{-2}\, P(t)]\, E(t) - [k_{-1} + k_2] C(t) = -\dot{E} \,, && (2.260)\\
\dot{P} &= k_2\, C(t) - k_2\, P(t)\, E(t) \,. && (2.261)
\end{aligned}$$

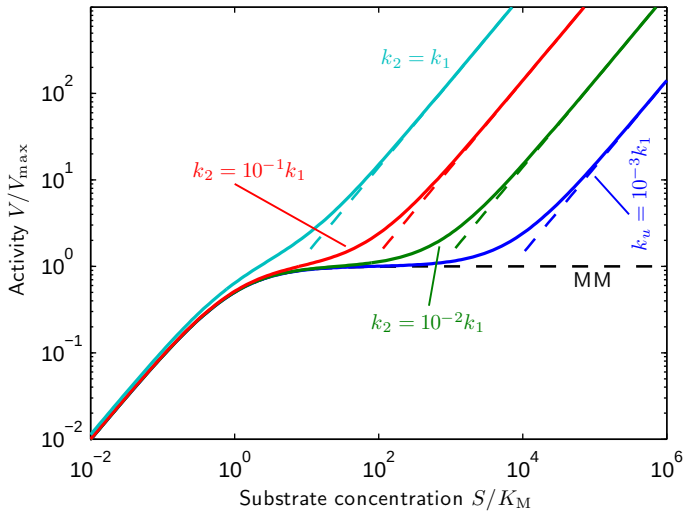


Figure 2.17: Activity for a combination of an enzyme kinetic reaction with a parallel uncatalyzed reaction with the same product as a function of the substrate concentration. The rate coefficient of the direct reaction is a multiple of the rate coefficient k_2 of the Michaelis-Menten model. The usual enzyme kinetic reaction is the dominant part in the conversion for small concentrations fulfilling relation (2.253) and the direct conversion for high concentrations (relation (2.254)). The dashed lines are the corresponding asymptotes.

The rate of change of enzyme, enzyme-substrate complex and product depends now on the product concentration, too. Again, there is a conservation for the enzyme

$$E(0) = E(t) + C(t)$$

relating the transient enzyme and complex concentration to a total enzyme concentration at $t = 0$. It enables us to reduce the system of differential equations. Considering the enzyme-substrate complex in steady state we obtain the balance equation

$$0 = (k_1\, S(t) + k_{-2}\, P(t))\,(E(0) - C(t)) - (k_{-1} + k_2)\, C(t)\,, \tag{2.262}$$

where we apply the above conservation law to replace the transient enzyme concentration. After some straight forward transformation we obtain

$$C(t) = \frac{E(0)\left(S/K_{\mathrm{M}}^{\mathrm{F}} + P/K_{\mathrm{M}}^{\mathrm{R}}\right)}{1 + S/K_{\mathrm{M}}^{\mathrm{F}} + P/K_{\mathrm{M}}^{\mathrm{R}}} \tag{2.263}$$

for the steady state concentration of the complex. In this expression we introduce the abbreviations

$$K_{\mathrm{M}}^{\mathrm{F}} = \frac{k_{-1} + k_2}{k_1} \tag{2.264}$$

and

$$K_{\mathrm{M}}^{\mathrm{R}} = \frac{k_{-1} + k_2}{k_{-2}}\,. \tag{2.265}$$

With this result we get for the reaction rate

$$\begin{aligned}\dot{P} &= -\dot{S} \\ &= \frac{k_2\, E(0)\, S(t)/K_{\mathrm{M}}^{\mathrm{F}} - k_{-1}\, E(0)\, P(t)/K_{\mathrm{M}}^{\mathrm{R}}}{1 + S/K_{\mathrm{M}}^{\mathrm{F}} + P/K_{\mathrm{M}}^{\mathrm{R}}} \\ &= \frac{V_{\max}^{\mathrm{F}}\, S(t)/K_{\mathrm{M}}^{\mathrm{F}} - V_{\max}^{\mathrm{R}}\, P(t)/K_{\mathrm{M}}^{\mathrm{R}}}{1 + S/K_{\mathrm{M}}^{\mathrm{F}} + P/K_{\mathrm{M}}^{\mathrm{R}}}\,,\end{aligned} \tag{2.266}$$

with the limiting rates $V_{\max}^{\mathrm{F}} = k_2\, E(0)$ and $V_{\max}^{\mathrm{R}} = k_{-1}\, E(0)$. The reversible Michaelis-Menten equation can be interpreted as a combination of two irreversible enzyme kinetic reactions. The forward process converts the substrate to the product and the reverse reaction degrades the product. The additional product-dependent terms increase the value of the denominator and decrease the nominator. The sum of both changes decelerates the reaction rate in comparison to the irreversible version.

Because we consider a reversible reaction scheme, the substrate is not completely converted into the product. A chemical equilibrium is established after a certain time. It is defined by the balance equation

$$0 = \frac{V_{\max}^{\mathrm{F}}\, S_{\mathrm{eq}}}{K_{\mathrm{M}}^{\mathrm{F}}} - \frac{V_{\max}^{\mathrm{R}}\, P_{\mathrm{eq}}}{K_{\mathrm{M}}^{\mathrm{R}}} \tag{2.267}$$

from which follows the equilibrium constant

$$K_{\mathrm{eq}} = \frac{P_{\mathrm{eq}}}{S_{\mathrm{eq}}} = \frac{V_{\max}^{\mathrm{F}}\, K_{\mathrm{M}}^{\mathrm{R}}}{V_{\max}^{\mathrm{R}}\, K_{\mathrm{M}}^{\mathrm{F}}} = \frac{k_2}{k_{-1}} \frac{k_1}{k_{-2}}\,. \tag{2.268}$$

2.8 Activation and Deactivation of Proteins by covalent Modifications

Signalling proteins often exist in an inactive form W and an active form W*. The interconversion of the forms is catalyzed by two converter enzymes E_1 and E_2. If we assume that the other substrates and products for modification and demodification are present at constant levels we obtain the reaction scheme [GKJ81]

$$\begin{aligned} W + E_1 &\underset{d_1}{\overset{a_1}{\rightleftharpoons}} (WE_1) \xrightarrow{k_1} W^* + E_1 \\ W^* + E_2 &\underset{d_2}{\overset{a_2}{\rightleftharpoons}} (W^*E_2) \xrightarrow{k_2} W + E_2\,. \end{aligned} \tag{2.269}$$

In the signalling the modification system often consists of a phosphorylation and a dephosphorylation. In this case enzyme E_1 is the kinase and enzyme E_2 the phosphatase. The kinase modifies protein W by attaching a phosphate group to an amino acid that have a free hydroxyl group. The phosphate group is removed from ATP which is converted to ADP during the process of phosphorylation. The phosphatase changes the

state of the protein due to hydrolysis taking away the phosphate group from the hydroxyl group. Hence, the rate constants in reaction scheme (2.269) depends on the concentrations of ATP and water.

Additionally to the usual conservation laws for enzymes

$$\begin{aligned} E_1^{\mathrm{T}} &= E_1(t) + (WE_1)(t) \\ E_2^{\mathrm{T}} &= E_2(t) + (W^*E_2)(t) \ , \end{aligned}$$

we furthermore assume a conservation of the protein itself

$$W^{\mathrm{T}} = W(t) + W^*(t) + (WE_1)(t) + (W^*E_2)(t) \ . \tag{2.270}$$

Due these relations we are able to reduce the system of coupled ordinary differential equations from primary six equations to three equations. As usual, we cancel the equations for the enzymes and, additionally, the equation for the inactive protein W. The remaining differential equations are

$$\frac{dW^*(t)}{dt} = -a_2\, W^*(t)\, E_2(t) + d_2\, (W^*E_2)(t) + k_1\, (WE_1)(t) \tag{2.271}$$

$$\frac{d(WE_1)(t)}{dt} = a_1\, W(t)\, E_1(t) - [d_1 + k_1]\, (WE_1)(t) \tag{2.272}$$

$$\frac{d(W^*E_2)(t)}{dt} = a_2\, W^*(t)\, E_2(t) - [d_2 + k_2]\, (W^*E_2)(t) \ . \tag{2.273}$$

A further simplification arises if we use the quasi-steady state approximation for the intermediate enzyme-substrate complexes and assume that the complex concentrations are negligible

$$W(t) + W^*(t) \gg (WE_1)(t) + (W^*E_2)(t) \tag{2.274}$$

in comparison to the protein concentration. The above assumption gives us the opportunity to relate directly the concentrations of the active and inactive protein to its overall concentration W^{T}. Alternatively, a transformation of rate equation (2.271) into expression in terms of the concentration of W* is not possible.

From (2.272) follows the balance equation for the first complex

$$0 = a_1 \left[W^{\mathrm{T}} - W^*(t)\right] \left[E_1^{\mathrm{T}} - (WE_1)(t)\right] - [d_1 + k_1]\, (WE_1)(t)$$

which has the quasi-steady state solution

$$(WE_1)(t) = \frac{\left[W^{\mathrm{T}} - W^*(t)\right] E_1^{\mathrm{T}}}{K_{\mathrm{M1}} + W^{\mathrm{T}} - W^*(t)} \ . \tag{2.275}$$

In the same way we obtain with

$$(W^*E_2)(t) = \frac{W^*(t) E_2^{\mathrm{T}}}{K_{\mathrm{M2}} + W^*(t)} \ . \tag{2.276}$$

a similar expression for the second intermediate complex. As in the previous section the assumption of the quasi-steady state introduces the Michaelis constants

$$K_{\mathrm{M1}} = \frac{d_1 + k_1}{a_1} \ , \qquad K_{\mathrm{M2}} = \frac{d_2 + k_2}{a_2}$$

into the mathematical representation. The rational form of solutions (2.275) and (2.276) results from the used enzyme conservation. We now insert the obtained expressions for both complexes into the differential equation (2.271) and obtain after some algebraic transformation[18] the dynamic Goldbeter-Koshland function

$$\frac{dW^*(t)}{dt} = \frac{V_1 \left[W^{\mathrm{T}} - W^*(t)\right]}{K_{\mathrm{M1}} + W^{\mathrm{T}} - W^*(t)} - \frac{V_2\, W^*(t)}{K_{\mathrm{M2}} + W^*(t)} \ , \tag{2.277}$$

where we introduce the abbreviations $V_1 = k_1 E_1^{\mathrm{T}}$ and $V_2 = k_2 E_2^{\mathrm{T}}$. This expression becomes slightly manageable, if we use dimensionless representations of the protein concentration and Michaelis constants

$$\overline{W^*}(t) = \frac{W^*(t)}{W^{\mathrm{T}}} \ , \qquad K_{1,2} = \frac{K_{\mathrm{M1,2}}}{W^{\mathrm{T}}} \ .$$

The mole fraction of active protein W* is then determined by the differential equation

$$\frac{d\overline{W^*}(t)}{dt} = \frac{V_1 \left[1 - \overline{W^*}(t)\right]}{K_1 + 1 - \overline{W^*}(t)} - \frac{V_2\, \overline{W^*}(t)}{K_2 + \overline{W^*}(t)} \ . \tag{2.278}$$

The corresponding steady state is defined by the balance equation

$$0 = (V_1 - V_2)\overline{W^*}^2 + (V_1 K_2 + V_2 K_1 + V_2 - V_1)\, \overline{W^*} - V_1 K_2 \tag{2.279}$$

which is a quadratic equation. An analysis of the solutions, see Figure 2.18, shows that only one fulfills all physical restrictions. The concentration of active protein W* can lie only in the range $[0, W^{\mathrm{T}}]$ and of course, it can not be negative. To distinguish between relevant and irrelevant solutions is a physical problem not a mathematical one.

The relevant solution

$$\begin{aligned} \frac{W^*_{\mathrm{SS}}}{W^{\mathrm{T}}} &= G(V_1, V_2, K_1, K_2) \qquad (2.280) \\ &= \frac{\left(\frac{V_1}{V_2} - 1\right) - \left(\frac{V_1}{V_2} + \frac{K_1}{K_2}\right) K_2 + \sqrt{\left[\left(\frac{V_1}{V_2} - 1\right) - \left(\frac{V_1}{V_2} + \frac{K_1}{K_2}\right) K_2\right]^2 + 4\frac{V_1}{V_2}\left(\frac{V_1}{V_2} - 1\right) K_2}}{2\left(\frac{V_1}{V_2} - 1\right)} \end{aligned}$$

Goldbeter-Koshland function

is the so-called Goldbeter-Koshland function [GKJ81] which is often abbreviated as $G(\ldots)$. The steady state is primarily determined by the ratio of V_1 and V_2. Because of the signal transduction in and between cells is carried out by changes of protein concentrations, this means that the ratio of the enzymes E_1 and E_2 or of kinase and

[18] Use the relation $a_2\, K_{\mathrm{M2}} = d_2 + k_2$ to merge the degradation terms in (2.271).

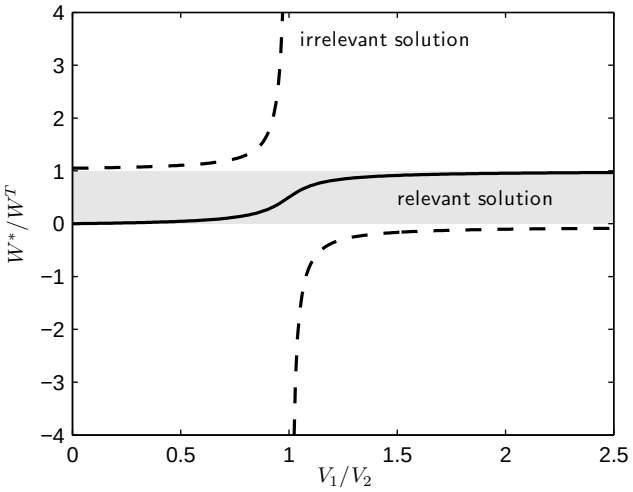


Figure 2.18: The two solutions of balance equation (2.279) as function of the ratio of the limiting rates V_1 and V_2. Only the solution inside the range $[0, W^{\mathrm{T}}]$ is relevant for biological systems. The second one (dashed) breaks physical restrictions and is therefore irrelevant.

phosphatase, respectively, plays a major role in the determination of the signaling state of a protein. The corresponding steady state (2.280) as function of the ratio of limiting rates is investigated in Figure 2.19. If we follow the previous assumption of negligible complex concentrations (2.274), the inactive protein can be calculated from the conservation relation (2.270) as

$$\frac{W}{W^{\mathrm{T}}} = 1 - \frac{W^*}{W^{\mathrm{T}}} \ .$$

Hence, both forms of protein W are directly linked as shown in Figure 2.19, where we use dashed lines for the inactive and solid lines for active state. The solution of the Goldbeter-Koshland function (2.280) has a typical sigmoidal shape. If the rate of deactivation is much greater than the activation corresponding to $V_1/V_2 \ll 1$ there is only a small fraction of activated protein W*. On the other hand, in the limit of a much faster production rate almost all proteins are in the active form. The steady state as function of the ratio V_1/V_2 shows a switch-like behavior [FJ96]. Thereby the sharpness of the transition between a low level and a high level protein concentration or the 'off' and 'on'-state, respectively, is determined by the values of the Michaelis constants K_1 and K_2. As shown in Figure 2.19 small values lead to a sharp transition. Such a behavior is called ultrasensitivity. Because of the reason is a rate law following asymptotically a zero order kinetics[19] this special feature of the covalent modification scheme is called zero-order ultrasensitivity. Apart from the discussed mechanism, there are further mechanisms leading to an ultrasensitive behavior, e.g. cooperativity and allosteric regulation (see for instance [CB04, Seg93]).

[19] In the limits $W^* \ll W \approx W^{\mathrm{T}}$ and $W \ll W^* \approx W^{\mathrm{T}}$ we can reduce the dynamic Goldbeter-Koshland function (2.277) to a zero order kinetic expression whereby we made the assumption $W^{\mathrm{T}} \gg K_{\mathrm{M1,2}}$ and neglect the term corresponding to the degradation of the protein form with small concentration.

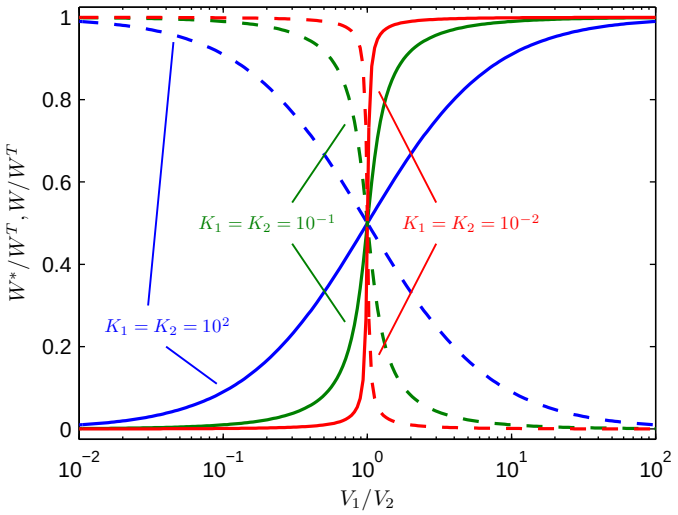


Figure 2.19: Steady state of active (solid lines) and inactive (dashed lines) form of protein W as function of the ratio V_1/V_2 for different Michaelis constants.

In Figure 2.19 we made with $K_1 = K_2$ a very special assumption. In fact, it is very improbably that two reactions have the same kinetic parameters. The steady state is a symmetric function of V_1/V_2. The transition can be characterised by the inflection point located at $V_1/V_2 = 1$ and $\overline{W} = \overline{W^*} = 0.5$. The consequences of different constants are shown in Figure 2.20 where we vary the ratio K_1/K_2. In the left plot we kept the parameter K_1 constant and in the right the parameter K_2. Differences in the plots occur due to the fact that K_1 enters only through the ratio K_1/K_2 into the Goldbeter-Koshland function (2.280) whereas K_2 is also a multiplicative factor. The ratio of both parameters affects shape and position of the curves. The transition between a low level and a high level of activated protein is shifted as function of K_1 and K_2. Furthermore the ultrasensitive properties are strongly influenced. Whereas some parameter values produce ultrasensitive behavior, this feature disappears for another combination of K_1 and K_2.

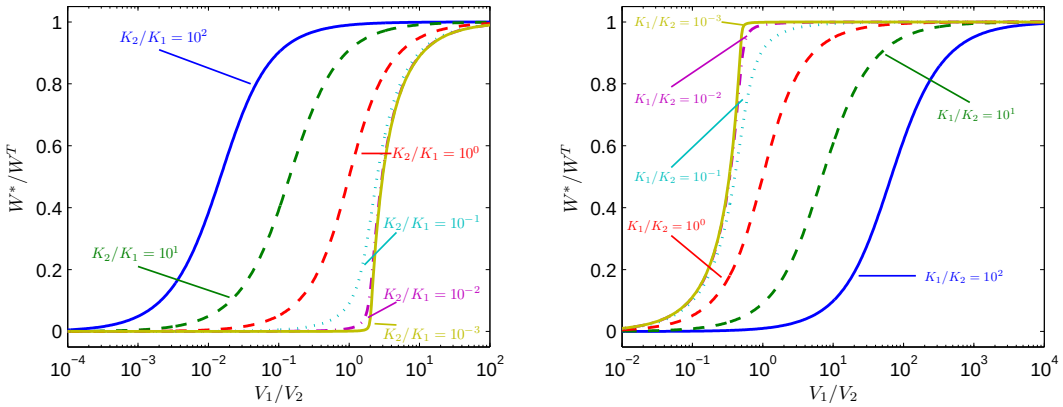


Figure 2.20: Steady state of active protein W* as function of the ratio V_1/V_2. In contrast to Figure 2.19 we now assume different Michaelis constants K_1 and K_2. In the left plot we vary K_2 and in the right plot K_1. The red dashed line corresponds to a symmetric set of parameters as in Figure 2.19.

3 Stochastic modelling

3.1 Why stochastic modelling?

At a coarse level, cell functions are largely determined by spatio-temporal changes in the abundance of molecular components. At a finer level, cellular events are triggered by discrete and random encounters of molecules [PE06]. The discreteness is typical of processes with only a few molecules. Gene transcription is an example of such discrete processes. Each gene is present in only one or two copies in a single cell. That cellular events are discrete and random is supported by many recent experiments [ARM98, ELSS02, RWA02] that have revealed cell-cell variations, even in isogenic cell populations, of transcription (when a gene is copied in the foam of an mRNA transcript) and translation (when the mRNA is used as a template to build proteins).

The above discussion may suggest a deterministic modelling approach at the coarse level (cell function) and a stochastic one at the finer level (gene regulation) [KEBC05, RO05]. However, stochastic modelling is necessary when noise propagation from processes at the fine level changes cellular behaviour at the coarse level.

Stochasticity is not limited to low copy numbers. The binding and dissociation events during transcription initiation are the result of random encounters between molecules [KEBC05]. If molecules are present in large numbers and the molecular events occur frequently, as in Figure 3.1 (left), the randomness would cancel out (both within a single cell and from cell to cell) and the average cellular behaviour could be described by a deterministic model. However, many subcellular processes, including gene expression, are characterised by infrequent (rare) molecular events involving small copy numbers of molecules, as in Figure 3.1 (right), [KEBC05, PE06]. Most proteins in metabolic pathways and signalling networks, realising cell functions, are present in the range 10-1000 copies per cell [BPE00, LKM07, Pau05]. For such moderate/large copy numbers, noise can be significant when the system dynamics are driven towards critical points in cellular systems which operate far from equilibrium [EE03, TJD05, ZYDQ06].

The notion of noise: The term *noise* can be confusing because it is not uniquely defined for all systems. Similarly the classification of noise (e.g. internal/external) can have different meanings for different system. However, noise and its various kinds in gene expression have been clearly defined in [KEBC05, RO05, Pau05]. Following [RO05], noise in gene expression refers to the stochastic variation of a (expressed) protein concentration within isogenic cells having the same history and conditions (environment). Placing two gene reporters in the same cell and quantifying their gene expression (by the abundance of their target proteins) allows the following categorisation of noise (see Figure 2 in [RO05]). *Intrinsic noise* arises from sources that create differences (in the gene expression) between

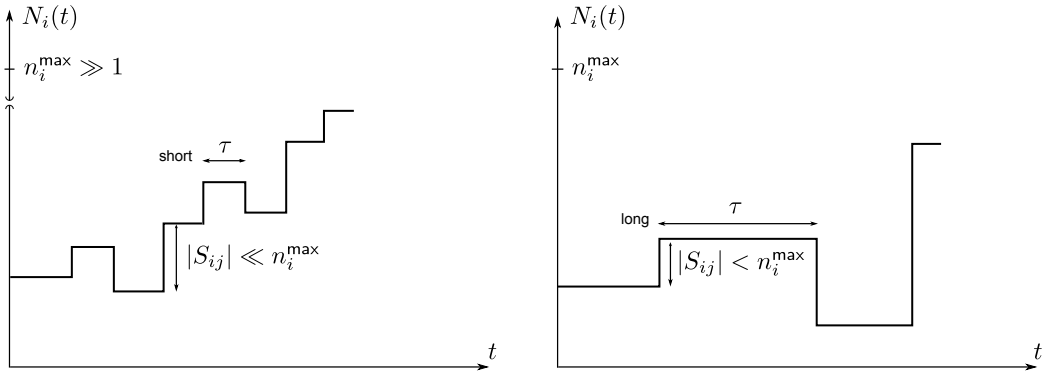


Figure 3.1: Discrete and random nature of chemical reactions. *Left*: large copy numbers and frequent reactions allows for a continuous approximation, which, for an infinitely large system, approaches deterministic rate equations. *Right*: small copy numbers and infrequent reactions requires discrete stochastic approaches leading to the chemical master equation and stochastic simulations. The Figure first appeared in our earlier work [UW09].

the two reporters in the same cell, and *extrinsic noise* arises from sources that have equal effect on the two reporters in the same cell but create differences between two cells. Stochastic events during gene expression would then emerge as intrinsic noise whereas differences between cells will appear as extrinsic noise. Extrinsic noise can be *global* when fluctuations in basic reaction rates affect expressions of all genes, or it can be *pathway-specific*. It is important to realise that extrinsic noise can be theoretically isolated from the system but intrinsic noise is the very essence (discrete nature) of the underlying molecular events and cannot be separated (even hypothetically) from the system. Finally, we like to add that the word "noise" has often negative associations as something undesirable, something that should be removed or avoided. In biology, noise can also have a role and "randomness" may be a better word. In this text the term (noise) is used with the understanding that it may well be something desirable.

3.2 Networks of reactions and species

Imagine molecules of s chemical species homogeneously distributed in a compartment of constant volume V at thermal equilibrium and interacting through r irreversible reaction channels. A reaction channel is either elementary, or may represent a simplification of multiple elementary steps into a single step. Any reversible (bidirectional) reaction can be listed as two irreversible reactions. We symbolise the ith species with X_i and the jth reaction channel with R_j. The abundance of X_i present in the system at time t can be described by the copy number $N_i(t)$. The total copy number n^{tot} of all species indicates how large is the system. Since n^{tot} is always large for large volumes (and small for smaller volumes), the volume V can also indicate the size of the system. Any such parameter can be used as the *system size* and is usually denoted by Ω. The copy number

is usually divide by the system size and the quantity thus obtained

$$X_i(t) = \frac{N_i(t)}{\Omega}$$

is referred to as the concentration. The choice of the system size Ω depends on the kind of concentration one would line to define.

Molar concentrations: For molar concentrations, in units M $\equiv$ mol/L, the system size is chosen as $\Omega = N_\mathrm{A} V$ where the Avogadro's constant

$$N_\mathrm{A} = 6.022 \times 10^{23}\,\mathrm{mol}^{-1}$$

(correct to 4 significant digits) is the number of molecules (or any elementary entities) in one mole. If the volume is given in litres (L) and concentration in molar (M), then the unit of system size Ω is $\mathrm{mol}^{-1} \times \mathrm{L} = \mathrm{M}^{-1}$. The molar unit (M) is too large for very Small concentrations which are better specified in smaller units, namely nanomolar (nM), with the correspond unit $(\mathrm{nM})^{-1}$ for the system size. Suppose the proteins in a cell of volume $V = 30\,\mathrm{fL}$ are measured in nanomolar $(\mathrm{nM})^{-1}$, then the computation of the system size proceeds like this:

$$\Omega = N_\mathrm{A} V = \left(6.022 \times 10^{14}\,(\mathrm{n\,mol})^{-1}\right) \times \left(3 \times 10^{-14}\,\mathrm{L}\right) \approx 18\,(\mathrm{nM})^{-1}\,.$$

Sometime, the volume is chosen so that $\Omega = 1\,(\mathrm{nM})^{-1}$ for the resulting convenience that each concentration is numerically equal to the corresponding copy number. It is left as an exercise to show that such a volume is $V \approx 1.66\,\mathrm{fL}$.

Relative concentrations: For relative concentrations the system size is chosen to give dimensionless concentrations. One simpler way to obtain relative concentrations is by choosing $\Omega = n^\mathrm{tot}$ so that each concentration is just a fraction of two copy numbers. Take the isomerisation reaction as an example where proteins are converted back and forth between the unmodified form U and the modified form W such that the total number n^tot of protein molecules remains constant. The relative concentrations in this example are the fractions,

$$X_\mathrm{U}(t) = \frac{N_\mathrm{U}(t)}{n^\mathrm{tot}} \quad \text{and} \quad X_\mathrm{W}(t) = \frac{n^\mathrm{tot} - N_\mathrm{U}(t)}{n^\mathrm{tot}}$$

of proteins in the inactive and active form, respectively. For some systems it is more appropriate to introduce different scaling parameter Ω_i for each component i if the copy numbers N_i differ in magnitude to keep X_i of the same order $\mathcal{O}(1)$. That can be obtained by defining relative concentration as

$$X_i = \frac{N_i}{C_i \Omega}$$

that is the concentration N_i/Ω divided by a characteristic concentration C_i. In that case, each scaling parameter can be expressed as $\Omega_i = C_i \Omega$. This will of concern to us in the following chapter. In this chapter, we stick to the simpler case.

The reaction channel R_j will be represented by the general scheme

$$\underline{S}_{1j}\mathrm{X}_1 + \cdots + \underline{S}_{sj}\mathrm{X}_s \xrightarrow{k_j} \bar{S}_{1j}\mathrm{X}_1 + \cdots + \bar{S}_{sj}\mathrm{X}_s\,. \tag{3.1}$$

The participation of individual species in the reaction is indicated by *stoichiometries* or *stoichiometric coefficients* written beside them. Thus the coefficient $\underline{S}_{ij}$ (on the left) represents the participation of X_i as reactant and $\bar{S}_{ij}$ (on the right) is the corresponding participation as product. The rate constant, or coefficient, k_j, written over the reaction arrow informs us about the assumed reaction kinetics, and will be explained later. The coefficient will be omitted when we do not want to attach any assumed reaction kinetics to the above reaction scheme. The progress of channel R_j is quantified by the *reaction count* $Z_j(t)$ defined as the number of occurrences of R_j during the time interval $[0, t]$. Another term used for $Z_j(t)$ is the degree of advancement. One occurrence of R_j changes the copy number of X_i by $S_{ij} = \bar{S}_{ij} - \underline{S}_{ij}$, the (i, j)th element of the *stoichiometry matrix* S. During the time interval $[0, t]$, the change in the copy number of X_i contributed by R_j is thus $S_{ij}Z_j(t)$. The total change in the copy number is the sum of contributions from all reactions:

$$N_i(t) = N_i(0) + \sum_{j=1}^{r} S_{ij}Z_j(t)\,. \tag{3.2}$$

Thus changes in copy numbers are determined by stoichiometries and reaction counts. Following the usual vector notation, we write $N(t)$ for the s-vector of copy numbers, $X(t)$ for the s-vector of concentrations and $Z(t)$ for the r-vector of reaction counts. The above conservation relation can be written in the vector notation:

$$N(t) = N(0) + S\,Z(t)\,. \tag{3.3}$$

Dividing by Ω gives the corresponding relation in concentrations:

$$X(t) = X(0) + \frac{S\,Z(t)}{\Omega}\,. \tag{3.4}$$

The copy number $N(t)$, the concentration $X(t)$ and the reaction count $Z(t)$ are alternative ways to describe our system. Description in terms of these *macroscopic variables* is done in the hope that they approximately satisfy an autonomous set of deterministic equations. Two problems stand in making such an effort. First, the reactions are discrete events in time which means that the copy numbers do not vary continuously with time. Secondly, the occurrence time of a reaction is a random quantity because it is determined by a large number of microscopic factors (e.g. positions and momenta of the molecules involved). Therefore, the deterministic description needs a few simplifying assumptions. Alternatively the macroscopic variables are formulated as stochastic processes. Such a stochastic description in terms of macroscopic variables is *mesoscopic*.

Throughout this chapter, we will use the following academic examples. They are chosen to demonstrate different ideas and methods in the discussion.

Standard modification: Consider a protein that can exist in two different conformations or forms, an *unmodified* form U and a *modified* form W. The protein changes between the two forms by the reversible isomerisation reaction

$$\mathrm{U} \underset{k_u}{\overset{k_w}{\rightleftharpoons}} \mathrm{W} \tag{3.5}$$

composed of a modification (forward) channel with rate constant k_u and a demodification (reverse) channel with rate constant k_w. The reaction scheme (3.5) also represents the opening and closing of an ion-channel and similar systems with two-state conformational change. Since the two reactions are not influenced by any external catalyst (e.g. an enzyme), the scheme (3.5) will be referred to as the *standard modification*.

Heterodimerisation: Consider the reversible heterodimerisation

$$\mathrm{X}_1 + \mathrm{X}_2 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{X}_3\,. \tag{3.6}$$

Here the forward reaction is the association of a receptor X_1 and a ligand X_2 to form a heterodimer (complex) X_3. The backward reaction is the dissociation of the heterodimer back into the two monomers. The parameters k_1 and k_2 are the respective association and dissociation rate constants. This example is the simplest one with a bimolecular reaction.

Lotka-Volterra model: Consider a system consisting of two interacting species: X_1 and X_2. The species can either be animals, chemical species or any interacting entities of two kinds. A large amount of a substance A is available for X_1 which reproduces immediately after consuming one unit A. An encounter between the two species results in the disappearance of X_1 and the replication of X_2. This is the only way X_1 dies (degrades) whereas X_2 has a natural death (degradation). The system can be represented by the following scheme

$$\left.\begin{aligned} \mathrm{X}_1 + \mathrm{A} &\xrightarrow{k_1} 2\mathrm{X}_1 + \mathrm{A} \\ \mathrm{X}_1 + \mathrm{X}_2 &\xrightarrow{k_2} 2\mathrm{X}_2 \\ \mathrm{X}_2 &\xrightarrow{k_3} \varnothing \end{aligned}\right\} \tag{3.7}$$

with rate constants k_1, k_2 and k_3. The symbol $\varnothing$ represents the dead (degraded) form of X_2. In general, $\varnothing$ represents any species not included in the model and is referred to as the "null species". The substance A is assumed to be constantly replenished so that the copy number n_A remains constant. This system was first investigated in the context of population biology by [Lot20] and [Vol26] where X_1 was considered as a *prey* to the *predator* X_2 in the same sense that a goat falls a prey to a tiger in a forrest.

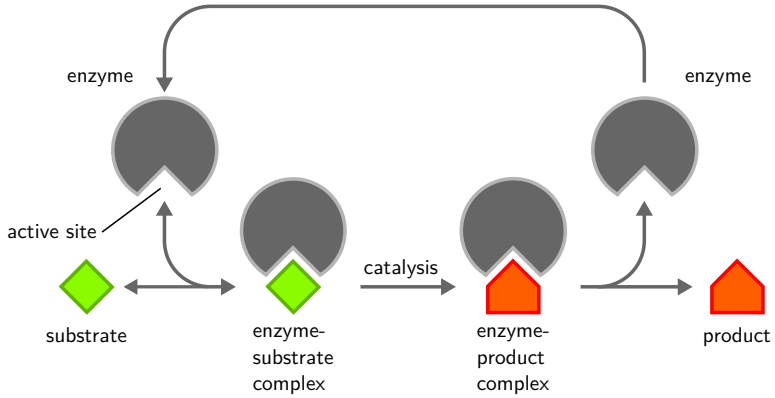


Figure 3.2: Enzyme catalysed conversion of a substrate to product. The enzyme binds to the substrate to make its conversion to product energetically favourable. Figured based on an illustration in [AJL+02].

Enzyme kinetic reaction: In biological systems, the conversion of a substrate to a product may not be a thermodynamically feasible reaction. However, specialised proteins called enzymes ease the job by binding to the substrate and lower the activation energy required for conversion to the product, as depicted in Figure 3.2. Represented in reaction notation

$$\mathrm{E} + \mathrm{S} \longrightarrow \mathrm{E} + \mathrm{P},$$

the the enzymatic reaction is thought to be accomplished in three elementary steps:

$$\mathrm{E} + \mathrm{S} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_3} \mathrm{E} + \mathrm{P}\,. \tag{3.8}$$

Here the enzyme E catalyses a substrate S into a product P that involves an intermediary complex ES. Note that we have not placed any rate constant over the arrow in the original reaction because we do not specify any assumed kinetics in that notation. Later we will learn that it is possible to approximate the three elementary reactions by a single reaction,

$$\mathrm{S} \xrightarrow{k_{\mathrm{eff}}} \mathrm{P}$$

with an effective rate coefficient k_{eff} that represents the assumed approximate kinetics. Intuitively, k_{eff} will be a function of the enzyme abundance. We include this example because this type of reaction appears frequently in the literature. It also serves the purpose of a simple system containing a bimolecular reaction and how a mass conservation leads to a simplified model.

Schlögl model: An autocatalytic, trimolecular reaction scheme, first proposed by [Sch72]

$$\mathrm{A} + 2\mathrm{X} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} 3\mathrm{X}, \quad \mathrm{B} \underset{k_4}{\overset{k_3}{\rightleftharpoons}} \mathrm{X} \tag{3.9}$$

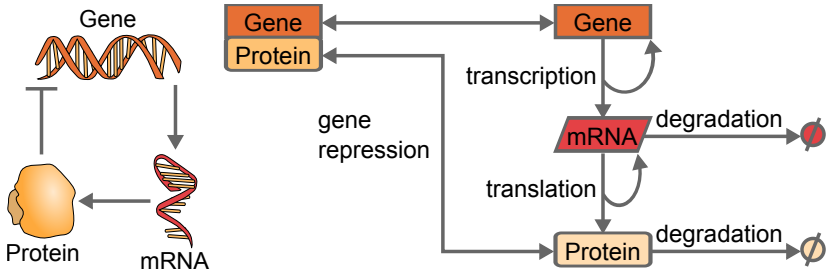


Figure 3.3: Gene regulation: a simpified model. *Left*: cartoon representation. *Right*: reaction pathways. Modified from documentation of the SimBiology toolbox from [Mat].

Here the concentrations A and B are kept constant (buffered). This example serves to illustrate the need for a stochastic approach to model systems with bistability and the associated behaviour known as "stochastic switching".

Gene regulation: This example is included to illustrate gene regulation at the simplest level where the protein product from translation controls (represses) transcription. The regulatory mechamism is simplified by not showing the contributions of RNA polymerase and any cofactors. The protein product from gene expression binds to a regulatory region on the DNA and represses trabscrption. Figure 3.3 illustrates a cartoon representation (on the left) side by side with the corresponding reaction pathways. The reaction scheme for the system is

$$\left.\begin{aligned} \mathrm{G} &\xrightarrow{k_m} \mathrm{G}+\mathrm{M} && \text{(transcription)}\\ \mathrm{M} &\xrightarrow{k_p} \mathrm{M}+\mathrm{P} && \text{(translation)}\\ \mathrm{G}+\mathrm{P} &\underset{k_u}{\overset{k_b}{\rightleftharpoons}} \mathrm{GP} && \text{(binding/unbinding)}\\ \mathrm{M} &\xrightarrow{k_m^-} \varnothing,\quad \mathrm{P} \xrightarrow{k_p^-} \varnothing && \text{(degradation)} \end{aligned}\right\} \tag{3.10}$$

where the gene G is transcribed to the mRNA M with rate constant k_m, the mRNA is translated to the protein P with rate constant k_p, the protein binds to (and represses) the gene with rate constant k_b and unbinds back with rate constant k_u. The mRNA and protein are degraded with respective rate constants k_m^- and k_p^-.

3.3 Deterministic description

Suppose that the reactions occur so frequently that the reaction count $Z(t)$ can be approximated by a continuous quantity $z(t)$. This assumption requires that a large number of reactant molecules are freely available (no crowding) in a large volume so that they can react easily. It also requires that the energy and orientation of reactant molecules favour the reaction, a fact summarised in a rate constant. Large numbers of molecules also mean that a change resulted from a single occurrence of a reaction is relatively small.

That means that the copy number $N(t)$ can be approximated by a continuous quantity $n(t)$. The concentration $X(t)$ is similarly approximated by a continuous quantity $x(t)$. In deterministic description, equations (3.3) and (3.4) respectively translate to

$$n(t) = n(0) + S\, z(t)\,. \tag{3.11}$$

and

$$x(t) = x(0) + \frac{S\, z(t)}{\Omega}\,. \tag{3.12}$$

Taking the time-derivatives we arrive at the deterministic *chemical kinetic equations*:

$$\dot{n}(t) = \Omega S\, v\left(\frac{n(t)}{\Omega}\right) \quad \text{and} \quad \dot{x}(t) = S\, v\,(x(t))\,, \tag{3.13}$$

where $v = \dot{z}/\Omega$ is the *reaction rate* vector whose jth element v_j is the reaction rate of channel R_j. The reaction rate v_j is the R_j reaction count per unit time divided by the system size. The notation $v\,(x(t))$ is based on the assumption that the reaction rate depends only on the concentrations of the reactants. This is a realistic assumption in many reactions at constant temperature. In general, the reaction rate can depend on temperature, pressure, and the concentrations or partial pressures of the substances in the system.

The functional form $v_j(\cdot)$ of the rate of R_j is called the *rate law* (or kinetic law) which is a result of the modelling assumptions about the particular reaction channels. It is only after specifying a rate law that the above ODEs can characterise a particular biochemical reaction network. Without that specification, the above ODEs only represent a consistency condition imposed by mass conservation of reactants and products. There is a large class of chemical reactions in which the reaction rate is proportional to the concentration of each reactant raised to some power:

$$v_j(x) = k_j \prod_{i=1}^{s} x_i^{g_{ij}}, \tag{3.14}$$

which is called *a rate law with definite orders* [Mor08]. The rate constant k_j summarises factors such as activation energy and proper orientation of the reactant molecules for an encounter leading to the reaction. The exponent g_{ij} is the order with respect to the species X_i. The sum of orders for a particular reaction channel is the overall order. For elementary reactions, the orders g_{ij} are the same as the reactant stoichiometries $\underline{S}_{ij}$:

$$v_j(x) = k_j \prod_{i=1}^{s} x_i^{\underline{S}_{ij}}\,. \tag{3.15}$$

This rate law is called *mass action kinetics* [HS96] and is justified by collision theory and transition state theory [Wri04, Hou07, Mor08]. The mass action kinetics should not be confused with the closely related law of mass action which is obtained by equating the forward and backward reaction rates (according to the above rate law) of a reversible

reaction. Reactions that cannot be described by rate laws like (3.14) are said not to have a definite order. For such a reaction, the rate law depends on the assumptions involved in the approximation of the constituent reaction channels. Examples of rate laws for the reactions are Michaelis-Menten kinetics, Hill kinetics and competitive inhibition [Fel97, CB04, HS96].

MATLAB implementation: To implement rate laws of the form (3.14) in MATLAB [Mat], the standard MATLAB data type *function handle* can be employed. We will need MATLAB representations of our mathematical quantities. Let us collect the species concentrations x_i (at a certain time) in an $s \times 1$ column vector `x`, the reaction rate constants k_j in an $1 \times r$ row vector `k`, the reaction rates v_j in an $1 \times r$ row vector `v` and the exponents g_{ij} of the rate-law (3.14) in an $s \times r$ matrix `G`. Then the MATLAB representation of the rate law (3.14), using function handle notation `@`, has the following form:

```
v = @(x) k.*prod(repmat(x,1,r).^G);
```

Note the MATLAB notations `.*` and `.^` for element-wise operations, multiplication and exponentiation. In general, a rate law may not be expressible in the form (3.14). In that case, MATLAB data types *nested function* or *subfunction* could be employed instead of function handles. A nested function representation of the above code will look like the following:

```
function vout = v(x)
  vout = k.*prod(repmat(x,1,r).^G);
end
```

which requires, in the MATLAB workspace, values of variables `k`, `r` and `G`, corresponding respectively to the rate constant k, the number r of reaction channels and the matrix G of exponents g_{ij}. Once such a function (or handle) has been written for the rate law, a MATLAB representation of the chemical kinetic equations (3.13) can be written and numerically solved with the following piece of MATLAB code:

```
dxdt = @(t,x) S*v(x)'; % concentration ODE
[tout,xout] = ode15s(dxdt, [0 tf], x0); % solution
```

Here `x0` is a column vector of initial concentrations and `tf` is the final (stop) time of simulation. The solver `ode15s` returns the column vector `tout` of time points and the solution array `xout` with a row of concentrations for each time point.

Standard modification: The reaction scheme (3.5) depicts the (de)modification of protein between two forms. Suppose there are n^{tot} copies of this protein in a container, $n(t)$

of them being unmodified (in form U) at time t. The two reaction channels progress at the following rates (list on the right)

$$\begin{array}{l|l} \mathrm{U} \xrightarrow{k_w} \mathrm{W} & \dot{z}_w = k_w n \\ \mathrm{W} \xrightarrow{k_u} \mathrm{U} & \dot{z}_u = (n^{\mathrm{tot}} - n)k_u \end{array}$$

and their difference gives the rate equation

$$\dot{n} = -\dot{z}_w + \dot{z}_u = k_u n^{\mathrm{tot}} - (k_w + k_u)n\,. \tag{3.16}$$

With a non-dimensional, $\tau = (k_w + k_u)t$, the ODE takes the form

$$\frac{\mathrm{d}n}{\mathrm{d}\tau} = \frac{k_u n^{\mathrm{tot}}}{(k_w + k_u)} - n\,.$$

The MATLAB implementation of this differential equation and its numerical solution will look like the following piece of code:

```
dndt = @(t,n) k(2)*ntot/(k(1)+k(2))-n; % ODE
[tout,nout] = ode15s(dndt, [0 tf], n0); % solution
```

which understands that values of variables `k`, `ntot`, `tf` and `n0`, corresponding respectively to the rate constant $k = [k_w, k_u]$, the total copy number n^{tot}, the simulation stop time and the initial copy number n^{init}, are available in the MATLAB workspace.

When the modification rate $\dot{z}_w$ is balanced by the demodification rate $\dot{z}_u$, the chemical equilibrium is said to have occurred. In other words, the reversible reaction equilibrates or reaches the steady state. Setting the right hand side of (3.16) to zero gives us the steady state copy number

$$n^{\mathrm{ss}} = \frac{k_u n^{\mathrm{tot}}}{(k_w + k_u)}\,.$$

Thus, in the steady state, a fraction $P_{\mathrm{U}} = {}^{k_u}/_{(k_u+k_w)}$ of proteins are in the unmodified form and a fraction $P_{\mathrm{W}} = {}^{k_w}/_{(k_u+k_w)}$ of them in the modified form. We can also say that a protein spends, on the average, a fraction P_{W} of time in the modified form and a fraction P_{U} of time in the unmodified form. This interpretation proves very useful in reducing complicated reactions into single steps. Suppose the W form participates in another reaction $\mathrm{W} \xrightarrow{k_b} \mathrm{B}$ which occurs on a much slower time scale than two-state conformational changes between U and W, then the complicated reaction scheme

$$\mathrm{U} \underset{k_u}{\overset{k_w}{\rightleftharpoons}} \mathrm{W} \xrightarrow{k_b} \mathrm{B}$$

can be reduced to a single step $\varnothing \xrightarrow{k_b P_{\mathrm{W}} n^{\mathrm{tot}}} \mathrm{B}$ under the fast equilibration assumption for the reversible reaction.

Heterodimerisation: The reaction scheme (3.6) depicts the reversible heterodimerisation: the forward reaction is the association of a receptor X_1 and a ligand X_2 to form the heterodimer X_3; the reverse reaction is the dissociation of the heterodimer back into the monomers. Let $x_1(t)$, $x_2(t)$ and $x_3(t)$ denote the respective time-dependent molar concentrations of X_1, X_2 and X_3.nThe reaction network has to satisfy two conservation relations

$$x_1 + x_3 = q_1, \quad x_2 + x_3 = q_2 \tag{3.17}$$

where q_1 and q_2 are constants determined by the initial conditions. Using these to express x_1 and x_2 in terms of x_3, the system state can be represented by tracking only species X_3. The reaction rates according to the mass action kinetics follow from (3.15) to be (each listed to the right of the corresponding reaction channel)

$$\begin{array}{rl|l} X_1 + X_2 & \xrightarrow{k_1} X_3 & v_1 = k_1 (q_1 - x_3)(q_2 - x_3) \\ X_3 & \xrightarrow{k_2} X_1 + X_2 & v_2 = k_2 x_3 \,. \end{array}$$

The concentration $x_3(t)$ of the complex thus evolves according to

$$\frac{\mathrm{d}x_3}{\mathrm{d}t} = v_1 - v_2 = k_1 (q_1 - x_3)(q_2 - x_3) - k_2 x_3 \,.$$

The MATLAB implementation of this differential equation and its numerical solution will look like the following piece of code:

```
dxdt = @(t,x) k(1)*(q(1)-x)*(q(2)-x)-k(2)*x; % ODE
[tout,xout] = ode15s(dxdt, [0 tf], x0); % solution
```

where values of variables `k`, `q` , `tf` and `x0` corresponding respectively to the rate constant k, the conservation constants q, the simulation stop time and the initial copy number x^{init}, are supposedly available in the MATLAB workspace.

Lotka-Volterra model: The reaction scheme (3.7) depicts the mutual interactions between two kinds of entities X_1 and X_2. Let $n_1(t)$ and $n_2(t)$ denote the copy number of X_1 and X_2, respectively. The copy number of the substance A is assumed to be so large that it is not changed by consumption during the time scale of our interest. The reaction rates according to the mass action kinetics follow from (3.15) to be (listed to the right)

$$\begin{array}{rl|l} X_1 + A & \xrightarrow{k_1} 2X_1 & v_1 = k_1 n_A n_1 \\ X_1 + X_2 & \xrightarrow{k_2} 2X_2 & v_2 = k_2 n_1 n_2 \\ X_2 & \xrightarrow{k_3} \varnothing & v_3 = k_3 n_2 \end{array}$$

Algebraic combination of these reaction rates according to stoichiometry leads the ODEs governing the time courses of $n_1(t)$ and $n_2(t)$:

$$\left.\begin{aligned} \frac{\mathrm{d}n_1}{\mathrm{d}t} &= v_1 - v_2 = \left(k_1 n_A - k_2 n_2\right) n_1, \\ \frac{\mathrm{d}n_2}{\mathrm{d}t} &= v_2 - v_3 = \left(k_2 n_1 - k_3\right) n_2 \,. \end{aligned}\right\} \tag{3.18}$$

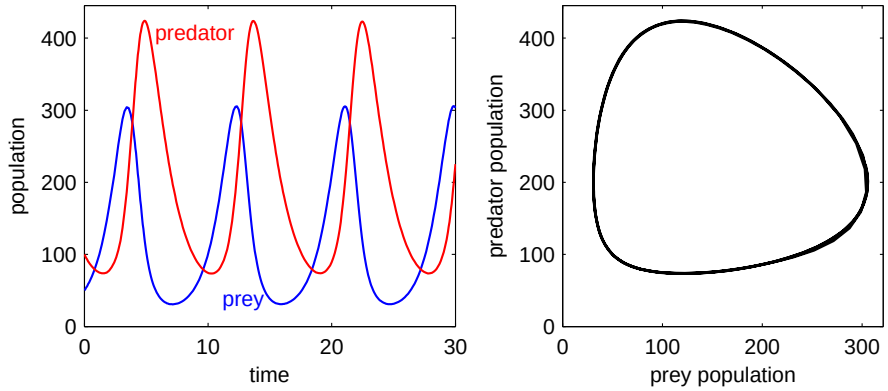


Figure 3.4: Deterministic simulation of the Lotka-Volterra model. *Left*: time course, *Right*: phase plot. Parameters (in $\sec^{-1}$): $k_1 = 1$, $k_2 = 0.005$, $k_3 = 0.6$, Initial population is taken as 50 individuals of prey for 100 individuals of predator.

The MATLAB implementation of this system of ODEs and its numerical solution will look like the following piece of code:

```
k = [1,0.005,0.6]; % rate constants
nA = 1; % substance for consumption
dndt = @(t,n) [k(1)*nA-k(2)*n(2);...
               k(2)*n(1)-k(3)].*n; % ODE
tf = 30; % stop time
n0 = [20;20]; % initial copy number
[tout,nout] = ode45(dndt, [0 tf], n0); % solution
```

Here the first line assigns value to (the array) `k` which corresponds to the rate constant k. The 2nd line assigns an arbitrary value to `nA`, which corresponds to the amount of the available substance n_A. This value is equivalent to incorporating n_A in k_1. The next line defines the function handle `dndt` to represent the system of ODEs in question. The next lines assign values to `tf`, corresponding to the stop time, and `n0`, corresponding to the initial copy number n^{init}. The last line calls an ODE solver to solve the problem and returns the outputs arrays `tout` of time points and `nout` of states (copy numbers) n. With these outputs, the time plot is shown in Figure 3.4 side by side with the associated phase plot.

Enzyme kinetic reaction: For the enzyme kinetic reaction (3.8), write $x_\mathrm{E}(t)$, $x_\mathrm{S}(t)$, $x_\mathrm{ES}(t)$ and $x_\mathrm{P}(t)$ for the respective time-dependent molar concentrations of E, S, ES and P. The solution is usually assumed to respect two conservation laws

$$x_\mathrm{E}(t) + x_\mathrm{ES}(t) = x_\mathrm{E}^{\mathrm{tot}}, \quad \text{and} \quad x_\mathrm{S}(t) + x_\mathrm{ES}(t) + x_\mathrm{P}(t) = x_\mathrm{S}^{\mathrm{tot}} \tag{3.19}$$

where $x_\mathrm{E}^{\mathrm{tot}}$ and $x_\mathrm{S}^{\mathrm{tot}}$ are, respectively, the total concentrations of the enzyme and substrate determined by the initial conditions. We can choose $x = (x_\mathrm{S}, x_\mathrm{ES})^T$ as the state vector sufficient to describe the system because the remaining two variables can be determined

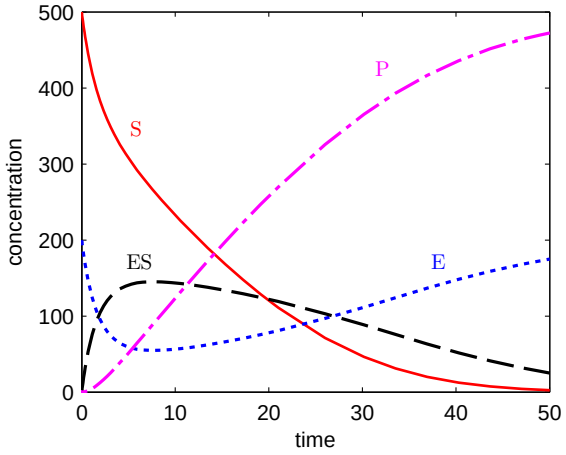


Figure 3.5: Deterministic time-course of the enzyme kinetic reaction. Parameters: $k_1 = 10^{-3}\,(\mathrm{nM\,sec})^{-1}$, $k_2 = 10^{-4}\,\mathrm{sec}^{-1}$, $k_3 = 0.1\,\mathrm{sec}^{-1}$, Initial concentrations: $x_\mathrm{S} = 500\,\mathrm{nM}$, $x_\mathrm{E} = 200\,\mathrm{nM}$, $x_\mathrm{ES} = x_\mathrm{P} = 0\,\mathrm{nM}$.

from the conservation relations above. The channel-wise mass action kinetic law for the reaction scheme (3.8) are (list on the right):

$$\begin{aligned} \mathrm{E} + \mathrm{S} &\xrightarrow{k_1} \mathrm{ES} \\ \mathrm{ES} &\xrightarrow{k_2} \mathrm{E} + \mathrm{S} \\ \mathrm{ES} &\xrightarrow{k_3} \mathrm{E} + \mathrm{P} \end{aligned} \quad \left| \quad \begin{aligned} v_1 &= \left(x_\mathrm{E}^\mathrm{tot} - x_\mathrm{ES}\right) k_1 x_\mathrm{S} \\ v_2 &= k_2 x_\mathrm{ES} \\ v_3 &= k_3 x_\mathrm{ES}\,. \end{aligned} \right.$$

The concentrations evolve according to the following set of nonlinear coupled ODEs

$$\begin{aligned} \frac{\mathrm{d}x_\mathrm{S}}{\mathrm{d}t} &= v_2 - v_1 = k_2 x_\mathrm{ES} - \left(x_\mathrm{E}^\mathrm{tot} - x_\mathrm{ES}\right) k_1 x_\mathrm{S}, \\ \frac{\mathrm{d}x_\mathrm{ES}}{\mathrm{d}t} &= v_1 - v_2 - v_3 = \left(x_\mathrm{E}^\mathrm{tot} - x_\mathrm{ES}\right) k_1 x_\mathrm{S} - \left(k_2 + k_3\right) x_\mathrm{ES}\,. \end{aligned} \tag{3.20}$$

The MATLAB implementation of system of ODEs (3.20) and its numerical solution will look like the following piece of code:

```
k = [1e-3,1e-4,0.1]; % rate constants
xEtot = 200; % total enzyme concentration
dxdt = @(t,x) [-(xEtot-x(2))*k(1), k(2);...
                (xEtot-x(2))*k(1),-(k(2)+k(3))]*x; % ODE
tf = 50;
xStot = 500; % total enzyme concentration
x0 = [xStot;0]; %initial concentrations
[tout,xout] = ode15s(dxdt, [0 tf], x0); % solution
```

Here the first line assigns value to `k` which corresponds to the rate constant k. The 2nd line assigns to `xEtot`, which corresponds to the total enzyme concentration $x_\mathrm{E}^\mathrm{tot}$.

The next line defines the function handle `dxdt` to represent the system of ODEs in question. The next lines assign values to `tf`, corresponding to the stop time, to `xStot`, corresponding to the total enzyme concentration $x_{\mathrm{E}}^{\mathrm{tot}}$, and to `x0`, corresponding to the initial concentration x^{init}. The last line calls an ODE solver to solve the problem and returns the outputs arrays `tout` of time points and `xout` of states (concentrations) x. Once we have these outputs, we can produce the time plot in Figure 3.5.

Schlögl model: For the Schlögl reaction scheme (3.9), write x_{A} and x_{B} for the constant respective concentrations of chemicals A and B, and $x(t)$ for the time-dependent concentration of chemical X. The reaction rates according to the mass action kinetics follow from (3.15) to be (listed on the right)

$$\begin{aligned} \mathrm{A} + 2\mathrm{X} &\xrightarrow{k_1} 3\mathrm{X} \\ 3\mathrm{X} &\xrightarrow{k_2} \mathrm{A} + 2\mathrm{X} \\ \mathrm{B} &\xrightarrow{k_3} \mathrm{X} \\ \mathrm{X} &\xrightarrow{k_4} \mathrm{B} \end{aligned} \qquad \left| \quad \begin{aligned} v_1 &= k_1 x_{\mathrm{A}} x^2 \\ v_2 &= k_2 x^3 \\ v_3 &= k_3 x_{\mathrm{B}} \\ v_4 &= k_4 x \end{aligned} \right.$$

The deterministic ODE then follows to read

$$\frac{\mathrm{d}x}{\mathrm{d}t} = v_1 - v_2 + v_3 - v_4 = k_1 x_{\mathrm{A}} x^2 - k_2 x^3 + k_3 x_{\mathrm{B}} - k_4 x\,. \tag{3.21}$$

The MATLAB implementation of this ODE and its numerical solution will look like the following piece of code:

```
dxdt = @(t,x) k(1)*xA*x^2-k(2)*x^3+k(3)*xB-k(4)*x; % ODE
[tout,xout] = ode15s(dxdt, [0 tf], x0); % solution
```

which assumes that values of variables `k`, `xA`, `xB`, `tf` and `x0`, corresponding respectively to the rate constant k, the fixed concentrations x_{A}, x_{B}, the simulation stop time and the initial concentration x^{init}, are available in the MATLAB workspace.

Gene regulation: For the gene regulation scheme (3.10):

$$\begin{aligned} \mathrm{G} &\xrightarrow{k_m} \mathrm{G} + \mathrm{M} && \text{(transcription)} \\ \mathrm{M} &\xrightarrow{k_p} \mathrm{M} + \mathrm{P} && \text{(translation)} \\ \mathrm{G} + \mathrm{P} &\underset{k_u}{\overset{k_b}{\rightleftharpoons}} \mathrm{GP} && \text{(binding/unbinding)} \\ \mathrm{M} &\xrightarrow{k_m^-} \varnothing,\quad \mathrm{P} \xrightarrow{k_p^-} \varnothing && \text{(degradation)} \end{aligned}$$

write $x_{\mathrm{M}}(t)$, $x_{\mathrm{G}}(t)$ and $x_{\mathrm{P}}(t)$ for the respective time-dependent molar concentrations of mRNA M, the unbound gene G and protein P. The total gene concentration $x_{\mathrm{G}}^{\mathrm{tot}}$ is assumed to be constant so that the bound (repressed) protein concentration is simply

$x_{\mathrm{G}}^{\mathrm{tot}} - x_{\mathrm{G}}$. The reaction rates based on mass action kinetics are $k_m x_{\mathrm{G}}$ for transcription, $k_p x_{\mathrm{M}}$ for translation, $k_b x_{\mathrm{G}} x_{\mathrm{P}}$ for the gene-protein binding, $k_u \left(x_{\mathrm{G}}^{\mathrm{tot}} - x_{\mathrm{G}}\right)$ for the gene-protein unbinding, $k_m^- x_{\mathrm{M}}$ for mRNA degradation and $k_p^- x_{\mathrm{P}}$ for protein degradation. The deterministic system of ODEs for the system can now be written:

$$\left.\begin{aligned} \frac{\mathrm{d}x_{\mathrm{M}}}{\mathrm{d}t} &= k_m x_{\mathrm{G}} - k_m^- x_{\mathrm{M}}, \\ \frac{\mathrm{d}x_{\mathrm{G}}}{\mathrm{d}t} &= k_u \left(x_{\mathrm{G}}^{\mathrm{tot}} - x_{\mathrm{G}}\right) - k_b x_{\mathrm{G}} x_{\mathrm{P}}, \\ \frac{\mathrm{d}x_{\mathrm{P}}}{\mathrm{d}t} &= k_p x_{\mathrm{M}} + k_u \left(x_{\mathrm{G}}^{\mathrm{tot}} - x_{\mathrm{G}}\right) - \left(k_b x_{\mathrm{G}} + k_p^-\right) x_{\mathrm{P}}\,. \end{aligned}\right\} \tag{3.22}$$

In this chapter, we present a stochastic framework for modelling subcellular biochemical systems. In particular, we make an effort to show how the notion of propensity, the chemical master equation (CME) and the stochastic simulation algorithm arise as consequences of the Markov property. This connection is not obvious from the relevant literature in systems biology. We review various analytical approximations of the CME, leaving out stochastic simulation approaches reviewed in [TSB04, Pah08]. Moreover, we sketch interrelationships between various stochastic approaches. The books by [PP01] and [Wil06] inspired section.

3.4 Stochastic formulation

The validity of deterministic macroscopic approaches for description of the averages is limited because the average of a nonlinear function is generally not the same as the function of the average. This was first demonstrated for bimolecular reactions in [Rn53].

Since the occurrence of reactions involve discrete events at the microscopic level, it is impossible to deterministically predict the progress of reactions in terms of the macroscopic variables such as $N(t)$ and $Z(t)$. To account for this uncertainty, one of the macroscopic quantities $N(t), Z(t), X(t)$ is formulated as a stochastic process. Choosing the copy number $N(t)$, a sample value n of the process is the state of our biochemical system under consideration.

How does the process $N(t)$ of copy numbers evolve in time? Starting at time $t = 0$ from some initial state $N(0)$, every sample path of the process remains in state $N(0)$ for a random amount of time W_1 until the occurrence of a reaction takes the process to a new state $N(W_1)$; it remains in state $N(W_1)$ for another random amount of time W_2 until the occurrence of another reaction takes the process to a new state $N(W_1 + W_2)$, and so on as shown in Figure 3.6. In other words, the time-dependent copy number $N(t)$ is a *jump process*.

The stochastic process $N(t)$ is characterised by a collection of state probabilities and transition probabilities. The state probability

$$P(n,t) = \Pr\Big[N(t) = n\Big],$$

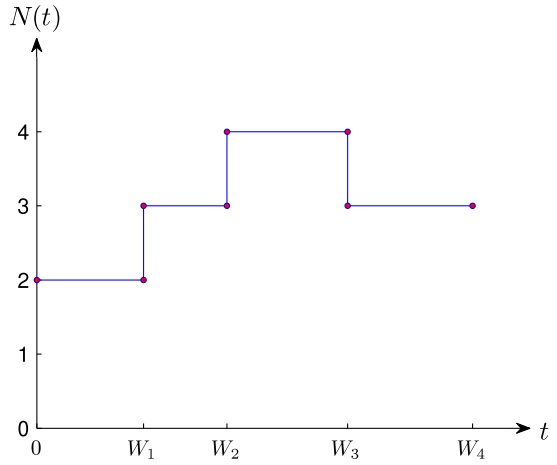


Figure 3.6: A time course realisation of a jump process. Labels W_i on the time axis denote the waiting times.

is the probability that the process $N(t)$ is in state n at time t. The transition probability

$$\Pr\Big[N(t_0+t)=n\,|\,N(t_0)=m\Big]$$

is the conditional probability that process $N(t)$ has moved from state m to state n during the time interval $[t_0, t_0+t]$. The analysis of a stochastic process becomes greatly simplified when the above transition probability depends on: (i) the starting state m but not on the states before time t_0 and (ii) the interval-length t but not on the start time t_0. Property (i) is the well-known *Markov property* and the process with this property is said to be a *Markov process*. The process holding property (ii) is said to be a *homogeneous process*. If the molecules are well mixed and are available everywhere for a reaction (space can be ignored), then the copy number $N(t)$ can be approximately formulated as a homogeneous Markov process in continuous time. In this text, all Markov processes will be assumed to be homogeneous unless stated otherwise. Now we use a simple notation for the above transition probability

$$P(n|m,t)=\Pr\Big[N(t_0+t)=n\,|\,N(t_0)=m\Big]=\Pr\Big[N(t)=n\,|\,N(0)=m\Big]\,. \tag{3.23}$$

It should be remembered that t in the above equation is the length of the time interval. The initial condition is usually fixed and the state probability can be written as a transition probability

$$P(n,t)=P(n|n^0,t)=\Pr\Big[N(t)=n\,|\,N(0)=n^0\Big]\,.$$

The Markov property has two important consequences, explained in the following two sections.

3.4.1 Chapman-Kolmogorov equation

The Markov property places a consistency condition on the transition probabilities. To see that, decompose the transition probability

$$\begin{aligned}
&\Pr\Big[X(t+w) = n \,|\, X(0) = m\Big] \\
&\quad = \sum_{n'} \Pr\Big[X(t+w) = n \,|\, X(t) = n' \cap X(0) = m\Big] \Pr\Big[X(t) = n' \,|\, X(0) = m\Big] \\
&\quad = \sum_{n'} \Pr\Big[X(t+w) = n \,|\, X(t) = n'\Big] \Pr\Big[X(t) = n' \,|\, X(0) = m\Big]
\end{aligned}$$

where the Markov property allows a simplification of the 2nd line leading to the 3rd line. In the compact notation for transition probabilities, the above consistency condition takes the form

$$P(n|m, t+w) = \sum_{n'} P(n|n', w) P(n'|m, t), \tag{3.24}$$

which is known as the "Chapman-Kolmogorov equation" (CKE) for continuous-time Markov processes. This equation expresses the probability of a transition $(m \to n)$ as the summation of probabilities of all transitions $(m \to n' \to n)$ via the intermediate states n'. Figure 3.7 illustrates the idea conveyed by the CKE. It is important to clarify that the CKE is only a consistency condition imposed on every stochastic process by Markov property and cannot characterise a particular process. We need dependence relations between random variables of the process to characterise it. Typically that is achieved by investigating the local behaviour of transition probabilities in a short time interval. Replacing the length w of the time interval of the transition probabilities in (3.24) by Δt and fixing the initial condition, the CKE (3.24) reduces to

$$P(n, t+\Delta t) = \sum_{n'} P(n|n', \Delta t) P(n', t), \tag{3.25}$$

where the transition probabilities away from the fixed initial state have been replaced by the state probabilities. Later we will see that the short-time transition probabilities $P(n|n', \Delta t)$ can be expressed in terms of parameters of the particular process under consideration when certain modelling assumptions about the underlying chemical reactions are made. This will open the door for an analytical characterisation of a particular Markov process.

3.4.2 Memoryless property

Suppose the Markov process $N(t)$ is in state n at time t_0 and let $T_j(n)$ denote the time, in state n, until the occurrence of a reaction R_j takes the process to state $n + S_{\cdot j}$. If the reaction has not occurred during $[t_0, t_0 + w]$, we can write $T_j(n) > w$. This knowledge, however, does not change the uncertainty in time until the next reaction. In other

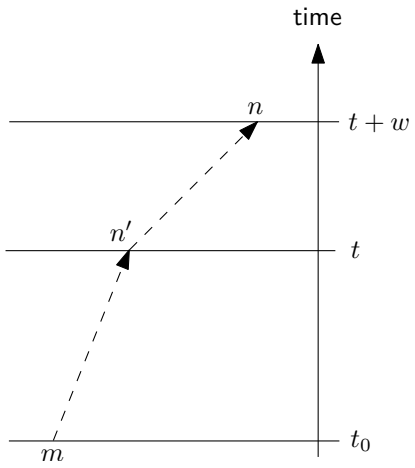


Figure 3.7: Graphical interpretation of the Chapman-Kolmogorov equation. The probability of a transition $m \to n$ can be obtained by summing up the probabilities of all transitions $m \to n' \to n$, via intermediate states n'. Drawing adopted from [Gil92].

words, the process is *memoryless* and its subsequent behaviour is independent of w. The memoryless property can be expressed mathematically as

$$\Pr\Big[T_j(n) > w + t \,|\, T_j(n) > w\Big] = \Pr\Big[T_j(n) > t\Big]\,.$$

and holds true only for the exponential distribution.

The memoryless property, and hence the fact that the times between reactions are exponentially distributed, opens the door for stochastic simulations of biochemical reaction networks. That will be our focus in the following section.

3.5 Propensity as the transition rate

It follows from the previous section that the time $T_j(n)$ until the occurrence of reaction R_j has an exponential distribution with a parameter, say $a_j(n)$. We can thus write

$$\Pr\Big[T_j(n) > t\Big] = \exp\Big(-a_j(n)t\Big), \tag{3.26}$$

for the probability that an R_j reaction will not occur in the next time interval of length t. Using a Taylor expansion, for arbitrarily short interval of length Δt, the above probability can be written as

$$\Pr\Big[T_j(n) > \Delta t\Big] = \exp\Big(-a_j(n)\Delta t\Big) = 1 - a_j(n)\Delta t + o(\Delta t)\,. \tag{3.27}$$

The probability of occurrence of an R_j reaction during the same short interval is complimentary to the above:

$$\Pr\Big[T_j(n) \le \Delta t\Big] = a_j(n)\Delta t + o(\Delta t)\,. \tag{3.28}$$

The parameter $a_j(n)$, which gives the probability per unit time of the occurrence of an R_j reaction in state n, is referred to as the *reaction propensity*.

In a vanishingly short interval, it is highly improbable that a particular reaction will occur more than once. To see that, the probability of two occurrences of R_j during a time interval $[t, t+\Delta t]$ is the joint probability of its first occurrence during $[t, t+\alpha\Delta t]$ and a second occurrence during $(t+\alpha\Delta t, t+\Delta t]$:

$$\begin{aligned}\Pr\Big[T_j(n) \le \alpha\Delta t\Big]\,\Pr\Big[T_j\left(n+S_{\cdot j}\right) \le (1-\alpha)\Delta t\Big] \\ = \Big(a_j(n)\alpha\Delta t + o(\Delta t)\Big)\Big(a_j\left(n+S_{\cdot j}\right)(1-\alpha)\Delta t + o(\Delta t)\Big) = o(\Delta t),\end{aligned}$$

where $0 < \alpha < 1$. Therefore, the probability in (3.28) is equivalent to the probability, in state n, of *one* occurrence (i.e. a unit increment in the reaction count) of R_j during $[t, t+\Delta t]$:

$$\Pr\Big[Z_j(t+\Delta t) - Z_j(t) = 1 \,|\, N(t) = n\Big] = a_j(n)\Delta t + o(\Delta t)\,.$$

The probability distribution, in state n, of the short-time R_j reaction count increment $\Delta Z_j = Z_j(t+\Delta t) - Z_j(t)$ during $[t, t+\Delta t)$ is

$$\Pr\Big[\Delta Z_j = z_j \,|\, N(t) = n\Big] = o(\Delta t) + \begin{cases} a_j(n)\Delta t & \text{if } z_j = 1 \\ 1 - a_j(n)\Delta t & \text{if } z_j = 0 \\ 0 & \text{if } z_j > 1 \end{cases} \tag{3.29}$$

The expected value, conditioned on $N(t) = n$, of this short-time R_j reaction count increment is

$$\begin{aligned}\langle \Delta Z_j \,|\, N(t) = n\rangle &= \sum_{j=0}^{r} z_j \Pr\Big[\Delta Z_j = z_j \,|\, N(t) = n\Big] \\ &= \overbrace{a_j(n)\Delta t}^{z_j=1} + \overbrace{o(\Delta t)}^{z_j>1}\,.\end{aligned}$$

The unconditional expectation of the short-time R_j reaction count increment can be obtained by summing the probabilities $P(n,t)$ weighted by the above conditional expectation over all possible states n:

$$\begin{aligned}\Big\langle \Delta Z_j \Big\rangle &= \sum_n \Big\langle \Delta Z_j \,|\, N(t) = n \Big\rangle P(n,t) \\ &= \sum_n a_j(n) P(n,t)\Delta t + o(\Delta t) \\ &= \big\langle a_j\big(N(t)\big)\big\rangle\, \Delta t + o(\Delta t)\end{aligned}$$

which for vanishingly small Δt leads to the ODE

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\langle Z_j(t)\rangle = \big\langle a_j\big(N(t)\big)\big\rangle\,, \tag{3.30}$$

Thus the mean propensity of a particular reaction can be interpreted as the *average reaction count per unit time.*
The state transition, when in state n, associated with channel R_j will be written as

$$n \xrightarrow{a_j(n)} n + S_{\cdot j}$$

because the completion of one R_j reaction simply adds the jth column of the stoichiometry matrix to the state. The completion of one R_j reaction could also bring the system in state n from another state. That state transition can be written as

$$n - S_{\cdot j} \xrightarrow{a_j\,(n - S_{\cdot j})} n\,.$$

The dependence relation of the propensity on the state n is determined by the system being modelled and reflects the assumptions made about the system. If R_j is an elementary reaction in a well-mixed system, it is reasonable to assume that each possible combination of the R_j reactant molecules has the same probability per unit time, c_j, to react. In other words $c_j \mathrm{d}t$ gives the probability that a particular combination of R_j reactant molecules will react in a short time interval $(t, t + \mathrm{d}t]$. In the literature, c_j is referred to as *stochastic (reaction) rate constant.* If there are $h_j(n)$ different possible combinations of R_j reactant molecules in state n, then the propensity $a_j(n)$ can be written as

$$a_j(n) = c_j h_j(n)\,. \tag{3.31}$$

The form of $h_j(n)$ depends on the order of the reaction R_j.

Zero-order reaction ($\varnothing \to \mathrm{X}$): Since the reaction rate does not depend on the reactant, the propensity is a constant $a(n) = c_j$ if the reaction is a single step. If the reaction is non-elementary, then the propensity is function of copy numbers of any enzymatic chemical species involved.

First-order reaction ($\mathrm{X} \to$): The stochastic reaction rate c_j of this reaction is the probability per unit time of a particular reactant molecule undergoing the reaction. Given n reactant molecules, the probability per unit time of any reactant molecule undergoing the reaction is obtained by summing up the individual probabilities of all n reactant molecules, that is $a_j(n) = c_j n$.

Bimolecular reaction with different species ($\mathrm{X}_1 + \mathrm{X}_2 \to$): The stochastic reaction rate c_j of this reaction is the probability per unit time of a particular pair of reactant molecules undergoing the reaction. Given n_1 copies of reactant X_1 and n_2 copies of reactant X_2, there are $n_1 n_2$ distinct possible pairs of reactant molecules available for the reaction. The probability per unit time of any pair of reactant molecules undergoing the reaction is obtained by summing up the individual probabilities of all $n_1 n_2$ pairs of reactant molecules, that is $a_j(n) = c_j n_1 n_2$.

Bimolecular reaction with repeating species ($2X_1 \rightarrow$): The stochastic reaction rate c_j of this reaction is the probability per unit time of a particular pair of reactant molecules undergoing the reaction. Given n copies of reactant X, there are $^{(n-1)n}/_2$ distinct possible pairs of reactant molecules available for the reaction. The reaction propensity is thus $a_j(n) = {}^{c_j(n-1)n}/_2$.

For an elementary reaction channel R_j of the general form (3.1), with S_{ij} molecules of reactant species X_i, we can write the combinatorial function

$$h_j(n) = \prod_{i=1}^{s} \binom{n_i}{S_{ij}} . \tag{3.32}$$

However, it is highly unlikely that a reaction of order higher than two will result from all its reactants coming together and reacting in one step, for example by collision. A more realistic model will decompose the high order reaction into two or more one step reactions. For non-elementary reactions, the propensity can be computed from the reaction rate by using (3.33). For elementary reactions, the stochastic rate constant c is closely related to the deterministic rate constant, as shown below.

Deterministic and stochastic reaction rates: Using the interpretation of propensity as the mean reaction count per unit time from (3.30), the propensity divided by the system size is analogous to the reaction rate v_j defined earlier in the deterministic framework. Hence, in the stochastic framework, the *stochastic reaction rate* can be defined as

$$\hat{a}(x) = \frac{a(n)}{\Omega} . \tag{3.33}$$

which is analogous to the deterministic reaction rate $v(x)$. The stochastic rate of a given elementary reaction can be computed from (3.33), (3.31) and (3.32) whereas (3.15) can be used for the deterministic reaction rate. The two kinds of reaction rates are given for a few example elementary reactions in Table 3.1. The condition under which the two reaction rates are equal is shown in the corresponding entry of the last column. This also provides the relationship between the stochastic rate constant c_j and the deterministic rate constant k_j. That relationship can be generalised in the following way.

Relationship between the deterministic and stochastic rate constants: Let us find the conditions under which the deterministic and stochastic reaction rates of a general elementary reaction are approximately the same. From (3.15), (3.33), (3.31) and (3.32) we can propose:

$$k_j \prod_{i=1}^{s} x_i^{S_{ij}} = v_j(x) \approx \hat{a}_j(x) = \frac{a_j(n)}{\Omega} = \frac{c_j}{\Omega} \prod_{i=1}^{s} \binom{n_i}{S_{ij}} .$$

The left-most expression is valid only in the deterministic framework which requires large system size, $\Omega \gg 1$. To the extent that this assumption is valid, the combinatorial

Table 3.1: Examples of elementary reactions R_j listed with their propensities $a_j(n)$, stochastic reaction rates $\hat{a}_j(x)$ and deterministic reaction rates $v_j(x)$. The last column shows the condition for equality of the two types of reaction rates. Note that the repeating reactant species require large system size, $\Omega \gg 1$.

R_j	$a_j(n)$	$\hat{a}_j(x)$	$v_j(x)$	$\hat{a}_j = v_j$ if
$\varnothing \xrightarrow{k_j} X$	c_j	$\frac{c_j}{\Omega}$	k_j	$c_j = \Omega k_j$
$X \xrightarrow{k_j} ?$	$c_j n$	$\frac{c_j n}{\Omega}$	$k_j x$	$c_j = k_j$
$X_1 + X_2 \xrightarrow{k_j} ?$	$c_j n_1 n_2$	$\frac{c_j n_1 n_2}{\Omega}$	$k_j x_1 x_2$	$c_j = \frac{k_j}{\Omega}$
$2X \xrightarrow{k_j} ?$	$c_j \frac{(n-1)n}{2}$	$c_j \frac{(n-1)n}{2\Omega}$	$k_j x^2$	$c_j = \frac{2k_j}{\Omega}$
$X_1 + X_2 + X_3 \xrightarrow{k_j} ?$	$c_j n_1 n_2 n_3$	$c_j \frac{n_1 n_2 n_3}{\Omega}$	$k_j x_1 x_2 x_3$	$c_j = \frac{k_j}{\Omega^2}$
$X_1 + 2X_2 \xrightarrow{k_j} ?$	$c_j n_1 \frac{(n_2-1)n_2}{2}$	$c_j \frac{(n_2-1)n_2 n_1}{2\Omega}$	$k_j x_1 x_2^2$	$c_j = \frac{2k_j}{\Omega^2}$

function can be approximated as

$$\begin{aligned}\binom{n_i}{\underline{S}_{ij}} &= \frac{(n_i - \underline{S}_{ij} + 1)\cdots(n_i - 1)\, n_i}{\underline{S}_{ij}!} \\ &= \left(\frac{\Omega^{\underline{S}_{ij}}}{\underline{S}_{ij}!}\right)\left(x_i - \frac{\underline{S}_{ij} - 1}{\Omega}\right)\cdots\left(x_i - \frac{1}{\Omega}\right) x_i \\ &\approx \left(\frac{\Omega^{\underline{S}_{ij}}}{\underline{S}_{ij}!}\right) x_i^{\underline{S}_{ij}} \quad \text{for} \quad \Omega \gg 1\end{aligned}$$

Inserted into the previous equation leads to the stochastic rate constant

$$c_j = \frac{k_j}{\Omega^{K_j - 1}} \prod_{i=1}^{s} (\underline{S}_{ij}!) \tag{3.34}$$

where $K_j = \sum_{i=1}^{s} \underline{S}_{ij}$ is the number of R_j reactant molecules required to collide and possibly result in a single occurrence of the reaction. The above derivation is a refinement of our earlier attempt in [WUKC04].

Relation between the deterministic and stochastic reaction rates: We saw that in general, the stochastic and deterministic reaction rates are not equal. Since the two are equal for infinitely large Ω, the difference between the two is of the order of Ω^{-1}, namely [Elf04]

$$\hat{a}_j(x) = v_j(x) + \mathcal{O}\left(\Omega^{-1}\right) . \tag{3.35}$$

Standard modification revisited: In the standard modification (3.5), the copy number $N(t)$ of the unmodified proteins is a simple birth-death process. Each copy of the unmodified protein U is modified at a rate k_w. Similarly, each copy of the modified protein W is demodified at a rate k_u. Both the modification and the demodification are monomolecular reactions. With $0 < n < n^{\text{tot}}$ unmodified proteins, expressions for the

reaction propensities $a(n)$ are listed here (on the right) together with the corresponding reactions (on the left)

$$\left.\begin{array}{l|l} \mathrm{U} \xrightarrow{k_w} \mathrm{W} & a_w(n) = k_w n \\ \mathrm{W} \xrightarrow{k_u} \mathrm{U} & a_u(n) = (n^{\mathrm{tot}} - n)k_u \end{array}\right\} \tag{3.36}$$

Heterodimerisation revisited: The reversible heterodimerisation (3.6) can be formulated as a 3-component 2-reaction network. Let $N_1(t)$,$N_2(t)$ and $N_3(t)$ denote, the respective copy numbers of the components X_1, X_2 and X_3. The full state has to respect the two conservation relations (3.17) which translate to:

$$N_1(t) + N_3(t) = \Omega q_1, \quad \text{and} \quad N_2(t) + N_3(t) = \Omega q_2$$

where q_1 and q_2 are the conserved concentrations and $\Omega = N_{\mathrm{A}}V$ is the system size. The Markov process $N(t) = N_3(t)$ having states $n = n_3$ is sufficient to describe the system, because the remaining two variables can be determined from the conservation relations above. Subject to those conservation relations, expressions for the channel propensities $a(n)$ in state $n = n_3$ are listed here (on the right) together with the corresponding reactions (on the left):

$$\left.\begin{array}{l|l} \mathrm{X}_1 + \mathrm{X}_2 \xrightarrow{k_1} \mathrm{X}_3 & a_1(n) = k_1 \left(q_1 - \dfrac{n}{\Omega}\right)\left(q_2 - \dfrac{n}{\Omega}\right) \\ \mathrm{X}_3 \xrightarrow{k_2} \mathrm{X}_1 + \mathrm{X}_2 & a_2(n) = k_2 n\,. \end{array}\right\} \tag{3.37}$$

Lotka-Volterra model revisited: The mutual interaction between two kinds of entities depicted in (3.7) is a 2-component 3-reaction network. Let $N_1(t)$ denote the population of the first kind X_1, and $N_2(t)$ that of the 2nd kind X_2. The prey replication and the predation, are of the 2nd order, whereas the predator death is of the first order. Expressions for the channel propensities $a(n)$ in state $n = (n_1, n_2)^T$ are listed here (on the right) together with the corresponding reactions (on the left):

$$\left.\begin{array}{l|l} \mathrm{X}_1 + \mathrm{A} \xrightarrow{k_1} 2\mathrm{X}_1 & a_1(n) = k_1 n_{\mathrm{A}} n_1 \\ \mathrm{X}_1 + \mathrm{X}_2 \xrightarrow{k_2} 2\mathrm{X}_2 & a_2(n) = k_2 n_1 n_2 \\ \mathrm{X}_2 \xrightarrow{k_3} \varnothing & a_3(n) = k_3 n_2 \end{array}\right\} \tag{3.38}$$

Enzyme kinetic reaction revisited: The enzyme kinetic model (3.8) is a 4-component 3-reaction network. Let $N_{\mathrm{E}}(t)$ denote the copy number of the enzyme, $N_{\mathrm{S}}(t)$ that of the substrate, $N_{\mathrm{ES}}(t)$ that of the complex and $N_{\mathrm{P}}(t)$ that of the product. The full state has to respect the two conservation relations (3.19) which translate to:

$$N_{\mathrm{E}}(t) + N_{\mathrm{ES}}(t) = \Omega x_{\mathrm{E}}^{\mathrm{tot}}, \quad \text{and} \quad N_{\mathrm{S}}(t) + N_{\mathrm{ES}}(t) + N_{\mathrm{P}}(t) = \Omega x_{\mathrm{S}}^{\mathrm{tot}},$$

where $x_{\mathrm{E}}^{\mathrm{tot}}$ and $x_{\mathrm{S}}^{\mathrm{tot}}$ are the conserved concentrations and $\Omega = N_{\mathrm{A}}V$. The Markov process

$$N(t) = \big(N_{\mathrm{S}}(t), N_{\mathrm{ES}}(t)\big)^T$$

having states $n = (n_S, n_{ES})^T$ is sufficient to describe the system, because the remaining two variables can be determined from the conservation relations above. The (enzyme-substrate) complex formation is a bimolecular reaction whereas the complex dissociation and the product formation are monomolecular reactions. Expressions for the reaction propensities $a(n)$ in state $n = (n_S, n_{ES})^T$ are listed here (on the right) together with the corresponding reactions (on the left):

$$\left.\begin{array}{ll} \mathrm{E} + \mathrm{S} \xrightarrow{k_1} \mathrm{ES} & a_1(n) = k_1 \left(x_E^{\mathrm{tot}} - \frac{n_{ES}}{\Omega} \right) n_S \\ \mathrm{ES} \xrightarrow{k_2} \mathrm{E} + \mathrm{S} & a_2(n) = k_2 n_{ES} \\ \mathrm{ES} \xrightarrow{k_3} \mathrm{E} + \mathrm{P} & a_3(n) = k_3 n_{ES} \,. \end{array}\right\} \tag{3.39}$$

Schlögl model revisited: For the Schlögl reaction scheme (3.9), write x_A and x_B denote the constant respective concentrations of chemicals A and B, and $N(t)$ for the time-dependent copy number of chemical X. The first two reaction channels, the autocatalysis and its backward dissociation, are trimolecular reactions with two and three identical species, respectively. The last two reaction channels, the synthesis/dissociation of X from/to B, are monomolecular reactions. Expressions for the reaction propensities $a(n)$ in state $n = (n_1, n_2)^T$ are listed here (on the right) together with the corresponding reactions (on the left):

$$\left.\begin{array}{ll} \mathrm{A} + 2\mathrm{X} \xrightarrow{k_1} 3\mathrm{X} & a_1(n) = \hat{k}_1 n\,(n-1) \\ 3\mathrm{X} \xrightarrow{k_2} \mathrm{A} + 2\mathrm{X} & a_2(n) = \hat{k}_2 n\,(n-1)\,(n-2) \\ \mathrm{B} \xrightarrow{k_3} \mathrm{X} & a_3(n) = \hat{k}_3 \\ \mathrm{X} \xrightarrow{k_4} \mathrm{B} & a_4(n) = k_4 n \end{array}\right\} \tag{3.40}$$

where the new rate parameters are defined as

$$\hat{k}_1 = \frac{k_1 x_A}{\Omega}, \quad \hat{k}_2 = \frac{k_2}{\Omega^2}, \quad \hat{k}_3 = k_3 x_B \Omega,$$

in terms of the system size $\Omega = N_A V$.

Reaction network reduction: The last example shows an interesting feature of some biochemical reaction networks. In this example, reaction channels R_1 and R_3 both have the same stoichiometry as far as the abundance $N(t)$, the only state variable, is concerned. A reaction of either of the two channels will take the system from state n to state $n+1$. Similarly, a reaction of either channel R_1 or R_4 will take the system from state n to state $n-1$. Thus, as far the state transitions are concerned, the reaction network (3.40) can be reduced to a birth-death process with birth rate $a^+(n)$ and death rate $a^-(n)$ given by

$$\left.\begin{array}{l} a^+(n) = \hat{k}_1 n\,(n-1) + \hat{k}_3, \\ a^-(n) = \hat{k}_2 n\,(n-1)\,(n-2) + k_4 n\,. \end{array}\right\} \tag{3.41}$$

In general, if the stoichiometry matrix S of a reaction network has identical columns, the network can be reduced by merging the set of reaction channels corresponding to those columns in the above manner.

Gene regulation: For the gene regulation scheme (3.10) write $n_{\mathrm{M}}(t)$, $n_{\mathrm{G}}(t)$ and $n_{\mathrm{P}}(t)$ for the respective time-dependent copy numbers of mRNA M, the unbound gene G and protein P. The total gene copy number $n_{\mathrm{G}}^{\mathrm{tot}}$ is assumed to be constant so that the bound (repressed) protein concentration is simply $n_{\mathrm{G}}^{\mathrm{tot}} - n_{\mathrm{G}}$. The reaction propensities based on mass action kinetics are (each to the right of the corresponding channel):

$$\left.\begin{array}{rcl|l} \mathrm{G} & \xrightarrow{k_m} & \mathrm{G}+\mathrm{M} & \\ \mathrm{M} & \xrightarrow{k_p} & \mathrm{M}+\mathrm{P} & a_m(n) = k_m n_{\mathrm{G}} \\ \mathrm{G}+\mathrm{P} & \xrightarrow{k_b} & \mathrm{GP} & a_p(n) = k_p n_{\mathrm{M}} \\ \mathrm{GP} & \xrightarrow{k_u} & \mathrm{G}+\mathrm{P} & a_b(n) = k_b n_{\mathrm{G}} n_{\mathrm{P}} \\ \mathrm{M} & \xrightarrow{k_m^-} & \varnothing & a_u(n) = k_u\left(n_{\mathrm{G}}^{\mathrm{tot}} - n_{\mathrm{G}}\right) \\ \mathrm{P} & \xrightarrow{k_p^-} & \varnothing & a_m^-(n) = k_m^- n_{\mathrm{M}} \\ & & & a_p^-(n) = k_p^- n_{\mathrm{P}} \end{array}\right\} \tag{3.42}$$

3.6 Stochastic simulation

Time until the next reaction: The probability in state n that no reaction has occurred in an interval of length t follows from (3.26) and independence of reaction channels:

$$\Pr\left[\bigcap_j T_j(n) > t\right] = \prod_j \exp\left(-a_j(n)t\right) = \exp\left(-t\sum_j a_j(n)\right).$$

Hence the time $T_0(n)$ until next reaction taking the process away from state n is exponential with rate parameter

$$a_0(n) = \sum_j a_j(n),$$

which is the *exit rate* (of the system away) from state n.

Index of the next reaction channel: If it is known that a reaction has occurred in state n, the (conditional) probability that it was an R_j reaction is determined as

$$\begin{aligned} \lim_{\Delta t\to 0} \Pr\left[T_j(n) \le \Delta t \,|\, T_0(n) \le \Delta t\right] &= \lim_{\Delta t\to 0} \frac{\Pr\left[T_j(n) \le \Delta t\right]}{\Pr\left[T_0(n) \le \Delta t\right]} \\ &= \lim_{\Delta t\to 0} \frac{a_j(n)\Delta t + o(\Delta t)}{a_0(n)\Delta t + o(\Delta t)} \\ &= \frac{a_j(n)}{a_0(n)}. \end{aligned}$$

Thus the index $J(n)$ of the next reaction known to have occurred in state n is a discrete random variable taking values j with probability

$$\Pr\Big[J(n) = j\Big] = \frac{a_j(n)}{a_0(n)}\,. \tag{3.43}$$

This result, together with the exponentially of T_0, allows a simple procedure to simulate the Markov process: 1) Pick a sample τ from the exponential distribution with rate $a_0(n)$ to realise the time until the next reaction will occur, and 2) pick a sample j from the discrete distribution with probabilities (3.43) to realise the reaction channel.

Simulating the time until the next reaction: The time $T_0(n)$ until the next reaction in state n is an exponential random variable with right-tail distribution function

$$G(t) = \Pr\Big[T_0(n) > t\Big] = \exp\Big(-a_0(n)t\Big)\,.$$

If u_1 is a uniform random number picked from $[0, 1]$, then

$$\tau = G^{-1}(u_1) = -\frac{\log u_1}{a_0(n)} \tag{3.44}$$

is a sample of the time until the next reaction.

Simulating the Index of the next reaction channel: The index $J(n)$ of the next reaction known to have occurred in state n is a discrete random variable with a probability mass function (3.43) and a cumulative distribution function

$$F(j) = \Pr\Big[J(n) \le j\Big] = \sum_{l=1}^{j} \frac{a_l(n)}{a_0(n)}\,.$$

If u_2 is a uniform random number picked from $[0, 1]$ then,

$$j = F^{-1}(u_2) = \min_w \{w : F(w) \ge u_2\}$$

is a sample of the random index $J(n)$. For the range of values taken by J, the above condition is equivalent to

$$F(j-1) < u_2 \le F(j)\,.$$

Multiplying both sides by $a_0(n)$ and plugging values for $F(j)$ gives the following criteria

$$\sum_{l=1}^{j-1} a_l(n) < u_2 a_0(n) \le \sum_{l=1}^{j} a_l(n)\,. \tag{3.45}$$

for j to be a sample of the index $J(n)$ of the next reaction known to have occurred in state n.

Algorithm 1 Gillespie stochastic simulation algorithm (direct method)

1. Initialise the system at $t = 0$ with initial numbers of molecules for each species, $n_1, \ldots, n_s$.
2. For each $j = 1, \ldots, r$, calculate $a_j(n)$ based on the current state n.
3. Calculate the exit rate $a_0(n) = \sum_{j=1}^{r} a_j(n)$. Terminate if $a_0(n) = 0$.
4. Compute a sample τ of the time until the next reaction using (3.44).
5. Update the time $t = t + \tau$.
6. Compute a sample j of the reaction index using (3.45).
7. Update the state n according to R_j. That is put $n = n + S_{\cdot j}$, where $S_{\cdot j}$ denotes jth column of the stoichiometry matrix S.
8. If $t < t_{\max}$, return to Step 2.

Gillespie algorithm [Gil77]: The above two results (3.44) and (3.45) are at the core of the stochastic simulation algorithm also known as the "Gillespie algorithm". The steps involved are listed in Algorithm 1. Over time, many improvements to the original SSA have been made for efficient computation. See [TSB04, Pah08] for extensive reviews.

MATLAB implementation of SSA: To implement the above SSA in MATLAB, we need a MATLAB representation of reactions and species. Suppose we have collected information about our chemical reaction network in a MATLAB structure `R` with fields: `R.S` represents the stoichiometry matrix (the static information - network topology) and `R.a` as a function handle to return the reaction propensity (the kinetic information - rate laws) when given the state `n` as an argument. If we pass this reaction object into a nested function `makessa` written in the M-code 3.1, a function handle `ssa` is returned which can be used as a function to generate sample trajectories according to the Gillespie SSA for the given chemical reaction network. This MATLAB implementation is only for illustration. Efficient implementations of the SSA and its variants is available in MATLAB *SimBiology toolbox* in the form of a stochastic solver. Numerous stochastic simulation packages (implemented in other programming languages) have been developed over time, including [BCF+07, BADG08, RLP07]. An alternative software, which allows arbitrary rate laws (and hence non-elementary reactions), is Cains http://cain.sourceforge.net, a free tool that specialises in computational efficient stochastic simulations.

Standard modification revisited: To use the function handle returned by the function makessa in M-code 3.1, we need to specify the fields of the reaction structure `R`. For the

M-code 3.1: A MATLAB implementation of Gillespie SSA (direct method).

```
function ssa = makessa(R)

rand('state', sum(100*clock));
S = R.S;
s = size(S,1);
a = R.a;
ssa = @gillespie;

    function [tt,nn] = gillespie(t,n,steps)
        tt = zeros(steps,1);
        nn = zeros(s,steps);
        tt(1) = t;
        nn(:,1) = n;
        for i=1:steps
            a0 = sum(a(n));
            if a0==0
                tt(i+1:end) = [];
                nn(:,i+1:end) = [];
                break;
            end
            tau = exprnd(1./a0);
            tt(i+1) = tt(i) + tau;
            j = find(cumsum(a(n))>a0*rand,1);
            n = n + S(:,j);
            nn(:,i+1) = n;
        end
        nn = nn.';
    end

end
```

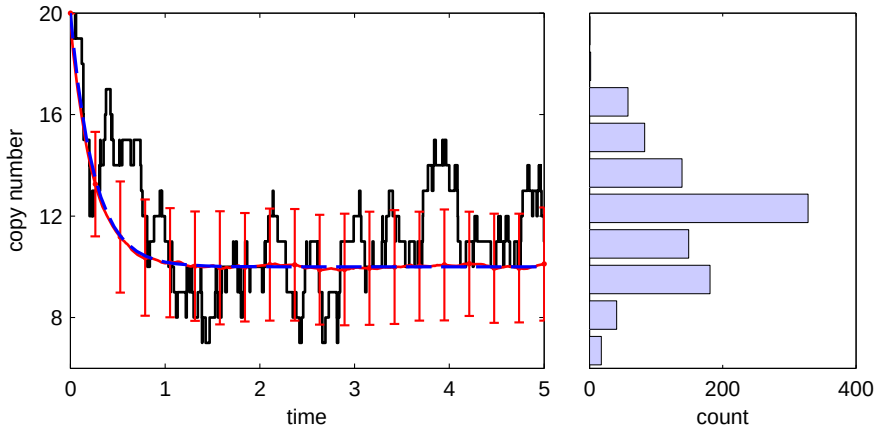


Figure 3.8: Stochastic simulation for the standard modification. *Left*: A single run of the SSA, mean over 1000 runs together with mean±SD (red thread), and solution of the deterministic ODE model (dashed). *Right*: End-point histogram. Parameters: $k_w = k_u = 2\,\mathrm{sec}^{-1}$, Initial conditions: $n = 20$.

isomerisation reaction (3.5) with propensities in (3.36), fields of the `reactions` structure are specified in the following piece of code:

```
R.S = [-1 1];
R.a = @(n)[k(1)*n; k(2)*(ntot-n)];
```

where we note that the propensity field is specified as a function handle unlike the stoichiometry field which is a matrix. Another point to note is that the state variable `n` here is a scalar which is the copy number of unmodified proteins because the copy number of modified proteins is just `ntot-n`. Of course, values of `k` and `ntot` respectively corresponding to the rate constant vector k and the total copy number n^{tot} must be available in the MATLAB workspace. The stochastic simulation results for the 2-species, 2-reaction network (3.5), with propensities (3.36), are shown in Figure 3.8. Identifiability of parameters from time-course data can be demonstrated through stochastic simulation, see the four cases in Figure wherein five sample trajectories are plotted, together with the associated deterministic time-course, for a each parameter value pair.

We see different patterns as the difference $k_w - k_u$ of parameters is changed, while keeping the sum $k_w + k_u$ the same. It can be shown that the sum $k_w + k_u$ determines the mean trajectory whereas the difference $k_w - k_u$ determines the spread of trajectories around the mean. The time-course measurements of mean alone provide information about one fraction $k_u/(k_w+k_u)$ only. To get information about the other fraction $(k_w-k_u)/(k_w+k_u)$ we need time-course measurements of variance as well.

Heterodimerisation revisited: The stochastic simulation results for the reversible heterodimerisation (3.6), with propensities (3.37), are shown in Figure 3.10.

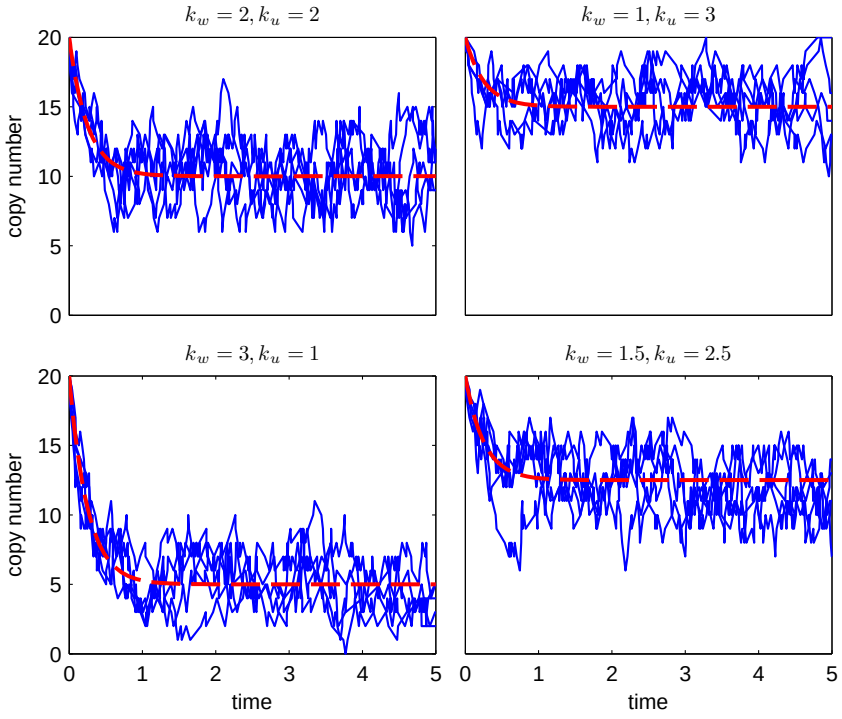


Figure 3.9: Identifiability illustrated through stochastic simulation. Five sample trajectories are shown, together with the associated deterministic time-course, for each parameter value pair. The parameter pairs (k_w, k_u) in sec^{-1} have been selected to satisfy $k_w + k_u = 4$. The total number of protein molecules was chosen to be $n^{\mathrm{tot}} = 10$, initially all unmodified, that is $N(0) = 10$.

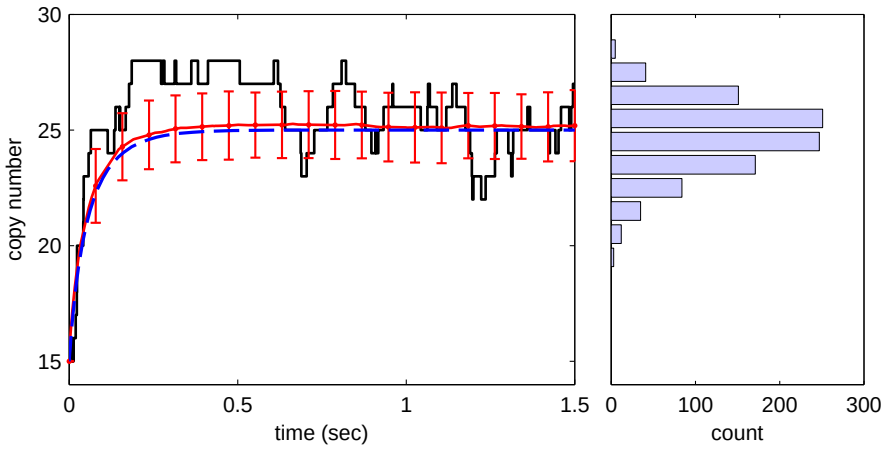


Figure 3.10: Stochastic simulation for the heterodimerisation. *Left*: A single run of the SSA (black stairs), mean over 1000 runs together with mean±SD (red thread), and solution of the deterministic ODE model (grean dashed). *Right*: End-point histogram. Parameters: $k_1 = 1\,\mathrm{sec}^{-1}\,(\mathrm{nM})^{-1}$, $k_2 = 1\,\mathrm{sec}^{-1}$, $V = 1.66\,\mathrm{fL}$ (chosen so that $\Omega = 1\,(\mathrm{nM})^{-1}$), $q_1 = q_2 = 30\,\mathrm{nL}$, Initial concentrations: $x_1 = x_2 = x_3 = 15\,\mathrm{nM}$.

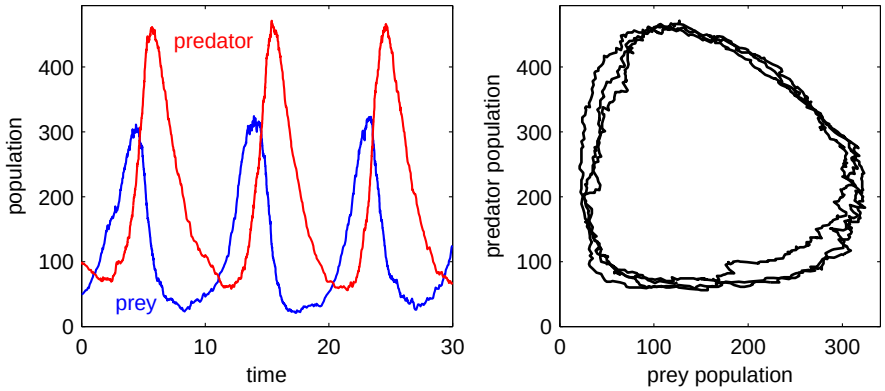


Figure 3.11: Stochastic simulation of the Lotka-Volterra model obtained by one SSA run. *Left*: time course, *Right*: phase plot. Parameters (in $\sec^{-1}$): $k_1 = 1$, $k_2 = 0.005$, $k_3 = 0.6$, Initial populations is taken as 50 individuals of prey and 100 individual of predator.

Lotka-Volterra model revisited: For the Lotka-Volterra system (3.7) with propensities in (3.38), fields of the `R` structure are specified in the following piece of code:

```
R.S = [1,-1,0; 0,1,-1];
R.a = @(n)[k(1)*nA*n(1); k(2)*n(1)*n(2); k(3)*n(2)];
```

with the values of the variables `k` and `nA`, respectively corresponding to the rate constant vector k and the constant copy number n_{A}, available in the MATLAB workspace. Five sample trajectories are shown in Figure 3.11 side by side with the associated phase plot. To see the possibility of species extinction, sample trajectories starting from different initial populations are plotted in Figure 3.12. It can be seen that for some initial populations, species extinction occurs quickly.

Enzyme kinetic reaction revisited: For the 4-species, 3-reaction enzymatic reaction (3.8) with propensities in (3.39), fields of the `R` structure are specified in the following piece of code:

```
R.S = [-1,1,0; 1,-1,-1];
R.a = @(n)[c(1)*n(1)*(nStot-n(2)); c(2)*n(2); c(3)*n(2)];
```

which assumes that the values of the variables `c` and `nStot`, respectively corresponding to the stochastic rate constant vector c and the total copy number $n_{\mathrm{S}}^{\mathrm{tot}}$ of molecules involving the substrate, are available in the MATLAB workspace. Remember that the stochastic rate constant c has to be computed from the deterministic rate constant k according to the relation (3.34). Since, in this example, only the first reaction channel is bimolecular, we have

$$c_1 = \frac{k_1}{\Omega}, \quad c_2 = k_2, \quad c_3 = k_3$$

with the MATLAB representation

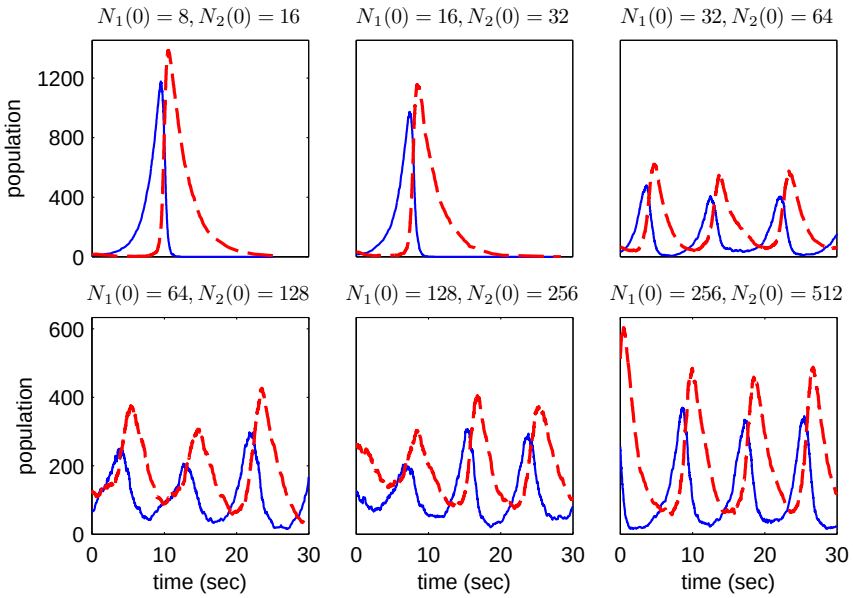


Figure 3.12: Stochastic trajectories of the Lotka-Volterra model of interacting species (3.7) for different initial species populations. The prey and predator population is plotted in *solid* and dashed *lines*, respectively. Note how extinction quickly occurs for some initial populations. Parameters are taken from Figure 3.11.

```
ssz = NA*V; % system size
c = [k(1)/ssz, k(2), k(3)];
```

which understands that values of variable `V` and `NA` respectively corresponding to the volume V and the Avogadro's number N_A are available in the MATLAB workspace. The volume is chosen to be $V == 1.66\,\mathrm{fL}$ so that $\Omega = 1\,(\mathrm{nM})^{-1}$ to have numerically identical values for a species concentrations and the corresponding copy number. To see how the variability among realisations, five different sample trajectories are shown in Figure 3.13 for two scenarios: small/large initial populations on the left/right. The mean species abundance, together with the error bars according to mean±SD, computed over an ensemble of 10000 realisations, are plotted side by side with the species-wise endpoint empirical distribution (PMF) in 3.14. Note that the distributions for enzyme and enzyme-substrate complex have exactly the same shapes and differ only in their means. This is not a coincidence, but a direct consequence of the conservation relation (3.19).

Schlögl model revisited: The Schlögl model (3.9) with propensities in (3.41) is a bistable system with two stable steady states separated by an unstable steady state. In a deterministic framework, such a system settles to that steady state whose basin of attraction is nearer to the initial condition. In a stochastic framework, however, the behaviour is more complex: either steady state may be reached in different realisations regardless of the initial condition. This behaviour, referred to as "stochastic switching"

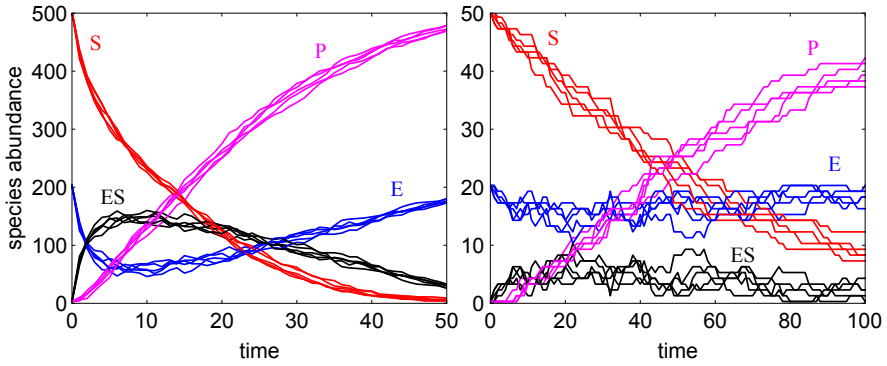


Figure 3.13: Five sample trajectories of species abundance in the enzymatic reaction (3.8). The volume is chosen as $V = 1.66\,\mathrm{fL}$ so that $\Omega = 1\,(\mathrm{nM})^{-1}$ and, hence, a species concentration is numerically the same as the corresponding copy number. Parameters are taken from Figure 3.5. *Left*: large copy numbers $n_S^{\mathrm{tot}} = 500$, $n_{\mathrm{E}}^{\mathrm{tot}} = 200$. *Right*: small copy numbers, $n_S^{\mathrm{tot}} = 50$, $n_{\mathrm{E}}^{\mathrm{tot}} = 20$.

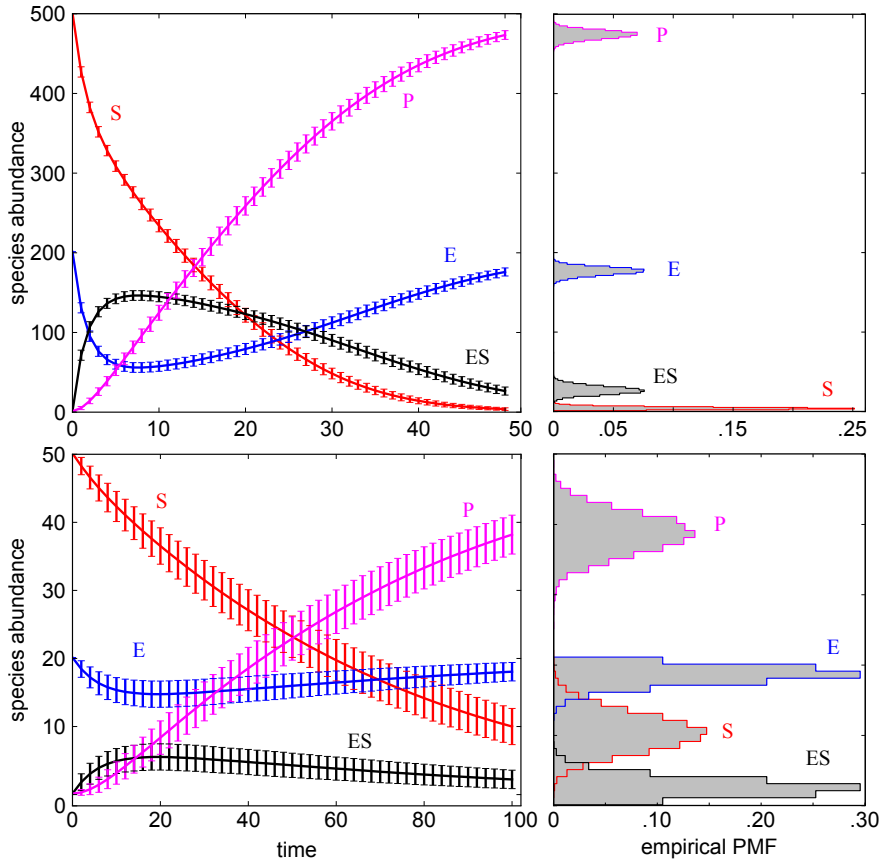


Figure 3.14: Ensemble of 10000 stochastic simulations for the enzymatic reaction (3.8). The parameters and initial copy numbers are taken from Figure 3.13. *Left*: The mean species abundance, together with the error bars according to mean±SD. *Right*: Species-wise end-point empirical distribution (PMF). *Top*: large copy numbers, $n_S^{\mathrm{tot}} = 500$, $n_{\mathrm{E}}^{\mathrm{tot}} = 200$. *Bottom*: small copy numbers, $n_S^{\mathrm{tot}} = 50$, $n_{\mathrm{E}}^{\mathrm{tot}} = 20$.

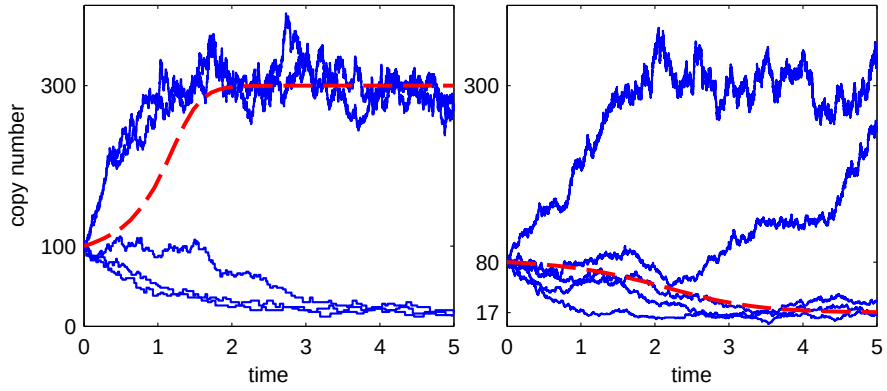


Figure 3.15: Stochastic simulation of the Schlögl model (3.9). Five sample trajectories together with the deterministic time-course (dashed). While the deterministic time-course settles to one of the two stable fixed points, some of the SSA trajectories spread out to other states. *Left*: Initial copy number $N(0) = 100$ is in the basin of attraction of the first stable fixed point $n = 300$. *Right*: Initial copy number $N(0) = 80$ is in the basin of attraction of the 2nd stable fixed point $n = 17$.

in [UIYS06, GUV07], is illustrated here in Figure 3.15 wherein two sets of five sample trajectories starting from, each set starting from a different initial copy number are plotted side by side. The associated deterministic time-course is overlaid on each set. It is easy to see that, while the deterministic time-course settles to one of the stable fixed points, some of the stochastic trajectories spread out to other states. This can be more easily seen in the histogram 3.16. The time varying histogram, which was obtained from 10000 realisations, is unimodal initially and has a bimodal pattern at the end. The MATLAB implementation of the Schlögl model (3.9) with propensities in (3.41), in terms of fields of the `R` structure is left as an exercise for the reader.

Gene regulation revisited: For the gene regulation scheme 3.10 with propensities in (3.42), stochastic simulation results are shown in Figures 3.17-3.18 for two different system sizes. The increased noise can be attributed to the low system size. A clearer picture is depicted in Figure 3.19 where three different measures of noise - the standard deviation, the coefficient of variation and the Fanno factor, are plotted side by side. The gene regulation model (3.10) with propensities in (3.42).

3.7 Chemical master equation

The occurrence of each reaction moves the system from one state to another in the state space. The possible state transitions from/to state n are usually sketched in a state transition diagrams like the one in 3.20, where a transition from one state to another is represented by an arrow that is labelled with the corresponding transition rate. The transition rate of a state transition resulting from a single reaction channel is equal to the reaction propensity of that channel. The transition rate of a state transition result-

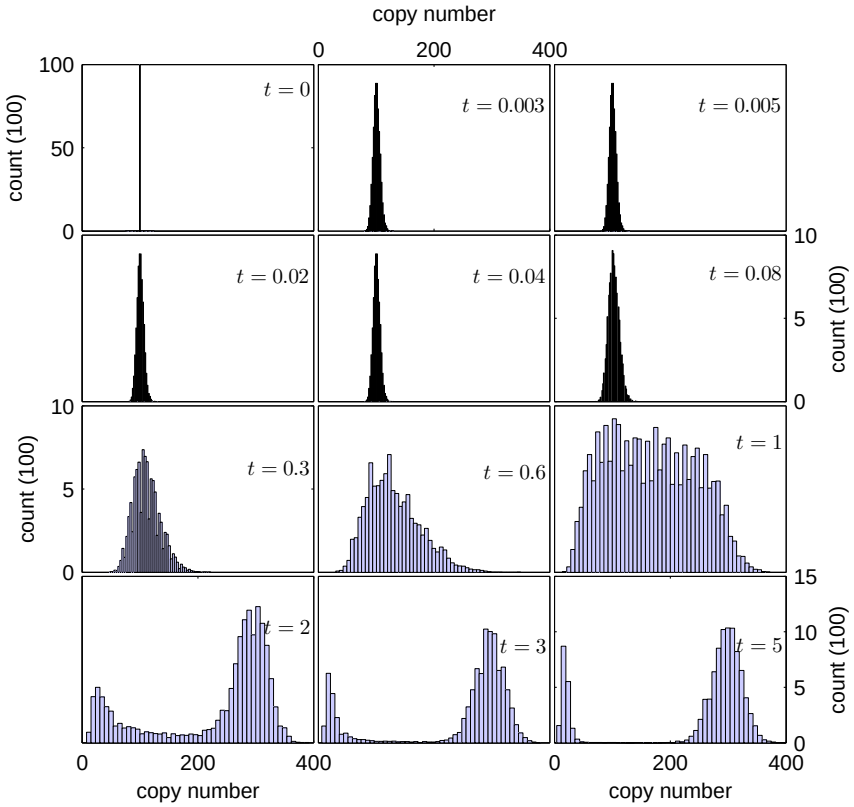


Figure 3.16: Temporal progress of the histogram for the Schlögl reaction.

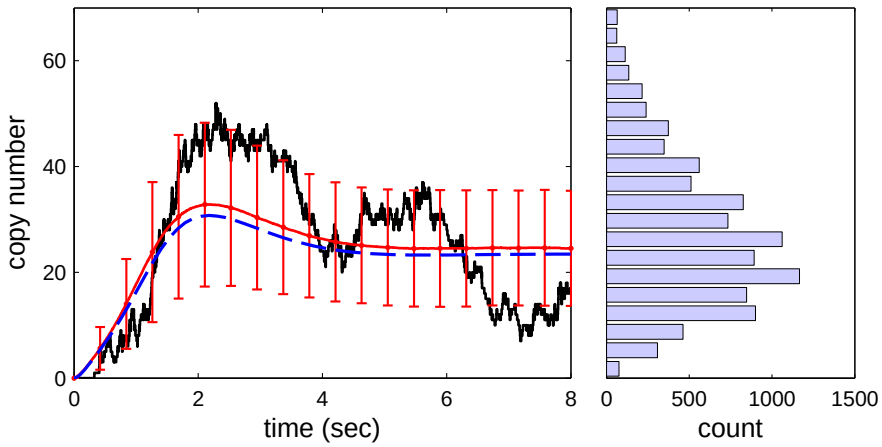


Figure 3.17: Protein abundance arising from large gene abundance (3.10). *Left*: A single run of the SSA. *Right*: End-point histogram.

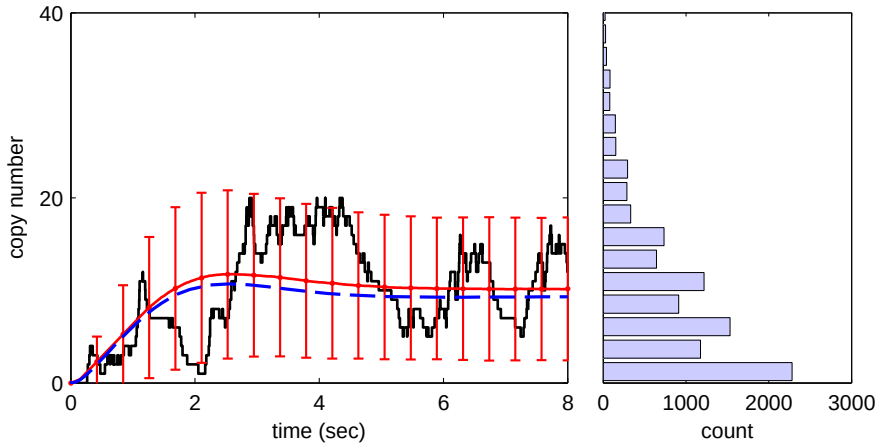


Figure 3.18: Protein abundance arising from small gene abundance (3.10). *Left*: A single run of the SSA. *Right*: End-point histogram.

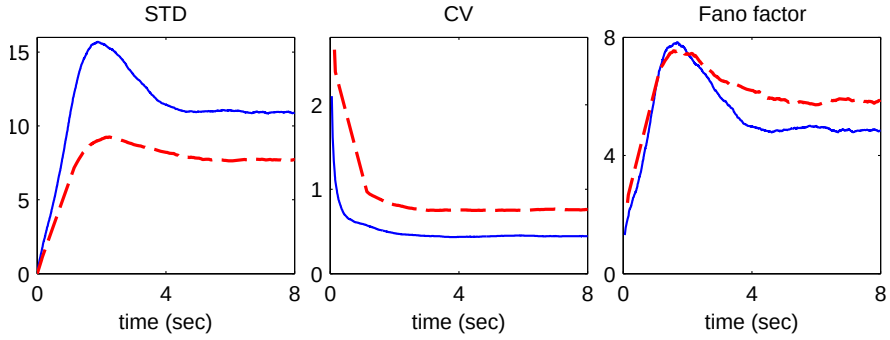


Figure 3.19: Measures of noise in gene regulation model.

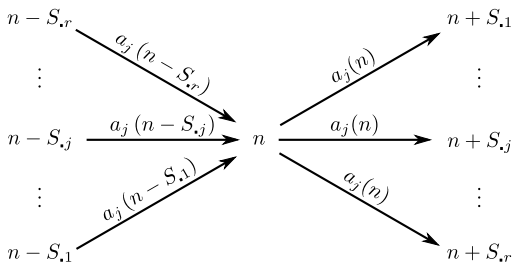


Figure 3.20: State transitions of a generic r-reaction network with network structure encoded in the stoichiometry matrix S and reaction kinetics encoded in the propensity function $a(n)$.

ing from more than one reaction channels is the sum of propensities of those reaction channels.

How does the state probability $P(n,t)$ change with time? To answer this, we need to find an expression for $P(n,t+\Delta t)$, the probability to be in state n after a short time-interval of length Δt. How can the system fall in state n at time $t+\Delta t$? One possibility is that the system was in state n at time t and no reaction occurred during the interval. Otherwise, as obvious from the state transition diagram in Figure 3.20, the state n was reached after the occurrence of one of r possible reactions. Mathematically we can write

$$P(n|n',\Delta t) = o(\Delta t) + \begin{cases} 1 - a_0(n)\Delta t & \text{if } n' = n \\ a_1\left(n - S_{\cdot 1}\right)\Delta t & \text{if } n' = n - S_{\cdot 1} \\ \vdots & \\ a_r\left(n - S_{\cdot r}\right)\Delta t & \text{if } n' = n - S_{\cdot r} \\ 0 & \text{elsewhere.} \end{cases}$$

The term $o(\Delta t)$ represents the probability of arriving in state n by the occurrence of more than one reaction during the interval. Recall that $a_0(n) = \sum_j a_j(n)$ is the exit rate from state n. Substituting the above expressions into (3.25) gives

$$P(n,t+\Delta t) = P(n,t)\left(1 - \sum_{j=1}^{r} a_j(n)\Delta t\right) + \sum_{j=1}^{r} P\left(n - S_{\cdot j}, t\right) a_j\left(n - S_{\cdot j}\right)\Delta t + o(\Delta t),$$

which for vanishingly short Δt can be re-arranged as the *chemical master equation* (CME):

$$\frac{\partial}{\partial t}P(n,t) = \sum_{j=1}^{r}\left[a_j\left(n - S_{\cdot j}\right)P\left(n - S_{\cdot j}, t\right) - a_j(n)P(n,t)\right]. \tag{3.46}$$

We will switch between the two alternative notations $\frac{\mathrm{d}}{\mathrm{d}t}\phi(t)$ and $\frac{\mathrm{d}\phi}{\mathrm{d}t}$ for any scalar quantity $\phi(t)$. We will prefer the later when dependence on time variable is implicitly clear.

Remember that the CME above has been written with an understanding that the functional form of the propensities $a_j(n)$ has been specified for the process under study. Without that specification, the CME, similar to the CKE, merely represents a consistency condition imposed by the Markov property.

Using a negative-shift operator $\mathbb{E}_j$ for each reaction channel defined by its effect

$$\mathbb{E}_j\, f(n) = f\left(n + S_{\cdot j}\right)$$

on an arbitrary scalar function $f(n)$ of s-vector n, the CME can be written in the alternative form

$$\frac{\partial}{\partial t}P(n,t) = \sum_{j=1}^{r}\left(\mathbb{E}_j^{-1} - 1\right)a_j(n)P(n,t). \tag{3.47}$$

$$n-1 \underset{k_w n}{\overset{k_u\,(n^{\mathrm{tot}}-n+1)}{\rightleftarrows}} n \underset{k_w(n+1)}{\overset{k_u\,(n^{\mathrm{tot}}-n)}{\rightleftarrows}} n+1$$

Figure 3.21: State transitions of the standard modification (3.5).

The CME is a differential-difference equation (*differential* in time t and *difference* in states n). In other words, there is one equation for each state n. Since there is potentially a large number of possible states, it is any attempt to solve the CME analytically or even numerically will be impractical, unless one is dealing with a very simple system such as the isomerisation reaction (3.5) that has just one state variable n and only two channels (state transitions).

Standard modification revisited: For the standard modification (3.5) with propensities in (3.36), the state transition diagram is given in Figure 3.21. Based on these state transitions, the CME for this example reads:

$$\begin{aligned}\frac{\partial}{\partial t}P(n,t) = k_w\Big[(n+1)P(n+1,t) - nP(n,t)\Big] \\ + k_u\Big[\big(n^{\mathrm{tot}}-n+1\big)\,P(n-1,t) - \big(n^{\mathrm{tot}}-n\big)\,P(n,t)\Big]\,. \quad (3.48)\end{aligned}$$

Note that this CME must respect the boundary conditions with respect to $n = 0, 1, \ldots, n^{\mathrm{tot}}$. That is, $P(n,t) = 0$ for $0 > n > n^{\mathrm{tot}}$. We can gain some insight into the dynamics described in the above CME by setting $n^{\mathrm{tot}} = 1$, which corresponds to a single molecule (in isolation) that can exist either in the unmodified form U with probability $P_{\mathrm{U}}(t) \overset{\mathrm{def}}{=} P(1,t)$, or in the modified form W with probability $P_{\mathrm{W}}(t) \overset{\mathrm{def}}{=} P(0,t) = 1 - P_{\mathrm{U}}(t)$. The single molecule version of the above CME turns out to be

$$\frac{\mathrm{d}}{\mathrm{d}t}P_{\mathrm{U}}(t) = -k_w P_{\mathrm{U}}(t) + k_u\,(1 - P_{\mathrm{U}}(t)) = k_u - (k_w + k_u)\,P_{\mathrm{U}}(t)$$

where we have used the boundary condition $P(2,t) = 0$. Suppose that the protein molecule is initially unmodified, that is $P_{\mathrm{U}}(0) = 1$. Then the above single molecule CME can be solved for $P_{\mathrm{U}}(t)$ to yield

$$P_{\mathrm{U}}(t) = \frac{k_u + k_w \mathrm{e}^{-(k_w+k_u)t}}{k_w + k_u}\,. \quad (3.49)$$

The probability $P_{\mathrm{U}}^{\mathrm{ss}}$ of ending up in the unmodified state can now be determined by setting t infinitely large (corresponding to the steady state distribution):

$$P_{\mathrm{U}}^{\mathrm{ss}} = P_{\mathrm{U}}(\infty) = \frac{k_u}{k_w + k_u}\,. \quad (3.50)$$

The probability $P_{\mathrm{W}}^{\mathrm{ss}}$ of ending up in the modified state is complementary to the above:

$$P_{\mathrm{W}}^{\mathrm{ss}} = 1 - P_{\mathrm{U}}^{\mathrm{ss}} = \frac{k_w}{k_w + k_u}\,. \quad (3.51)$$

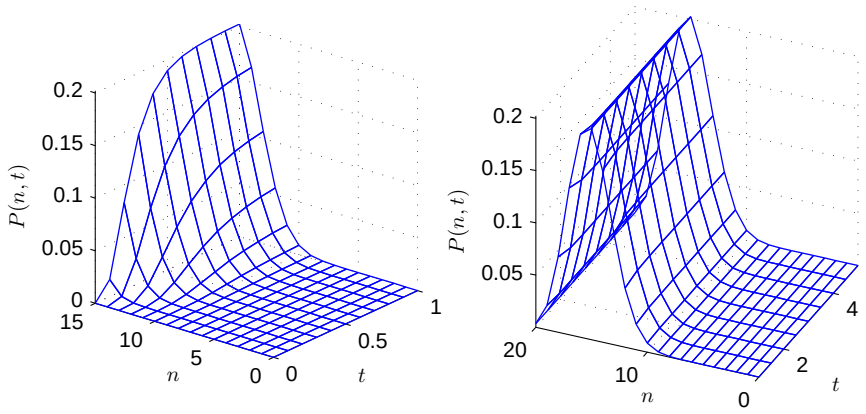


Figure 3.22: Temporal progress of the probability distribution for the standard modification. The PMF $P(n,t)$, for the copy number of unmodified proteins, is plotted during two time subintervals: $0 < t < 1$ (*left*) and $1 < t < 5$ (*right*). The parameters were chosen as $k_w = 3$ and $k_u = 1$ both in sec^{-1}. Initially, all the proteins are assumed to be unmodified.

$$n-1 \underset{k_2 n}{\overset{k_1\left(q_1-\frac{n-1}{\Omega}\right)\left(q_2-\frac{n-1}{\Omega}\right)}{\rightleftarrows}} n \underset{k_2(n+1)}{\overset{k_1\left(q_1-\frac{n}{\Omega}\right)\left(q_2-\frac{n}{\Omega}\right)}{\rightleftarrows}} n+1$$

Figure 3.23: State transitions of the heterodimerisation reaction (3.6).

The last two results make intuitive sense when chemical equilibrium is assumed: the respective fractions of time the protein spends in the unmodified and modified states are $k_u/(k_u+k_w)$ and $k_w/(k_u+k_w)$. Having determined the probability P_{U} of a single molecule to be unmodified, the probability $P(n,t)$ that n out of all the available n^{tot} are unmodified is simply the PMF of the binomial distribution namely

$$P(n,t) = \binom{n^{\text{tot}}}{n} \left(P_{\text{U}}(t)\right)^n \left(1 - P_{\text{U}}(t)\right)^{n^{\text{tot}}-n} . \tag{3.52}$$

We have thus found the solution (3.52) to the original CME (3.48) through an indirect, but insightful, procedure. This will, however, not be tractable for every case. The progress, in time, of the probability distribution of the copy number $N(t)$ (of molecules in inactive form) is shown in Figure 3.22, wherein the PMF is plotted during two time subintervals.

Hetrodimerisation revisited: For the reversible hetrodimerisation (3.6) with propensities (3.37), the state transition diagram is shown in Figure 3.23. Based on these state

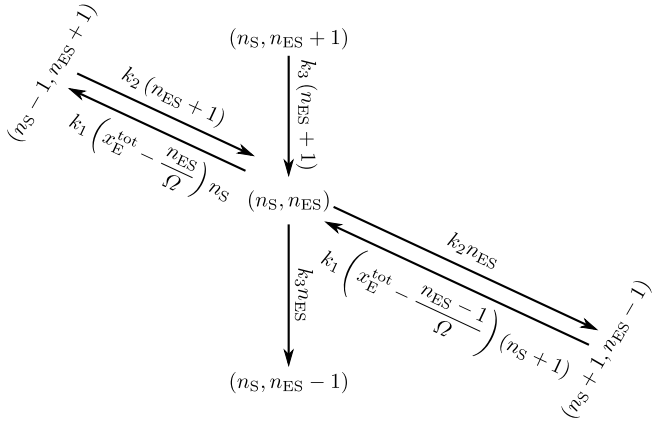


Figure 3.24: State transitions of the enzyme kinetic reaction (3.8).

transitions, the CME for this example reads:

$$\frac{\partial}{\partial t}P(n,t) = k_1\left(q_1-\frac{n-1}{\Omega}\right)\left(q_2-\frac{n-1}{\Omega}\right)P(n-1,t) - k_1\left(q_1-\frac{n}{\Omega}\right)\left(q_2-\frac{n}{\Omega}\right)P(n,t) + k_2\Big[(n+1)P(n+1,t)-nP(n,t)\Big].$$

Enzyme kinetic reaction revisited: For the enzymatic reaction (3.8) with propensities in (3.39), the state transition diagram is given in Figure 3.24. Following these state transitions, the CME can be written as:

$$\begin{aligned}\frac{\partial}{\partial t}P\left(n_S,n_{ES},t\right) = {} & k_1\left(x_S^{tot}-\frac{n_{ES}-1}{\Omega}\right)\left(n_S+1\right)P\left(n_S+1,n_{ES}-1,t\right)\\ & - k_1\left(x_S^{tot}-\frac{n_{ES}}{\Omega}\right)n_S P\left(n_S,n_{ES},t\right)\\ & + k_2\Big[\left(n_{ES}+1\right)P\left(n_S-1,n_{ES}+1,t\right)-n_{ES}P\left(n_S,n_{ES},t\right)\Big]\\ & + k_3\Big[\left(n_{ES}+1\right)P\left(n_S,n_{ES}+1,t\right)-n_{ES}P\left(n_S,n_{ES},t\right)\Big].\end{aligned}$$

Lotka-Volterra model revisited: For the Lotka-Volterra model (3.7) with propensities in (3.38), the state transition diagram is given in Figure 3.25. From these state transitions, the CME for this example reads:

$$\begin{aligned}\frac{\partial}{\partial t}P\left(n_1,n_2,t\right) = {} & k_1\Big[\left(n_1-1\right)P\left(n_1-1,n_2,t\right)-n_1P\left(n_1,n_2,t\right)\Big]\\ & + k_2\Big[\left(n_1+1\right)\left(n_2-1\right)P\left(n_1+1,n_2-1,t\right)-n_1n_2P\left(n_1,n_2,t\right)\Big]\\ & + k_3\Big[\left(n_2+1\right)P\left(n_1,n_2+1,t\right)-n_2P\left(n_1,n_2,t\right)\Big].\end{aligned}$$

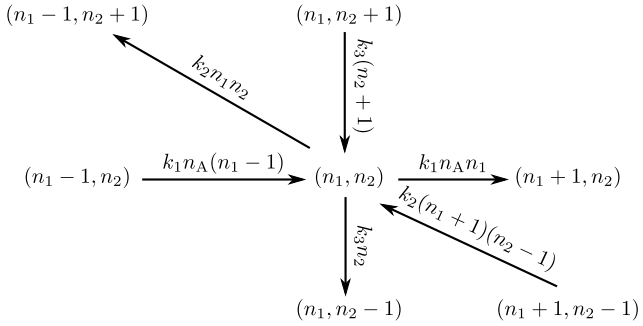


Figure 3.25: State transitions of the Lotka-Volterra model (3.7).

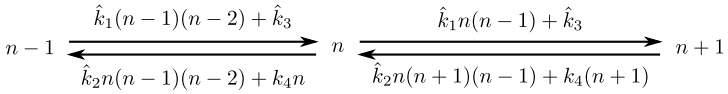


Figure 3.26: State transitions of the Schlögl model (3.9).

Schlögl model revisited: For the Schlögl model (3.9) with propensities in (3.40), the state transition diagram is given in Figure 3.26. But recall that the same transition diagram also corresponds to the reduced reaction network (3.41). Following these state transitions, the CME for both the original Schlögl reaction (3.40) and the reduced reaction network (3.41) reads the same:

$$\frac{\partial}{\partial t}P(n,t) = \Big[\hat{k}_1(n-2)(n-1)+\hat{k}_3\Big]P(n-1,t) - \Big[\hat{k}_1(n-1)n+\hat{k}_3\Big]P(n,t)$$
$$+\hat{k}_2\Big[(n+1)P(n+1,t)-(n-2)P(n,t)\Big](n-1)n + k_4\Big[(n+1)P(n+1,t)-nP(n,t)\Big].$$

While the stochastic simulation algorithm and extensions provide a way to generate sample paths of copy numbers for a biochemical system, the need for repeating many simulation runs to get an idea of the probability distribution in terms of its moments (mean and (co)variance) become increasing time consuming and even impractical for larger systems. Therefore attempts have been made towards continuous approximations of the CME [Gil96, HJ04**?** , MK06].

Gene regulation revisited: For the gene regulation scheme 3.10 with propensities in (3.42), the state transition diagram is given in Figure 3.27. Following these state trans-

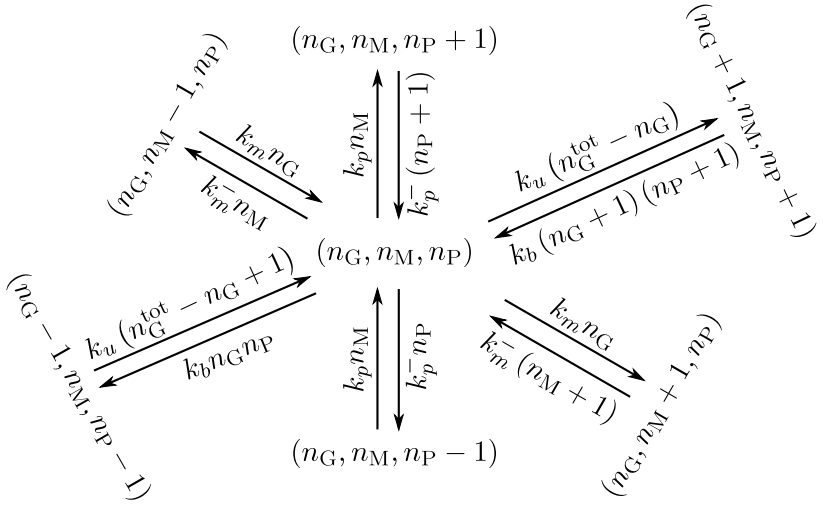


Figure 3.27: State transitions of the gene regulation 3.10.

itions, the CME can be written as:

$$\begin{aligned}
\frac{\partial}{\partial t} P(n_G, n_M, n_P, t) = {} & k_m n_G \left[P(n_G, n_M - 1, n_P, t) - P(n_G, n_M, n_P, t) \right] \\
& + k_m^- \left[(n_M + 1) P(n_G, n_M + 1, n_P, t) - n_M P(n_G, n_M, n_P, t) \right] \\
& + k_p n_M \left[P(n_G, n_M, n_P - 1, t) - P(n_G, n_M, n_P, t) \right] \\
& + k_p^- \left[(n_P + 1) P(n_G, n_M, n_P + 1, t) - n_P P(n_G, n_M, n_P, t) \right] \\
& + k_u \left[\left(n_G^{tot} - n_G + 1 \right) P(n_G - 1, n_M, n_P - 1, t) - \left(n_G^{tot} - n_G \right) P(n_G, n_M, n_P, t) \right] \\
& + k_b \left[(n_G + 1)(n_P + 1) P(n_G + 1, n_M, n_P + 1, t) - n_G n_P P(n_G, n_M, n_P, t) \right]
\end{aligned}$$

4 Modelling of Cell Communication Networks

The life of multicellular animals begins with the fertilisation of an oocyte (egg cell) by a sperm cell. Thereafter the fertilised egg undergoes a series of cell divisions. In the early stages of development, individual cells in the embryo are *totipotent*, i.e., each cell retains the capacity to differentiate into any one of the many different cell types in the body. As development proceeds cells become *pluripotent*, i.e., they become more restricted in their capacity to generate different types of descendent cells.

cell differentiation

The processes of *cell differentiation* lead to individual cells acquiring specialised structures and functions. Some mature and terminally differentiated cells do not undergo cell division, while others (e.g. osteoblasts, chondroblasts, myoblasts,...) divide actively and thereby act as precursors of terminally differentiated cells. Those precursor cells that are also capable of self-renewal are known as *stem cells* (e.g. pluripotent hematopoietic stem cells in the bone marrow). The process of differentiation is closely related to *morphogenesis*, the process by which the structure of the cell is modified through regulated growth.

stem cells

morphogenesis

While genes clearly have a role in these fundamental processes, by which cells grow, divide and differentiate, this role is primarily to provide information for the molecules whose dynamic interactions determine the *structure* and *function* of cells. The *cell cycle* is a sequence of events take the cell from division to division (*mitosis*). Progression through the cell cycle determines *proliferation* (the increase of the number of cells in a population). For example, the essence of cancer is that cells no longer act and react in a regulated fashion within the context of the organ that defines their environment.

cell cycle

mitosis

proliferation

The concept by which interactions of proteins in cell functions are organised are pathways. A *pathway map* exhibits the names of the molecular components, whose interactions govern the basic cell functions. This leads us to a definition of pathways as biochemical networks. One motivation for systems biology is to bring these static diagrams to life by modelling and simulating the biochemical reactions that underlie cell function, development, and disease. To combine into networks that realise higher levels of organisation, such as tissue and organs, cells must communicate. The physical interface between the inside and outside of a cell is comprised, amongst other things, of receptors, which can sense extracellular signals and transduce a signal to the genome where it can effect the transcription of genetic information. The biochemical reactions that relay signals are organised as signal transduction pathways in which regulatory feedback loops play a central role. Many cancer and neurodegenerative diseases are considered a failure of communication at molecular level.

pathway map

This chapter is to consider mathematical modelling and simulation of pathways, i.e.,

networks

networks of biochemical reactions, focussing on dynamic or transient changes.

We have so far encountered a range of representations including the biologist's graphical pathway map and the biochemist's formal reaction equations that describe the interactions of those components referred to in a pathway map. We hereafter seek a similar approach, devising a graphical representation in form of block diagrams as a representation of mathematical equations. The behavior of a formal system is then in turn visualised through simulation plots, phase planes, and bifurcation diagrams. In many ways both, the experimentalists and theoretician rely on visualisations to help an understanding. This does not come as a surprise if one accepts the philosophical arguments put forward in the first chapters of this text.

Figure 4.1: Visualisation plays an important role in all sciences. The biologist visualises his understanding with a pathway map. Properties of mathematical models are visualised as simulation plots, phase plane and bifurcation diagrams.

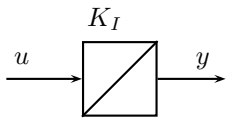
4.1 Pathways as Dynamic Systems

Systems theory and cell biology have enjoyed a long relationship that has received renewed interest in recent years in the context of *systems biology*. The term 'systems' in systems biology comes from *systems theory* or *dynamic systems theory*: Systems biology is defined through the application of systems- and signal-oriented approaches for an understanding of inter- and intra-cellular dynamic processes. The aim of the present text is to review the systems and control perspective of dynamic systems. The biologist's conceptual framework for representing the variables of a biochemical reaction network, and for describing their relationships, are pathway maps. A principal goal of systems biology is to turn these static maps into dynamic models which can provide insight into the temporal evolution of biochemical reaction networks. Towards this end we review the case for differential equation models as a 'natural' representation of causal entailment in pathways. Block-diagrams, commonly used in the engineering sciences, are introduced and compared to pathway maps. The stimulus-response representation of a molecular system is a necessary condition for an understanding of dynamic interactions among the components that make up a pathway. Using simple examples, we show how biochemical reactions are modelled in the dynamic systems framework and visualised using block-diagrams.

Pathway maps used are for most cases a graphical representation that lacks a standard and for which it is not clear which mathematical model should/could be used to simulate

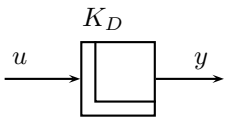
the system. We here introduce a *block diagram* representation of nonlinear dynamic systems, which is an unambiguous translation of the mathematical model. Admittedly it is therefore only suitable for differential equations. The biologist's conception of a pathway map is similar to block diagrams that are widely used in the physical- and engineering sciences. Arbitrary complex systems can be built up from four basic building blocks:

Integrator:



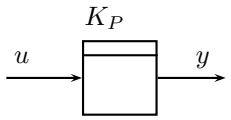
$$y(t) = K_I \int_0^t u(t)dt$$

Differentiator:



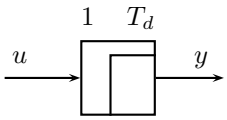
$$y(t) = K_D \frac{d}{dt} u(t)$$

Gain:



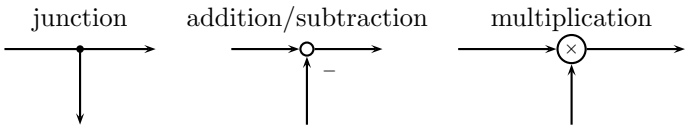
$$y(t) = K_P u(t)$$

Transport Delay:



$$y(t) = u(t - T_d)$$

The most important block we are going to focus on is that of an *integrator*, which describes an accumulation or growth process. The differentiator is simply the reverse operation to the integrator. As alluded to above, the *transport delay* block is of particular importance in simulating the effect of protein translocation, nucleocytoplasmic export and related spatial effects. Block diagrams differ to pathway maps in that they show the processing of *signals*. Block-diagrams are thus a *signal-oriented* approach, an arrow in these diagrams is associated with a variable that is changing over time. The arrows are thus not simply defining 'associations', plus/minus signs indicating amplification/inhibition but instead they are numbers that are added or subtracted. Towards this end, blocks or subsystems are connected through signals via the following nodes:



For the addition/subtraction node, if there is no sign, a "+" is assumed. These basic building blocks form a de facto standard for graphical modelling of control systems circuits. While the value and use of diagrammatic representations of pathway models and tools to visualise them are discussed, for example[1], in [Kit02, A+04, Laz02], there are no established standards for pathway maps. Given that we are discussing the value of control concepts in pathway modelling, we hereafter consider a couple of well studied examples of biochemical systems and investigate (a) how control block diagram representations might be used and (b) how a control analyst might incorporate feedback loops in pathway models. A discussion of how the more conventional pathway maps can serve as information organisers and simulation guides is discussed in [Koh01].

[1] See also http://discover.nci.nih.gov/kohnk/interaction_maps.html

4.2 The Role of Feedback

Differential equations models are particulary suited to study the role of feedback loops. One of the first biologists who recognised the importance of biological feedback was René Thomas [TD90]. For any process that is to maintain, optimise or adapt a condition or value, information about the 'is-state' has to be fed back into the decision on any change that should occur. In other words, *feedback loops* are the basis for any form of regulation and/or control.

Control engineers distinguish between two principal kinds of control systems with different purposes: a) *reference tracking*, and b) *disturbance rejection*. We hereafter refer to the first case, where the system is *sensitive* to inputs, as the ability to make changes as required, e.g., to track or follow a reference signal, as *control*. On the other hand, we refer to *regulation* as the maintenance of a regular or desirable state, making the system *robust* against perturbations. Regulation that maintains the level of a variable is also referred to as *homoeostasis*. Here we should distinguish two forms of robustness in a control system. The first is robustness against external disturbances (disturbance regulation). In a biochemical pathway, a disturbance might be caused by unwanted cross-talk from a neighboring signalling pathway. The second form of robustness, is one which tolerates parameter changes in a system, without significantly changing the system performance. Both forms of robustness are important properties in understanding pathways.

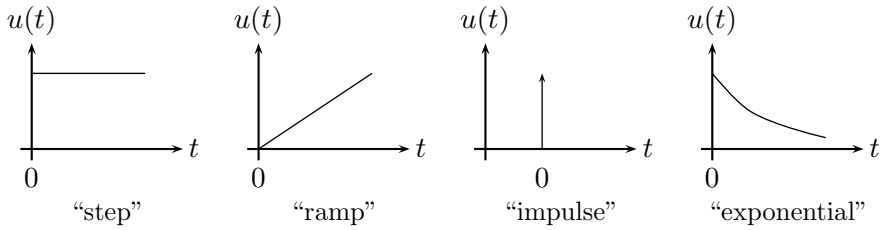


Figure 4.2: Test signals that can be used to investigate the dynamic behavior of pathways.

A central objective of systems biology is to devise methods that allow the detection and description of feedback loops in pathways [K+02, SK04]. An important result from systems theory is that this is only possible through perturbation studies, where the system is stimulated with a well defined signal. Unfortunately, experiments in molecular and cell biology are difficult to set up in a way that suits systems-theoretic approaches. A major hurdle for the success of systems biology arises therefore from the need to conduct expensive, time consuming, complex perturbation experiments.

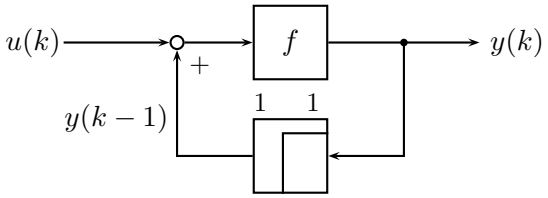
A superficial view of feedback would say that positive feedback is bad (destabilising) and negative feedback is good (stabilising). Indeed, the description of the role of feedback often implies that in the absence of negative feedback, a system is unbounded, unstable and not resistant to perturbations. In fact this is not the case, most dynamical systems exist in a stable manner without the need for feedback. A better way in which to describe the role of feedback is as a modifier of the dynamical behavior of a system. Depending upon the nature of the feedback, it can either stabilise, destabilise, sensitise or desensitise the behavior of a process. While positive feedback is conventionally associated

with destabilisation the truth is more complex, and in many circumstances negative feedback can have unwelcome effects. However, in the context of the special dynamical model forms found in pathway modelling, there are certain special dynamical features induced by feedback that are important to understand. The following simple models of accumulation or growth processes will illustrate some of these features as they manifest themselves within cells.

As an initial demonstration of the features associated with feedback, consider the simple model of growth (e.g. of a cell or of a population of molecules in the cell). Let $u(k)$ denote the stimulus of the system at time k and y the response. Let us take the view that the present depends not only on the current state but also on the past, leading to a discrete version of a differential equation, called *difference equation*:

$$y(k) = f\big(y(k-1), u(k)\big) \ . \tag{4.1}$$

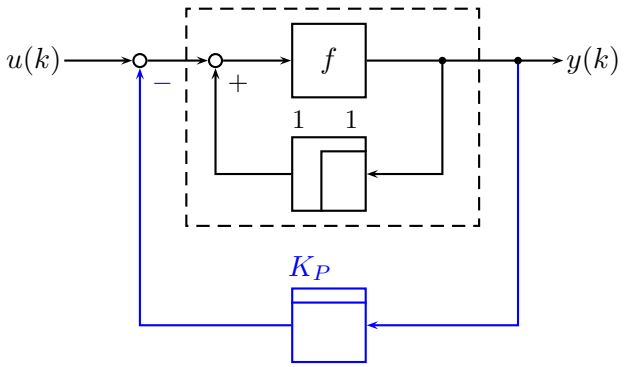
where f describes the detailed functional relationship between the stimulus, the past of y and the current response $y(k)$. One way to illustrate this is by the following block-diagram:

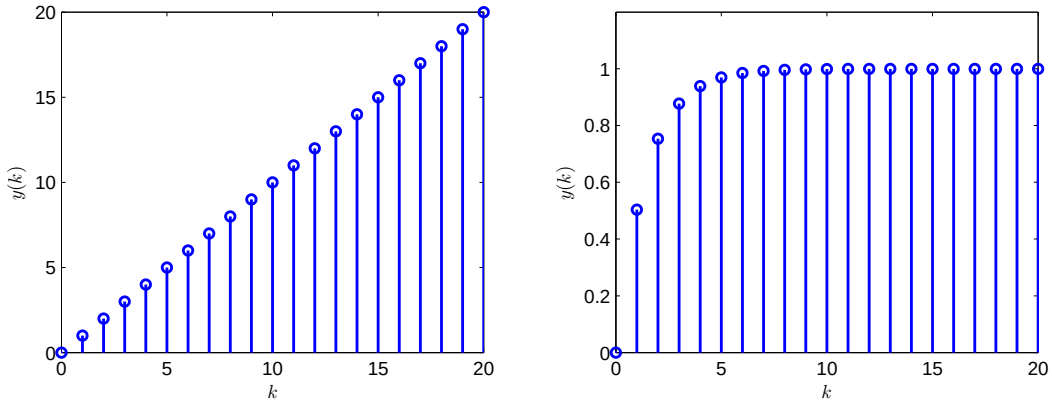


In the diagram the two numbers above the transport delay block denote an amplification of the signal, respectively the unit sampling time delay. For instance, let us look at a linear system, where f is realised by the following law

$$y(k) = u(k) + y(k-1) \ .$$

For initial conditions $y_0 = 0$, $u_0 = 0$ if we stimulate the system with a step input, $u(k) = 1$ for $k \geq 1$, a simulation reveals a linearly increasing, unbounded signal (Figure 4.3(a)). Whatever the initial conditions, $y_0 \geq 0$, the system is unstable and an unrealistic model for most purposes. Let us therefore see what happens if we add a *negative* feedback loop to the system:





(a) (b)

Figure 4.3: Illustration of the stabilising effect of a negative feedback loop in a discrete-time system. Left: Unstable system with a positive feedback loop. Right: Negative feedback loop with $K_P = 1$. In both cases the response to unit step input signal is shown.

The temporal evolution of the response signal is modelled by the following equation

$$y(k) = \big(u(k) - K_P \cdot y(k)\big) + y(k-1) \ .$$

A simulation reveals a bounded signal (Figure 4.3(b)).

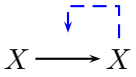
4.3 Tutorial Examples

In the following we present very simple examples of biochemical reactions, which are subsequently translated into a set of mathematical (differential) equations. These in turn maybe related to a standard positive/negative feedback representation drawn from control engineering. In general, we say a component or variable of a system is subject to *negative feedback* when it inhibits its own level of activity. For example, a gene product that acts as a repressor for its own gene is applying negative feedback. Likewise, a component of a system is subject to *positive feedback* when it increases its own level of activity. Through these examples we are going to review the concepts of the biochemist's reaction equation, pathway maps, differential equations and block diagrams.

Returning to our proteolysis example from the introductory section, we generalise it in the context of the framework outlined above. Consider a simple monomolecular reaction where chemical species X is transformed. The change in concentration of X at time t depends on the concentration of X at time t in that the rate by which the reaction proceeds is proportional to the concentration at each time instant,

$$\frac{dx(t)}{dt} \propto x(t)$$

with a certain positive *rate constant* k. A diagrammatic representation of this biochemical process illustrates the fact that chemical species X *"feeds back"* on itself:



A linear mathematical ODE model of the process is given by

$$\frac{d}{dt}x(t) = k \cdot x(t) \ .$$

Here X acts as a *substrate* being converted and the *product*. There is *positive feedback* in that the larger the product X, the greater the rate of change by which substrate X is transformed. A simulation of this system reveals the expected unbounded growth in the concentration of X,

$$x(t) = x_0 \cdot e^{kt} \ ,$$

where $x_0 = x(t = 0)$ denotes the initial condition. With increasing x, the growth rate dx/dt also increases in this system, leading to an unbounded growth. Next we look at the *autocatalytic* reaction

$$X + A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} 2X$$

where for a given X molecule, A facilitates the doubling. A pathway map of this process would be

In pathway maps we use a bar at the end of the arrow to denote an inhibition or negative feedback loop. If A is considered to have a constant concentration, generalising the law of mass action, we arrive at the following differential equation model:

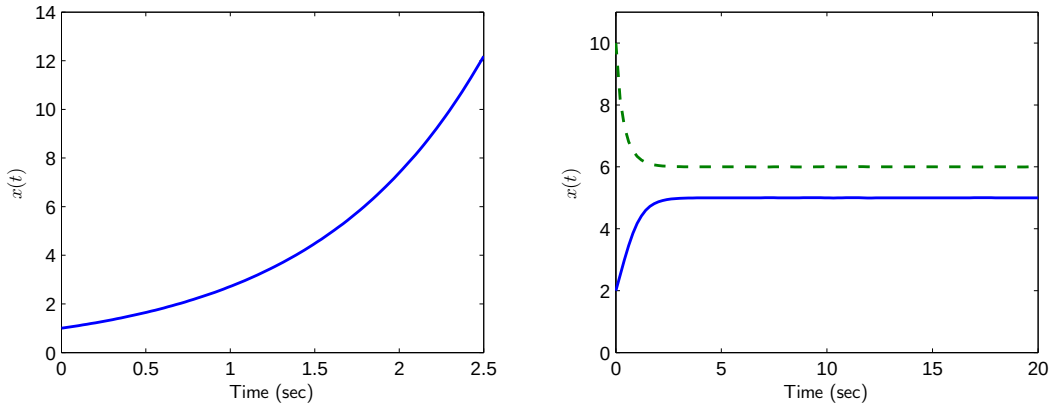
$$\begin{aligned}\frac{d}{dt}x(t) &= k_1 a x(t) - k_2 x^2(t) \\ &= k_1 a x(t) \left(1 - \frac{k_2}{a k_1} x(t)\right) \ .\end{aligned}$$

Why we rewrote the equation in the form given in the second line will be clarified below. In this autocatalytic reaction the 'product' has a strong inhibitory effect on the rate at which X is transformed. In order to indicate the internal feedback mechanisms at work in this system, we will label the right-hand bracketed term $(1 - k_2 x(t)/(a k_1))$ as a control input variable $u(t)$

$$\frac{d}{dt}x(t) = k_1 a u(t) x(t) \ .$$

Because of the product term on the right-hand side this equation is also referred to as a model of a *bilinear system*. If we consider variable x to represent the state of the system, and we write $dx(t)/dt = \dot{x}$ for short, this system becomes

$$\begin{aligned}&\dot{x} = f(x) + g(x)u \ , \qquad x(t_0) = x_0 \ , \\ &y = h(x) \ .\end{aligned}$$



(a) (b)

Figure 4.4: Unbounded and limited growth. Left: Simulation of the monomolecular reaction with positive feedback. Right: Simulation of an autocatalytic reaction (logistic equation) with negative feedback. For the solid line $x_0 = 2$, $a = 2$, $b = 1/2.5$ and for the dashed line $x_0 = 10$, $a = 2$, $b = 1/3$.

We can alternatively write:

$$u(x) = \alpha - \beta x \ ,$$

where the constant α is called the *intrinsic growth rate* of the population and α/β corresponds to the maximum attainable population. The model we thus obtain is specified by the equation

$$\begin{aligned} \frac{dx}{dt} &= \alpha x \left(\frac{\alpha/\beta - x}{\alpha/\beta} \right) \\ &= \alpha x(t) \left(1 - \frac{\beta}{\alpha} x(t) \right) \ . \end{aligned} \tag{4.2}$$

This model form is called the *logistic growth model* and is equivalent to the autocatalytic reaction introduced above. The model describes the real growth rate as a proportion of the intrinsic growth rate. This proportion however decreases with an increase in the population, leading to a more realistic scenario of a system that remains within bounds (Figure 4.4). Both previous examples, echo the observations made in the discrete-time example of a simple growth process with added negative feedback.

For two molecular species we can generalise the control of the system into

$$\begin{aligned} \dot{x}_1 &= u_1(x_1, x_2)\, x_1 \ , \\ \dot{x}_2 &= u_2(x_1, x_2)\, x_2 \ . \end{aligned}$$

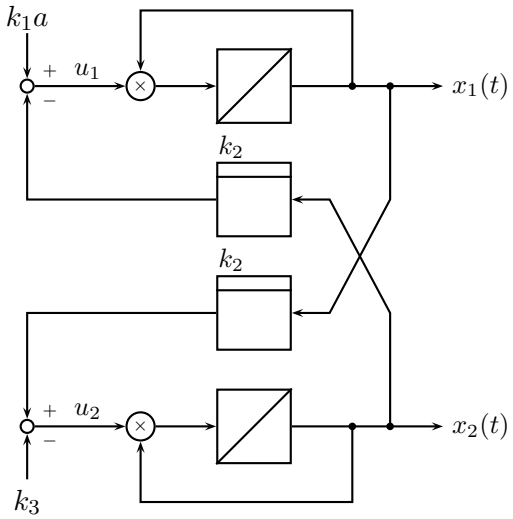
If we specify for u_1 and u_2,

$$\begin{aligned} u_1(x_1, x_2) &= k_1 a - k_2 x_2 \ , \\ u_2(x_1, x_2) &= k_2 x_1 - k_3 \ , \end{aligned}$$

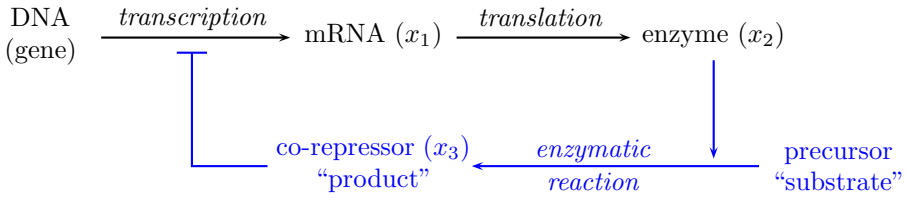
we obtain the well known *Lotka-Volterra model* of two competing populations. If variables x_1 and x_2 correspond to the chemical species X_1 and X_2, the biochemical representation of this system is

$$\begin{aligned} X_1 + A &\xrightarrow{k_1} 2X_1 \\ X_1 + X_2 &\xrightarrow{k_2} 2X_2 \\ X_2 &\xrightarrow{k_3} B \end{aligned}$$

where A is maintained at a constant concentration and B corresponds to the degradation of X_2. The first two reactions are autocatalytic. Compared to the limited growth model from above, this system is capable of showing oscillatory behavior. The block diagram for the Lotka-Volterra model can be drawn directly from those equations:



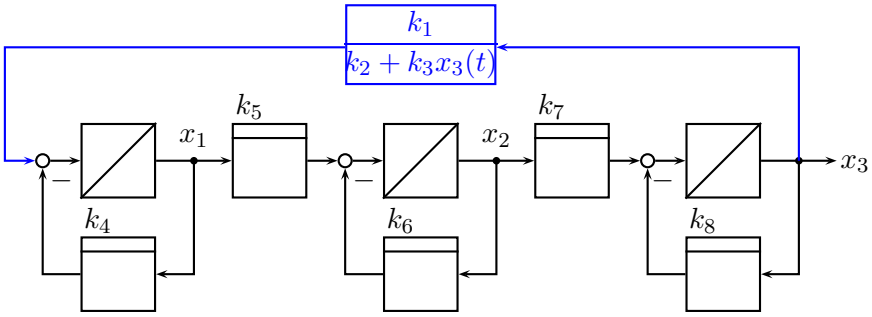
The Lotka-Volterra model of competing species gives an opportunity to discuss the purpose of mathematical models as a mechanism for illuminating basic principles, while not necessarily describing the details of a particular case. Specifically, the Lotka-Volterra model would nowadays be considered an *unrealistic* model for modelling animal population dynamics. However as an abstraction it has proven very *useful*, helping scientists to establish a conceptual approach and ask the right questions [Mur02]. It is in this spirit that models of intracellular dynamics are, or should be, developed in systems biology. The systems considered here are frequently used for an introduction to differential equations. The prototypical biological example of a regulatory system is the protein synthesis model of Jacob and Monod [JM61]. The conceptual model explains how the production of mRNA (x_1), is feedback controlled by a repressor (x_3). A simplified pathway map of this process is shown in the following diagram:



A differential equation model of this regulatory mechanism of protein synthesis is:

$$\begin{aligned}\frac{d}{dt}x_1 &= \frac{k_1}{k_2 + k_3 x_3(t)} - k_4 x_1(t)\\ \frac{d}{dt}x_2 &= k_5 x_1(t) - k_6 x_2(t)\\ \frac{d}{dt}x_3 &= k_7 x_2(t) - k_8 x_3(t)\ .\end{aligned}$$

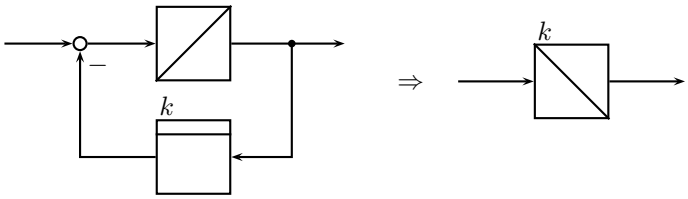
For each of these equations, the last term describes degradation of the molecules. k_5 is the rate of synthesis for the protein that facilitates the production of the co-repressor. Note that there is no minus sign to indicate the negative feedback as in previous examples. The greater x_3 in the numerator of the first term of the rate equation for x_1, the smaller its contribution towards the rate of change of x_1. In contrast to the previous example where the feedback was linear, i.e., a simple additive or negative term, in this example the feedback is nonlinear. To illustrate the use of block-diagrams more clearly, let us consider the block-diagram for the Jacob-Monod model of protein synthesis.



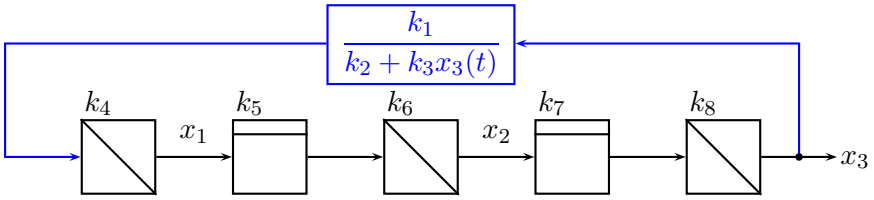
We are now alerted to the fact that *negative feedback* does not necessarily coincide with an explicit form of *negative feedback loop*. Specifically, we have in the block diagram arbitrarily chosen to arrange the figure such that $x_3(t)$ appears as the term fed-back to $x_1(t)$ and that because of the nonlinear form of the feedback it will in fact for small perturbations be negative. The arbitrary nature of the feedback variable is because there is no explicit control input. In such autonomous systems, it is the physical/biological structure that will determine what we (the analyst) chose to call the feedback signal. When the differential equation for x_1 is linearised by Taylor series expansion the $x_3(t)$ appears as a negative feedback term. Whether or not linearisation is feasible depends on the system considered. A more comprehensive discussion of this system and further references can be found in [Mur02].

In the block diagram above we have also noticed that degradation is represented by an integrator with a negative feedback loop around it. This motif we can summarise

into a single block:



Note that this is not just an arbitrary graphical simplification, the inner structure of the block remains unambiguously defined. That is, we do not lose information or accuracy in presentation by scaling the block diagram in this way. Finally, the protein synthesis model can be simplified to



Although the systems we have considered here are fairly simplistic, the consequences of feedback loops we have observed remain akin for more complex processes. For a related instructive discussion of the dynamic systems approach in biochemical pathway modelling we recommend [TO78, FMWT02, TCN03].

4.4 Discussion

Although a pathway or pathway map describes molecules, their physical state and interactions, it is an *abstraction*, with no physical embodiment. A pathway map is thus a model; which proteins and what physical states of the molecules should be considered for experiments and the model is what we call the *art* of modelling.

Feedback loops are the essence of control and regulation, for only if information about the consequences of some output is fed back, the system can adjust itself or respond in an appropriate way. Using ordinary differential equations to model biochemical networks, we have shown that feedback loops can stabilise and destabilise a system, keep its variables and signals bounded, they can make the system robust against perturbations, they allow the system to adapt to changes, or track an input stimulus.

Another relevant feature of control systems is that they have specific intent, and control systems analysts have theories for understanding intent and methods for achieving a required intent or purpose [YHS00]. In a modelling framework, the causal structure of a control system provides a framework for the dynamical manipulation of information with a purposeful objective. This is topical and relevant in the light of recent discussion of the value of systems biology compared with mathematical biology [Lan04]. In this same spirit, feedback loops lie at the heart of the causal/purposeful mechanisms of control and regulation in dynamic systems. Specifically, it is only if information about

the consequences or some output is fed back, can the system automatically adjust itself or respond in an appropriate way. Feedback is not always beneficial, for feedback loops can stabilise or destabilise a system. Feedback can keep a system's variables and signals bounded, or it can induce oscillations or unbounded growth. Likewise, feedback loops can make a system robust against perturbations, but at the same time they allow the system to adapt to changes, or track an input stimulus.

Apart from the role of feedback loops, we surveyed alternative and complementary representations and visualisations, including the biochemist's reaction equations, the mathematician's differential equation models, the control engineer's block diagrams and the biologist's pathway maps. Block diagrams are well established in the engineering sciences as a means of describing dynamic systems in general. Through the integrators used, these diagrams are inherently linked to differential equation models and are therefore less general than those molecular interaction maps [A+04], commonly used to visualise relationships in pathways. On the other hand, block diagrams are a direct and unambiguous visualisation of the mathematical model. These diagrams also do not explicitly represent spatial aspects. While the transport of a protein from the nucleus to the cytosol can be modelled, compartments are realised by introducing more than one variable in the model for the same molecular species in different regions of the cell. For the analysis of the nonlinear differential equations models we only used time plots. Visualisation is no less important to theoreticians than it is to biologists and so there are a range of tools available we have not mentioned here, including stimulus-response curves, phase-plane and bifurcation analysis (e.g. [GH83, KG95, Mur02, TCN03]). For an application of these mathematical tools applied to a model of the yeast cell cycle, we refer to the expositions of Novak and Tyson [Tys01, TN01, TCN03]. The building block approach to an understanding of systems, when associated with purpose, is very similar to the causality principles that are embedded in the dynamical system modelling methods of control engineering. One question we investigated here was whether the control engineer's proficiency with block diagram models and modular representations can contribute to systems biology by facilitating the translation of biological concepts into mathematical representations.

The cell is made up of molecules, like a car is made up from plastic and metal. But a soup of molecules is no more a cell than a heap of plastic and metal is a car. To understand the *functioning* and *function* of a cell we need to know the *relations* and *interactions* of the components that constitute it. If the central dogma of systems biology is that *it is dynamics that determines biological function*, we would refine this statement and argue that *the dynamical manifestation of feedback determines the development and maintenance of biological processes.*

4.5 Phase-Plane Analysis

Phase-plane analysis is an important technique in studying the behavior of linear and nonlinear dynamic systems. It is a graphical approach which allows the study of the behavior of the system for a large range of initial conditions. It is for this reason that

this method is referred to as providing a *qualitative analysis* of the dynamic system. For linear systems the approach is truly global, while for nonlinear systems it is only locally applicable. The main purpose of this section is to illustrate the diversity of behavior nonlinear dynamics can generate and to introduce a tool for its analysis in the plane. For a more comprehensive description of phase-plane analysis see one of the many books available on differential equations (e.g. [BD01, EP01]).

The general form of systems considered for phase-plane analysis is

$$\frac{dx}{dt} = f(x, y) \ , \qquad \frac{dy}{dt} = g(x, y) \ . \tag{4.3}$$

A system in which t does not explicitly occur in f and g is called an *autonomous system*. The two differential equations determine the velocity of two variables x and y moving in the xy-plane referred to as the *phase-plane*. As time increases, the system state moves along a curve in the xy-plane, called the trajectory. While for non-autonomous systems trajectories could cross in the plane, for autonomous systems the trajectories cannot cross. The totality of all trajectories describes the *phase portrait*. Points (x, y) of the trajectory for which

phase portrait

$$f(x, y) = g(x, y) = 0$$

are called *critical points* or *fixed points*, often also referred to as steady-states or equilibrium points[2]. To see the appearance the phase portrait, we consider a *direction field* on a grid of points in the $\mathbb{R}^2$ plane and determine velocity vectors defined by

critical point

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}} = \frac{g(x, y)}{f(x, y)} \ .$$

As a first simple example let us consider the system

$$\frac{dx}{dt} = y \qquad \text{and} \qquad \frac{dy}{dt} = 4x \ .$$

The only fixed point for this system is the origin, $(0, 0)$, of the plane. We can solve the system by separation of variables:

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}} = \frac{4x}{y}$$

which implies

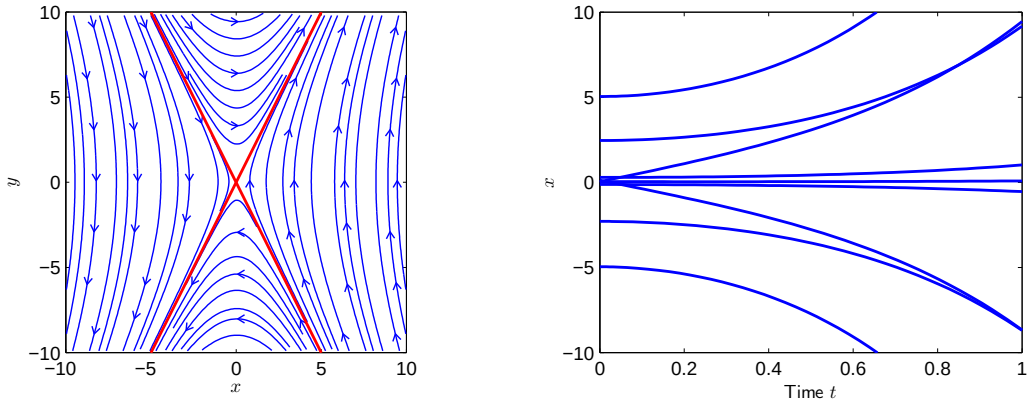
$$\int y \ dy = \int 4x \ dx \qquad \text{i.e.,} \qquad \frac{y^2}{2} = 2x^2 + c$$

leading to trajectories that are hyperbolas:

$$4x^2 - y^2 = c \ .$$

The phase-plane and some sample solutions are shown in Figure 4.5.

[2] In the engineering literature fixed points are also referred to as *equilibrium points*. In the context of biochemical networks in cells this can however lead to confusion.



(a) (b)

Figure 4.5: Left: Phase portrait of the system $dx/dt = y$, $dy/dt = 4x$. The fixed point $(0, 0)$ is said to be *unstable* because trajectories close to the fixed point move away from it. From the appearance of graph, the fixed point is also referred to as a *saddle point*. Right: Sample solutions for $x(t)$ for different initial conditions.

A second motivating example, leading to trajectories that are circles is

$$\frac{dx}{dt} = y \qquad \text{and} \qquad \frac{dy}{dt} = -x \ ,$$

The only fixed point is again $(0, 0)$. By separation of variables

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}} = \frac{-x}{y}$$

which implies

$$\int y \, dy = -\int x \, dx \qquad \text{i.e.,} \qquad \frac{y^2}{2} = -\frac{x^2}{2} + c$$

leading to circular trajectories:

$$x^2 + y^2 = c \ .$$

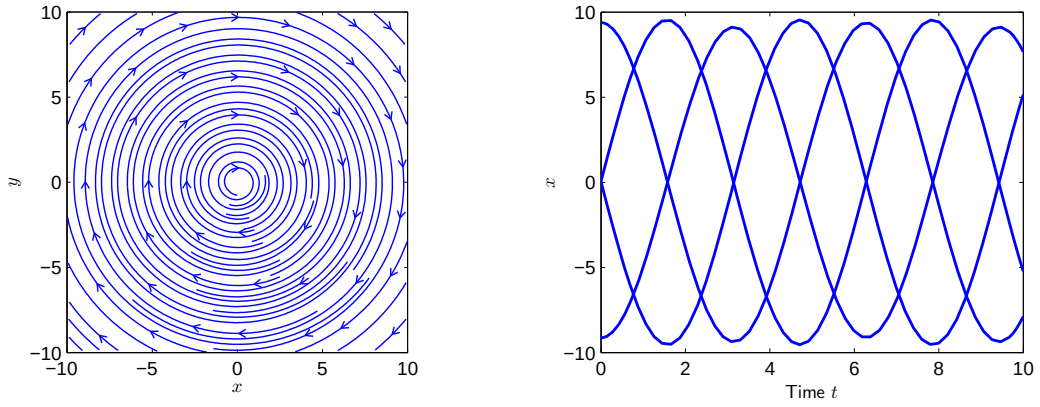
The phase-portraits is shown in Figure 4.6.

From these two examples we now consider a more comprehensive survey of linear dynamics, followed by nonlinear systems. If the derivatives are linear functions of the variables, we deal with a linear (autonomous) system:

$$\left.\begin{aligned} \frac{dx}{dt} &= a \cdot x + b \cdot y \\ \frac{dy}{dt} &= c \cdot x + d \cdot y \end{aligned}\right\} \tag{4.4}$$

In matrix form we can rewrite this system as

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \ .$$



(a) (b)

Figure 4.6: Phase portrait of the system $dx/dt = y$, $dy/dt = -x$. The fixed point $(0, 0)$ of this system is *stable*, i.e., all trajectories close to it remain close for all t. This fixed point is called a *centerpoint*.

The matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is called the *system matrix* of coefficients. If the determinant of A system matrix

$$\det(A) \equiv |A| = ad - bc \ ,$$

is nonzero, there is a unique solution to the equations. For a linear system, the origin of the phase-plane is this fixed point. If $\det(A) = 0$ there either aren't any solutions or there are infinitely many. In this case have to solve the system of linear algebraic equations

$$\begin{aligned} a \cdot x + b \cdot y &= 0 \\ c \cdot x + d \cdot y &= 0 \ . \end{aligned}$$

The solutions to the linear system differential equations are

$$\begin{bmatrix} x \\ y \end{bmatrix} = c_1 \vec{v}_1 e^{\lambda_1 t} + c_2 \vec{v}_2 e^{\lambda_2 t} \ ,$$

where $\vec{v}_1$ and $\vec{v}_2$ are the *eigenvectors* of matrix A corresponding to *eigenvalues* λ_1 and λ_2 of A. The eigenvectors and eigenvalues are found by asking whether there are exists a nonzero vector $\vec{v}$ such that the result $A\vec{v}$ is a simple scalar multiple of $\vec{v}$. With the eigenvalues on the diagonal of a diagonal matrix Λ and the corresponding eigenvectors $\vec{v}$ forming the columns of a matrix L, we have eigenvalues/vectors

$$AL = \Lambda L$$

If L is nonsingular[3], this becomes the *eigenvalue decomposition* of A:

$$A = L\Lambda L^{-1} \ .$$

The eigenvectors and values have a geometric interpretation. The length $|A\vec{v}|$ of vector $A\vec{v}$ is $\pm\lambda|\vec{v}|$. The multiplication of $\vec{v}$ by the matrix A expands or contracts vector $\vec{v}$, while a positive eigenvalue preserves its direction, a negative value reverses it. In application of dynamic systems in the plane, the eigenvalue corresponds to the speed of response, while the eigenvector determines the principal direction. A line in the phase-plane that is not crossed by any trajectory is called a *separatrix*. Eigenvectors determine the separatrices.

separatrix

The eigenvalues λ_1 and λ_2 of A are determined as solutions of the *characteristic equation*

characteristic equation

$$\det(A - \lambda I) = \det \begin{bmatrix} a-\lambda & b \\ c & d-\lambda \end{bmatrix} = 0$$

where I is the identity matrix with ones on the diagonal and zeros elsewhere. The eigenvalues are thus the roots of the *characteristic polynomial* $\det(A - \lambda I)$

characteristic polynomial

$$(a-\lambda)(d-\lambda) - bc = \lambda^2 - (a+d)\lambda + (ad - bc) = 0 \ .$$

The constant term is equal to $\det(A)$ and the coefficient $(a+d)$ corresponds to the *trace* of A, denoted $\mathrm{tr}(A)$. Let $A = [a_{ij}]$ be an $n \times n$ matrix, the trace of A is defined to be the sum of the diagonal entries $\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}$. The eigenvalues are then given by

$$\begin{aligned} \lambda_{1,2} &= \frac{1}{2}\left(\mathrm{tr}(A) \pm \sqrt{[\mathrm{tr}(A)]^2 - 4\,\det(A)}\right) \\ &= \frac{1}{2}\left((a+d) \pm \sqrt{(a+d)^2 - 4(ad-bc)}\right) \ . \end{aligned}$$

Given the eigenvalues, the eigenvectors can be calculated by

$$\vec{v}_i = \frac{1}{\sqrt{1+p_i^2}} \begin{bmatrix} 1 \\ p_i \end{bmatrix}$$

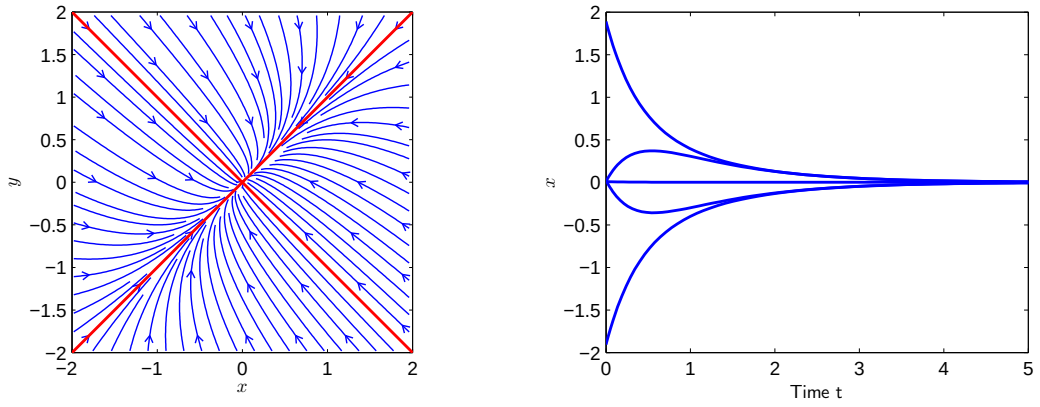
where

$$p_i = \frac{\lambda_i - a}{b}, \quad b \neq 0, \quad i = 1, 2 \ .$$

The sign of $[\mathrm{tr}(A)]^2 - 4\,\det(A)$ determines whether the eigenvalues λ are:

1. complex with nonzero imaginary part if $[\mathrm{tr}(A)]^2 - 4\,\det(A) < 0$
2. real and distinct if $[\mathrm{tr}(A)]^2 - 4\,\det(A) > 0$
3. real and repeated if $[\mathrm{tr}(A)]^2 - 4\,\det(A) = 0$.

[3] A matrix is *singular* if its determinant is zero. It is *regular* if the determinant is nonzero, and in which case an inverse exist.



(a) (b)

Figure 4.7: Left: Phase portrait of the system $dx/dt = -2x+y$, $dy/dt = x-2y$. Right: Sample solutions for x. Eigenvalues $\lambda_1 = -3, \lambda_2 = -1$. The critical point $(0,0)$ is called a nodal sink. It is asymptotically stable node.

If in the first case $[\mathrm{tr}(A)]^2 - 4\ \det(A) < 0$, then the real part of the eigenvalues is $\mathrm{tr}(A/2) \equiv (a+d)/2$, determining a

... *spiral sink* if $\mathrm{tr}(A) < 0$

... *spiral source* if $\mathrm{tr}(A) > 0$

... *center* if $\mathrm{tr}(A) = 0$.

If for the second case above $[(\mathrm{tr}(A)]^2 - 4\ \det(A) > 0$, if $\det(A) < 0$ we have a *saddle* and for $\mathrm{tr}(A) > 0$ and $\det(A) > 0$ we have a *source*. In terms of the eigenvalues of matrix A we can distinguish five cases, discussed hereafter.

Case 1: Unequal, real eigenvalues of the same sign: The general solution of (4.4) is

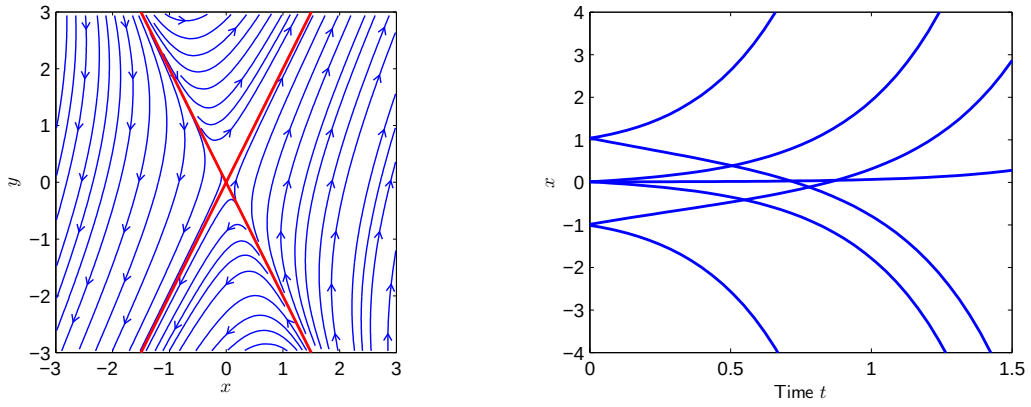
$$[x\ y]^T = c_1\vec{v}_1 e^{\lambda_1 t} + c_2\vec{v}_2 e^{\lambda_2 t}\ .$$

The eigenvalues can be either positive or negative. In Figure 4.7 the case for

$$A = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$$

with $\lambda_1 < \lambda_2 < 0$ is shown. From the general solution we see that both variables approach zero as time goes to infinity, regardless of the constants c_1 and c_2. This means that all solutions approach the critical point at the origin as $t \to \infty$. The eigenvectors are in this case $\vec{v}_1 = [0.71 \quad -0.71]^T$ and $\vec{v}_2 = [0.71 \quad 0.71]^T$, forming a cross through the origin. Notice that all solutions approach the critical point tangent to $\vec{v}_2$, except those solutions that start on the line through $\vec{v}_1$. This critical point is called a *node* or *nodal sink*. If λ_1 and λ_2 are both positive and $0 < \lambda_2 < \lambda_1$, the trajectories in the phase-plane have the same pattern as in Figure 4.7 but they are moving away from the

nodal sink



(a) (b)

Figure 4.8: Left: Phase portrait of the system $dx/dt = x + y$, $dy/dt = 4x + y$. Right: Sample solutions for x. Eigenvalues $\lambda_1 = 3, \lambda_2 = -1$. The critical point $(0, 0)$ is called a saddle point.

critical point. $x(t)$ and $y(t)$ grow exponentially in this case. The critical point is, in this case, called *nodal source*.

nodal source

Case 2: Unequal, real eigenvalues of the opposite sign: The general solution of (4.4) is again

$$[x\ y]^T = c_1\vec{v}_1 e^{\lambda_1 t} + c_2\vec{v}_2 e^{\lambda_2 t}\ .$$

In Figure 4.8 the system with matrix

$$A = \begin{bmatrix} 1 & 1 \\ 4 & 1 \end{bmatrix}$$

is illustrated. The eigenvectors for this system are $\vec{v}_1 = [0.45 \quad 0.89]^T$ and $\vec{v}_2 = [-0.45 \quad 0.89]^T$. The eigenvectors are again forming a cross through the origin. If a solution starts on the line along $\vec{v}_1$ (which goes from the bottom left to top right corner of the plane), it will remain there for all time and $c_2 = 0$. The only solutions that approach the critical point in the origin are those that start on the line determined by $\vec{v}_2$. For all other initial conditions the positive exponential term in the solution will eventually dominate. The origin is called a *saddle point*. The origin is also an unstable fixed point since no solution will remain there.

saddle point

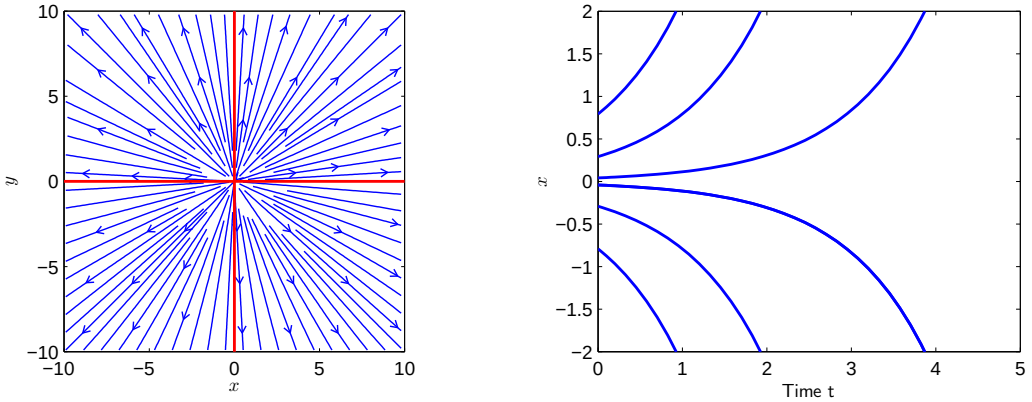
Case 3: Equal eigenvalues: In case $\lambda_1 = \lambda_2 = \lambda$, we have to distinguish two cases.

Two independent eigenvectors: The general solution is

$$[x\ y]^T = c_1\vec{v}_1 e^{\lambda t} + c_2\vec{v}_2 e^{\lambda t}\ .$$

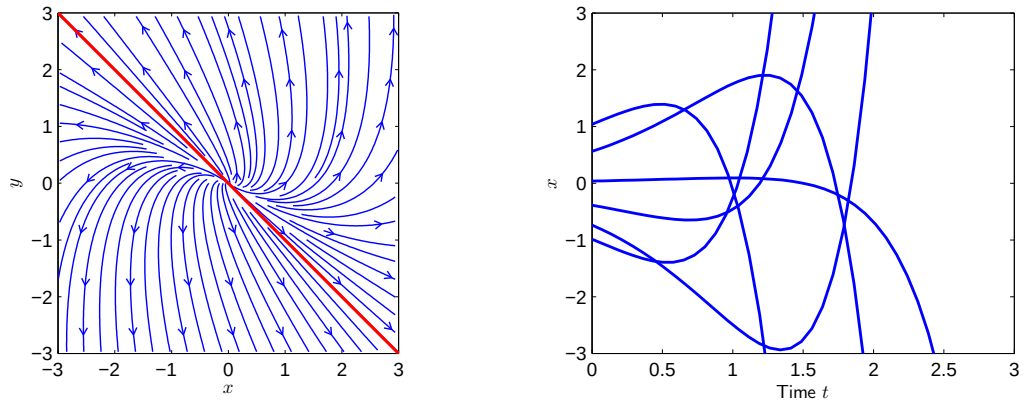
In this case the ratio y/x is only dependent on $\vec{v}_1$, $\vec{v}_2$ and independent of t. The trajectories generate a star-shaped pattern of the phase-plane. The fixed point is called a *proper node* or *star point*. Figure 4.9 illustrates a system with eigenvectors $\vec{v}_1 = [1 \quad 0]^T$ and $\vec{v}_2 = [0 \quad 1]^T$. The node is asymptotically stable or unstable, depending on whether

proper node



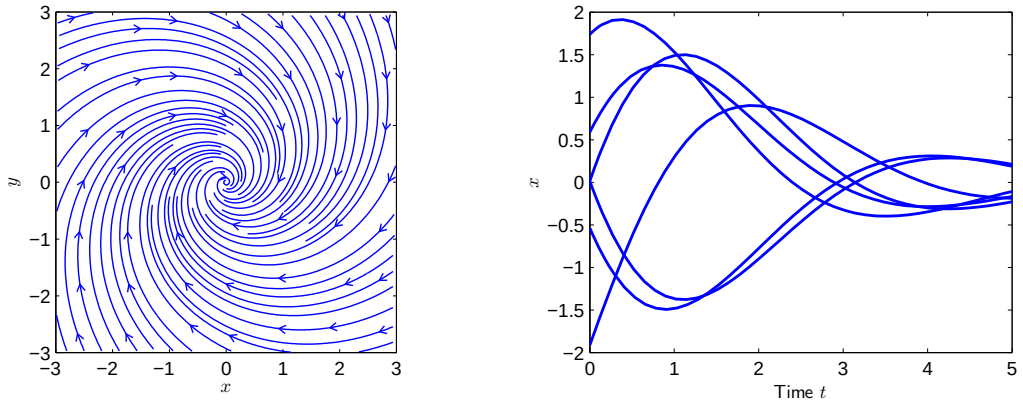
(a) (b)

Figure 4.9: Left: Phase portrait of the system $dx/dt = x$, $dy/dt = y$. Right: Sample solutions for x. Eigenvalue $\lambda = 1$. The critical point is called a proper node.



(a) (b)

Figure 4.10: Left: Phase portrait of the system $dx/dt = x - y$, $dy/dt = x + 3y$. Right: Sample solutions for x. Eigenvalue $\lambda = 2$. There is only one independent eigenvector. The critical point is called an improper node.



(a) (b)

Figure 4.11: Left: Phase portrait of the system $dx/dt = -0.5x + y$, $dy/dt = -x - 0.5y$. Right: Sample solutions for x. Eigenvalues $\lambda_1 = -0.5 + i$, $\lambda_2 = -0.5 - i$. Since the real part is negative the trajectories spiral inwards.

the eigenvalue is negative or positive.

One independent eigenvector: The general solution is in this case

$$[x\ y]^T = c_1\vec{v}_1 e^{\lambda t} + c_2(\vec{v}_1 t e^{\lambda t} + \vec{v}_2 e^{\lambda t})$$

where $\vec{v}_1$ is the one independent eigenvector and $\vec{v}_2$ denotes the generalised eigenvector associated with the repeated eigenvalue. For a large t the dominant term is $c_2\vec{v}_1 t e^{\lambda t}$, which means that for $t \to \infty$ all trajectories approach the origin tangent to the line through the eigenvector. The orientation of the trajectories depends on the relative positions of $\vec{v}_1$ and $\vec{v}_2$. Figure 4.10 illustrates one situation for a system with eigenvectors $\vec{v}_1 = [-0.71 \quad 0.71]^T$ and $\vec{v}_2 = [-0.71 \quad 0.71]^T$. When a repeated eigenvalue has only one independent eigenvector, then the critical point is called an *improper node* or *degenerate node*. An improper node is asymptotically stable or unstable, depending on whether the eigenvalues are negative or positive.

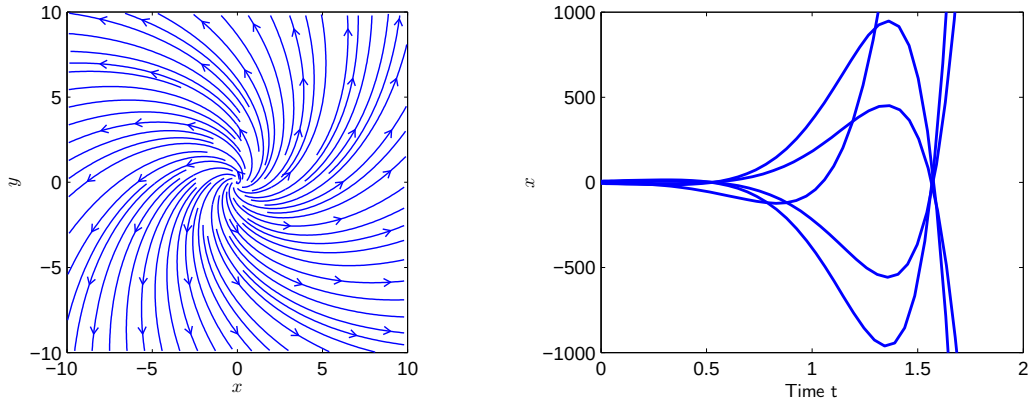
improper node

Case 4: Complex eigenvalues: In this case the eigenvalues are $a \pm ib$, where a is the real part and b denotes the imaginary part. The critical point is called a *spiral point*, *spiral sink* or *spiral source*. Whenever $a \neq 0$, the trajectories are spirals. They are directed inward or outward, depending on whether a is positive or negative. Figures 4.11 and 4.12 provides an illustration.

spiral point

Case 5: Pure imaginary eigenvalues: In case $a = 0$ for the eigenvalues, the trajectories become circles around the origin, that are traversed clockwise if $b > 0$ and anticlockwise if $b < 0$. Figure 4.6 provides an illustration for the system $dx/dt = y$, $dy/dt = -x$, with eigenvalues $\lambda_{1,2} = \pm i$.

We have summarised the dynamic properties or *stability* of the linear system $[x\ y]^T = A[x\ y]^T$ in Table 4.1. Before we continue with study nonlinear dynamics using phase-plane analysis, we look at a particular nonlinear autonomous system which is almost



(a) (b)

Figure 4.12: Left: Phase portrait of the system $dx/dt = 4x - 3y$, $dy/dt = 3x + 4y$. Right: Sample solutions for x. Eigenvalues $\lambda_{1,2} = 4 \pm 3i$. Since the real part is positive the trajectories spiral outwards.

Table 4.1: Stability of the linear system $[\dot{x}\ \dot{y}]^T = A[x\ y]^T$ with $\det(A) \neq 0$.

Eigenvalues	Type of Critical Point	Stability
$\lambda_1 > \lambda_2 > 0$	Node	Unstable
$\lambda_1 < \lambda_2 < 0$	Node	Asymptotically stable
$\lambda_2 < 0 < \lambda_1$	Saddle point	Unstable
$\lambda_1 = \lambda_2 > 0$	Proper or improper node	Unstable
$\lambda_1 = \lambda_2 < 0$	Proper or improper node	Asymptotically stable
$\lambda_{1,2} = \alpha \pm i\beta$	Spiral point	
$\alpha > 0$		Unstable
$\alpha < 0$		Asymptotically stable
$\lambda_{1,2} = \pm i\beta$	Center	Stable

linear around the origin:

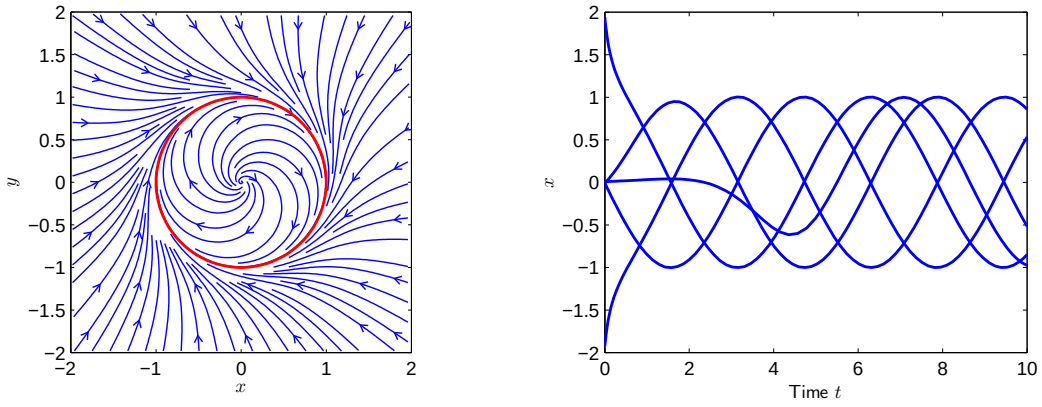
$$\begin{aligned} \frac{dx}{dt} &= y + x - x(x^2 + y^2) \\ \frac{dy}{dt} &= -x + y - y(x^2 + y^2) \ . \end{aligned}$$

The only critical point of this system is the origin $(0, 0)$. The corresponding linear system has the system matrix

$$A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \ .$$

and eigenvalues $\lambda_{1,2} = \pm i$, which suggests the origin is an unstable spiral point for the linear as well as the nonlinear system. However, rather than spiralling out completely, as the linear analysis would suggest, the system exhibits what is known as a *limit cycle*. Figure 4.13 illustrates the limit cycle behavior.

limit cycle



(a) (b)

Figure 4.13: Illustration of an autonomous nonlinear system, which is almost linear around the origin but displays a limit cycle.

4.6 Nonlinear Dynamics

As a motivating example for nonlinear systems let us find the trajectories of the following system with two coupled nonlinear equations:

$$\frac{dx}{dt} = 4 - 2y \ , \quad \frac{dy}{dt} = 12 - 3x^2 \ .$$

To find critical points we set the derivatives to zero:

$$4 - 2y = 0 \ , \quad 12 - 3x^2 = 0$$

and find that there are two critical points at $(-2, 2)$ and $(2, 2)$. For the trajectories we write

$$\frac{dy}{dx} = \frac{12 - 3x^2}{4 - 2y} \ .$$

Separation of variables and integration provides us with the solution

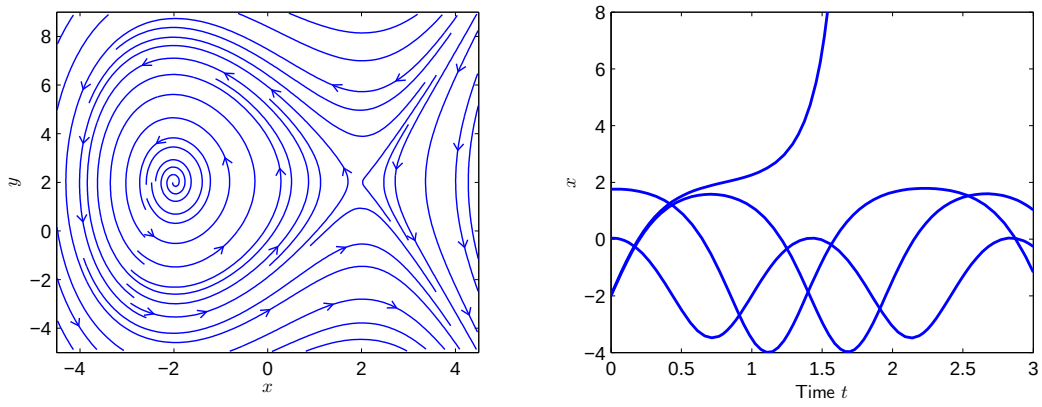
$$4y - y^2 - 12x + x^3 = c$$

where c is some arbitrary constant. Figure 4.14 illustrates the phase portrait.

The phase-plane analysis introduced above does work for nonlinear systems by linearising a system around a point of interest. The analysis in this case applies locally. Points of particular interest are critical- or fixed points. Let us denote such point of particular interest in the plane as (x^*, y^*).

Taylor expansion

Linearising a nonlinear system is done in the neighborhood of the fixed points using a Taylor expansion of $f(x, y)$ and $g(x, y)$. The Taylor expansion for a function of two



(a) (b)

Figure 4.14: Phase portrait of the nonlinear system $dx/dt = 4 - 2y$, $dy/dt = 12 - 3x^2$.

variables $f(x, y)$ is given by

$$f(x,y) = f(x^*,y^*) + \left.\frac{\partial f}{\partial x}\right|_{x^*,y^*}(x-x^*) + \left.\frac{\partial^2 f}{\partial x^2}\right|_{x^*,y^*}\frac{(x-x^*)^2}{2!} + \left.\frac{\partial^3 f}{\partial x^3}\right|_{x^*,y^*}\frac{(x-x^*)^3}{3!} + \left.\frac{\partial f}{\partial y}\right|_{x^*,y^*}(y-y^*) + \left.\frac{\partial^2 f}{\partial y^2}\right|_{x^*,y^*}\frac{(y-y^*)^2}{2!} + \left.\frac{\partial^3 f}{\partial y^3}\right|_{x^*,y^*}\frac{(y-y^*)^3}{3!} + \cdots \tag{4.5}$$

If we neglect terms higher than first order the Taylor expansion is

$$f(x,y) \approx f(x^*,y^*) + \left.\frac{\partial f}{\partial x}\right|_{x^*,y^*}(x-x^*) + \left.\frac{\partial f}{\partial y}\right|_{x^*,y^*}(y-y^*) \ . \tag{4.6}$$

Introducing new variables u, v,

$$u \doteq x - x^* \ , \qquad v \doteq y - y^* \ ,$$

we can write for the expansion of (4.14),

$$\begin{aligned} \frac{du}{dt} &= au + bv + \cdots \ , \\ \frac{dv}{dt} &= cu + dv + \cdots \ , \end{aligned} \tag{4.7}$$

where

$$\begin{aligned} a &= \left.\frac{\partial f}{\partial x}\right|_{x^*,y^*} \ , \qquad & b &= \left.\frac{\partial f}{\partial y}\right|_{x^*,y^*} \ , \\ c &= \left.\frac{\partial g}{\partial x}\right|_{x^*,y^*} \ , \qquad & d &= \left.\frac{\partial g}{\partial y}\right|_{x^*,y^*} \ . \end{aligned} \tag{4.8}$$

The trick is then to assume that in the neighborhood of the fixed points, the higher-order terms in (4.7) are small enough to be neglected. If we collect the partial derivatives in matrix form this leads us to what is called the *Jacobian matrix*:

Jacobian m

$$J^* = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{bmatrix} , \tag{4.9}$$

with the partial derivatives evaluated at (x^*, y^*), treating in each case the other variable as a constant. Therefore in the neighborhood of a fixed point, the nonlinear system (4.14) can be approximated by the set of linear equations

$$\begin{bmatrix} \dot{u} \\ \dot{v} \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} . \tag{4.10}$$

The dynamics in the neighborhood of (x^*, y^*) are now determined by the eigenvalues of the system matrix, as illustrated above.

For more examples, we return to the population models referred to earlier in the text. On page 101 of Chapter 2 we considered a simple population model

$$\frac{dS(t)}{dt} = (k^+ - k^-)S(t) , \tag{2.5}$$

where $S(t)$ denotes the molecular population size at time t, k^+ the formation or birth rate and k^- the decay or death rate. In systems and control theory or the study of differential equations such model is more commonly written using the following notation

$$\frac{dx(t)}{dt} = u(t)x(t) , \tag{4.11}$$

where $u(t)$ may be considered a *control input variable*. Because of the product of the state variable and input variable, $u(t)x(t)$, such a system is also called a *bilinear* system, and which is an example of the more general system

bilinear system

$$\begin{aligned} \dot{x} &= \phi\big(t, x(t)\big) + u(t)g\big(t, x(t)\big) , \qquad x(t_0) = x_0 \\ y(t) &= h\big(t, x(t)\big) . \end{aligned} \tag{4.12}$$

In many realistic situations $u(t)$ may depend on $x(t)$, which in effect leads to some population constraints. A simple example of such a *feedback control loop* is the following definition for u

$$u\big(x(t)\big) = a - bx(t) ,$$

where for a population model, a may be considered as the *intrinsic* growth rate such that a/b corresponds to the maximum population level that can be reached [Ton90]. The input-output description and the description of feedback loops is central to the control engineering approach.

We can extend the idea of the bilinear formulation (4.11) to two molecular species x and y,

$$\frac{dx}{dt} = u_1(x,y)x \ , \qquad \frac{dy}{dt} = u_2(x,y)y \ , \tag{4.13}$$

where the dependence on t is suppressed to simplify the notation. If we define

$$\begin{aligned} u_1(x,y) &= \alpha - \beta y \\ u_2(x,y) &= \gamma x - \delta \ , \end{aligned}$$

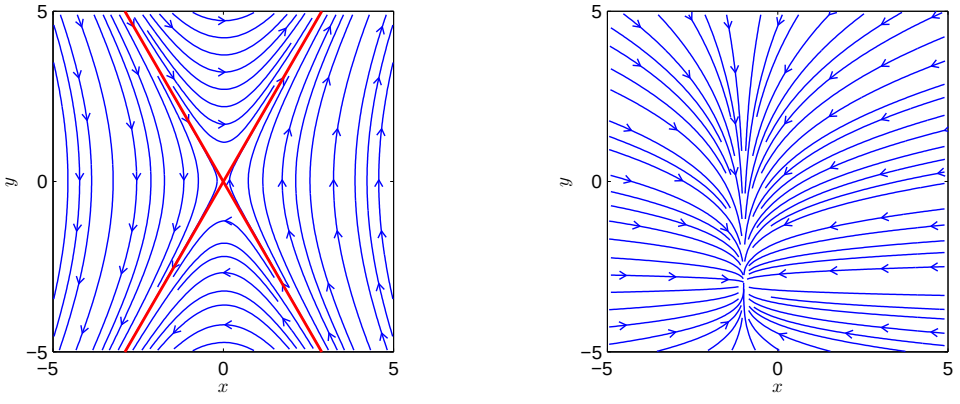
we obtain the well known Lotka-Volterra model. This model has been rather useful, certainly for educational purposes. We encountered this system in Chapter 3 (Equation 3.7, Figures 3.4 and 3.11) and will further discuss it hereafter. But first let us consider the phase-portrait of a bilinear system with $u_1(x,y) = y + y/x$ and $u_2(x,y) = x + 3x/y$:

$$\begin{aligned} \frac{dx}{dt} &= xy + y \\ \frac{dy}{dt} &= xy + 3x \ . \end{aligned}$$

The system has two critical points, a trivial fixed point (x_1^*, y_1^*) at $(0,0)$ and another at $(x_2^*, y_2^*) = (-1,-3)$. The Jacobian matrix is

$$J^* = \begin{bmatrix} y_{1,2}^* & x_{1,2}^* + 1 \\ y_{1,2}^* + 3 & x_{1,2}^* \end{bmatrix} \ .$$

Considering first the trivial fixed point. The eigenvalues are $\lambda_1 = -\sqrt{3}$, $\lambda_2 = \sqrt{3}$, with associated eigenvectors $\vec{v}_1 = [-0.5 \ 0.866]^T$ and $[0.5 \ 0.866]^T$. For the nontrivial fixed point the eigenvalues are $\lambda_1 = -3$, $\lambda_2 = -1$, with associated eigenvectors $\vec{v}_1 = [1 \ 0]^T$



(a) (b)

Figure 4.15: Phase portrait of the nonlinear system $xy + y$, $dy/dt = xy + 3x$, around the critical points $(0,0)$ on the left and $(-1,-3)$ on the right.

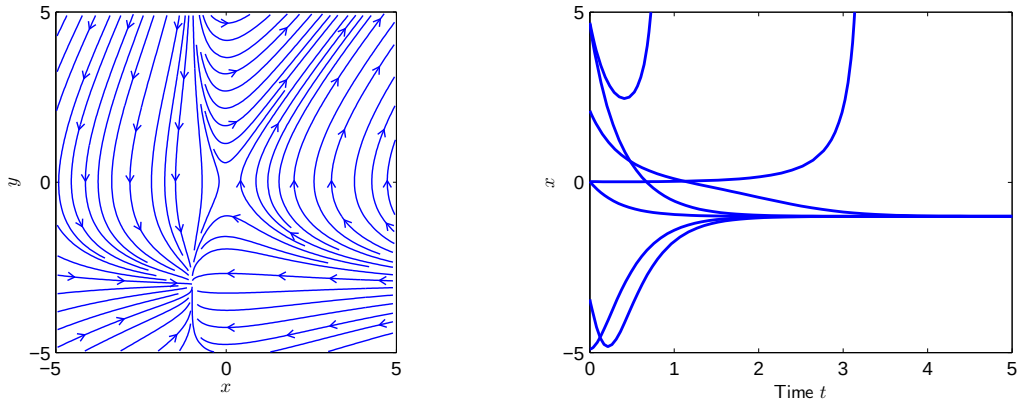


Figure 4.16: Left: Phase portrait of the nonlinear system $dx/dt = xy + y$, $dy/dt = xy + 3x$. Right: sample solutions.

and $[0\ 1]^T$. Figure 4.15 illustrates the two phase-portraits of the system linearised around the critical points and Figure 4.16 shows the combined phase portrait of the nonlinear system.

What the Lotka-Volterra model is for the theoretician is the enzyme kinetic reaction for the experimentalist. The vast majority of biochemical reactions in a cell are catalyzed by enzymes. This type of reaction is therefore of particular interest in modelling and in experiments. We first introduced this system on in Section 2.7 and will have a further discussion of this reaction in Section 4.8.2. The biochemists diagrammatic representation is

$$E + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_3} E + P$$

where substrate S is under the action of enzyme E turned first into an intermediate complex C before further decomposed into a product P and the enzyme. The mass action kinetic equations for changes in the concentrations of the substrate and complex are

$$\begin{aligned} \frac{d[S]}{dt} &= -k_1[S][E] + k_2[C], \\ \frac{d[ES]}{dt} &= -(k_2 + k_3)[C] + k_1[S][E] \ . \end{aligned}$$

Since the enzyme is, in a sense, *controlling* the reaction, we may consider it as an input to the system. Rewriting these equations, using input variable u and state variables x_1 and x_2 for substrate and complex respectively, gives us the following compact representation that emphasises an input-output representation and where state-variables and inputs are bilinearly related:

$$\begin{aligned} \dot{x}_1 &= -k_1 x_1 u + k_2 x_2, \\ \dot{x}_2 &= -(k_2 + k_3)x_2 + k_1 x_1 u \ . \end{aligned}$$

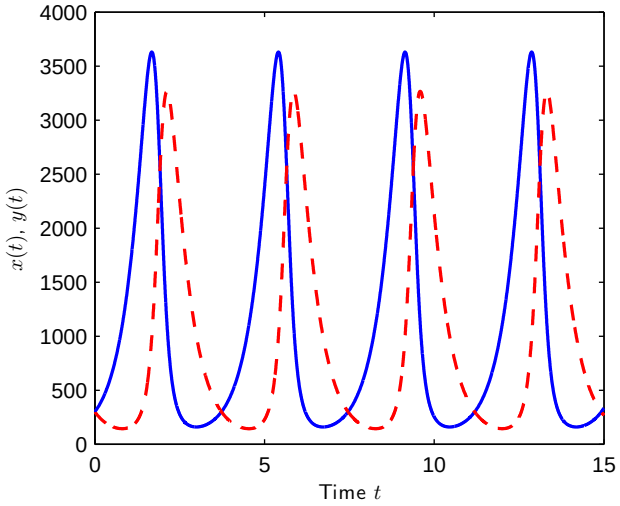


Figure 4.17: Time series plot of the Lotka-Volterra system for $\alpha = 2$, $\beta = 0.002$, $\gamma = 0.0018$, $\delta = 2$ and initial conditions $x(0) = 300$, $y(0) = 300$. The solid line is for the prey population, $x(t)$, and the dashed line represents the predator population, $y(t)$.

Let us continue with a closer look at the predator-prey model for two competing or interacting populations, introduced by Lotka and Volterra:

$$\begin{aligned} \frac{dx}{dt} &= \alpha x - \beta xy \ , \\ \frac{dy}{dt} &= \gamma xy - \delta y \ , \end{aligned} \tag{4.14}$$

where $x, y \geq 0$, and α, β and δ are all positive constants. In (4.14) variable x is to represent the prey population and y the predator population. The structure of the Lotka-Volterra equations imply that in the absence of predators, the prey population will grow unbounded and in the absence of any prey, the predators will be extinguished. δ denotes the natural death rate of the predator and the term involving β describes the death of prey as being proportional to the encounters with predators.

Figure 4.18 shows two visualisations of the dynamic behavior to the Lotka-Volterra system. The x-isoclines is described by those points in the phase-plane for which $dx/dt = 0$, i.e., isocline

$$f(x, y) = \alpha x - \beta xy = 0$$

which is true for

$$x = 0 \quad \text{and} \quad y = \frac{\alpha}{\beta} \ .$$

These are two lines, equal to the y-axis going through the origin and a horizontal line at height α/β. The y-isoclines is defined in the same fashion,

$$g(x, y) = \gamma xy - \delta y$$

M-code 4.1: Matlab function to simulate the Lotka-Volterra equations.

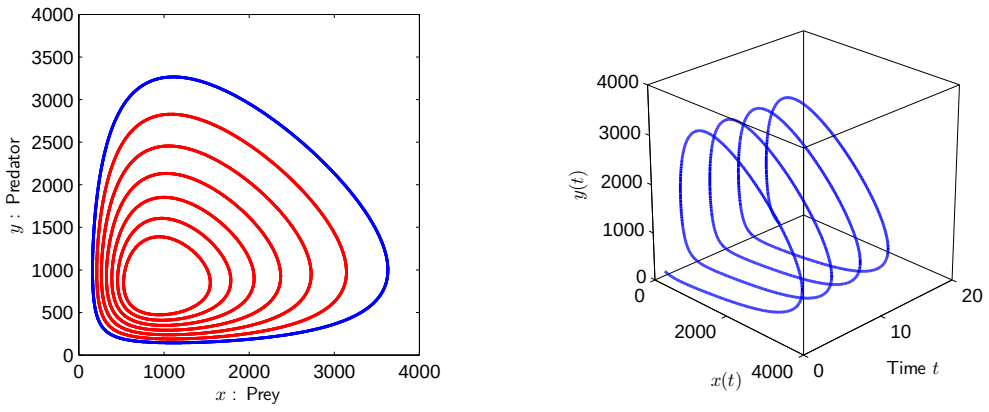
```
x0 =300; y0 = 300; % Initial conditions. x: prey, y: predator
tspan = [0 15]; % Simulation time.
% ODE solver for numerical solution:
[t,z] = ode45(@LV,tspan,[x0 y0],dlt,gamm,alph,bt);
figure, plot(t,z(:,1),t,z(:,2)); % Plot time series.
figure, plot(z(:,1),z(:,2));     % Trajectory in phase-plane.
figure, plot3(z(:,1),t,z(:,2));  % Integral curve.
zlabel('predator'); ylabel('time'); xlabel('prey');
% Subfunction for LV equations:
function dzdt = LV(t,z,dlt,gamm,alph,bt)
%  z(1) : prey, z(2) : predator
dzdt = [alph*z(1)-bt*z(1)*z(2);
gamm*z(1)*z(2)-dlt*z(2)];
```

which is true for

$$y = 0 \quad \text{and} \quad x = \frac{\delta}{\gamma} \, .$$

These are again two perpendicular lines. There are two points of intersection between the x-isoclines and y-isoclines. There are therefore two fixed points for the system. The first fixed point is the origin of the plane, where $x = y = 0$. The second fixed point is given by $y = \alpha/\beta$, $x = \delta/\gamma$, which is a point at which the two populations are balanced so that there is no change to either population.

Plotting trajectories in the phase-plane, as in Figure 4.18, requires a software tool for numerical integration (Matlab code in listing 4.1). A quick way to visualise the flow of a two-dimensional nonlinear systems is to plot for a grid of (x, y)-values the gradient



(a) (b)

Figure 4.18: Phase plane (left) and integral curve (right) of the Lotka-Volterra equations, (4.14).

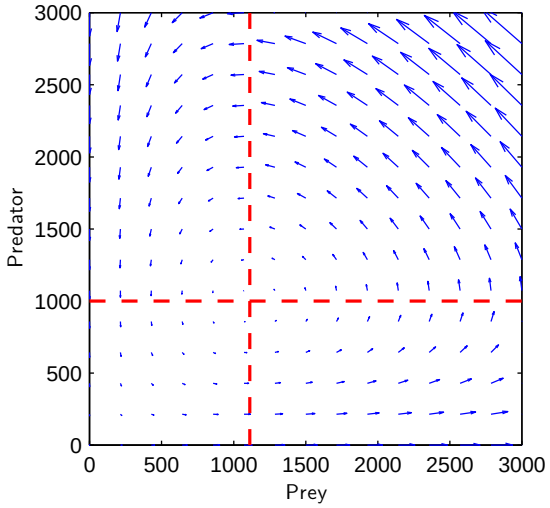


Figure 4.19: Visualisation of the flow of the Lotka-Volterra equations. The horizontal second x-isocline and vertical second y-isocline are shown as dotted lines. The fixed point lies where the isoclines meet. The first isoclines are the axis of the plot going through the origin.

dy/dx as an arrow as shown in Figure 4.19, plotted using the few lines of Matlab code in listing 4.2. As an exercise, you may want to explain why the trajectories in Figure 4.18 are not circles or ellipsoids?

For the Lotka-Volterra system, (4.14), the fixed points occur at $x^* = 0$, $y^* = 0$ and at $x^* = \delta/\gamma$, $y^* = \alpha/\beta$. The constants of the system, linearised at (x^*, y^*), are

$$\begin{aligned} a &= \left.\frac{\partial f}{\partial x}\right|_{x^*,y^*} = \alpha - \beta y^* \ , & b &= \left.\frac{\partial f}{\partial y}\right|_{x^*,y^*} = -\beta x^* \ , \\ c &= \left.\frac{\partial g}{\partial x}\right|_{x^*,y^*} = \gamma y^* \ , & d &= \left.\frac{\partial g}{\partial y}\right|_{x^*,y^*} = \gamma x^* - \delta \ . \end{aligned} \tag{4.15}$$

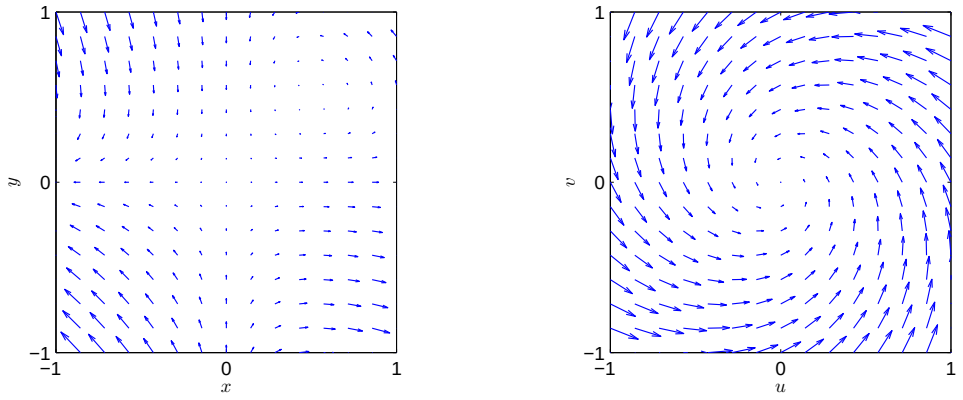
For the fixed point at the origin, $a = \alpha$, $b = 0$, $c = 0$, $d = -\delta$, such that the eigenvalues are

$$\lambda_1 = \alpha \ , \qquad \lambda_2 = -\delta \ .$$

Since α and δ are positive, this fixed point is a saddle point, i.e., trajectories going towards it will drift off just before it. For the predator-prey model this means that even

M-code 4.2: Matlab function to visualise the flow of the Lotka-Volterra system.

```
meshgrid(linspace(0,3000,15));
   dx = alph.*X - bet.*X.*Y;
   dy = gamm.*X.*Y - delt.*Y;
quiver(X,Y,dx,dy,1.5);
xlabel('prey'); ylabel('predator');
```



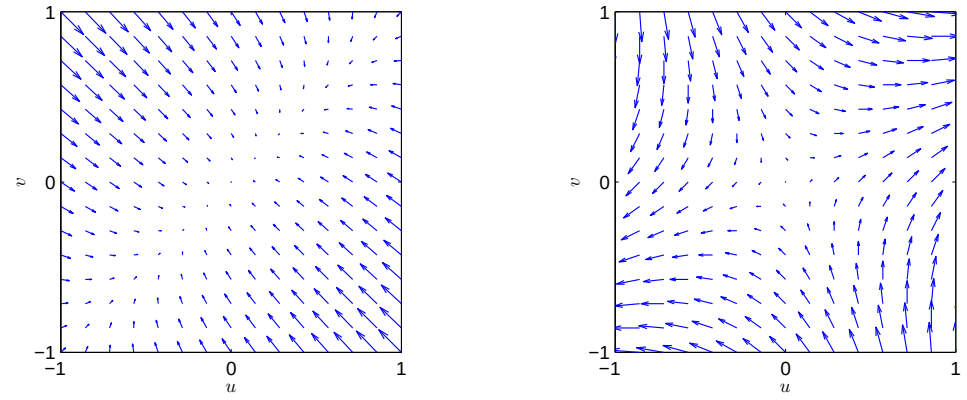
(a) (b)

Figure 4.20: Left: The flow of the Lotka-Volterra system, (4.14), with parameters $\beta = \gamma = 2$, $\alpha = \delta = 1$. Right: stable focus for $a = -1$, $b = -1$, $c = 1.9$, $d = -1$, which leads to eigenvalues $\lambda_1 = -1 + i1.378$, $\lambda_2 = -1 - i1.378$.

if the populations get near the extinction point, the populations will eventually grow again. For the second fixed point, we have $a = 0$, $b = -\beta\delta/\gamma$, $c = \gamma\alpha/\beta$, $d = 0$. The eigenvalues are

$$\lambda = \pm\sqrt{-\alpha\delta}\ .$$

Taking the square root of a negative number will lead to a complex number. The eigenvalues are therefore both imaginary numbers, which implies that the predator and prey populations oscillate around the fixed point, as can be seen well in Figure 4.19.



(a) (b)

Figure 4.21: Left: stable node of the linearised system, (4.10), with parameters taken from [KG95]: $a = -1.5$, $b = 1$, $c = 1$, $d = -1.5$, leading to $\lambda_1 = -2.5$, $\lambda_2 = -0.5$. Right: saddle point of the linearised system, (4.10), for $a = 1$, $b = 1$, $c = 1$, $d = -1$, leading to eigenvalues $\lambda_1 = -\sqrt{2}$, $\lambda_2 = \sqrt{2}$.

This section could only serve as a rudimentary introduction to nonlinear modelling. Other more comprehensive books at introductory level are [Str00a], and [JS07]. The book by Kaplan and Glass [KG95] is particularly suitable for biologists. Advanced texts are [Wig03] and [GH83]. The control engineering perspective of nonlinear systems is described in [Nv90], [Isi89], and [Sas99], all of which are advanced texts. The standard text on mathematical biology by Murray [Mur02] is an excellent source of examples for nonlinear modelling, applied to biological systems.

4.7 Receptor Modelling

In this section we pick up the thread of Section 1.5 and discuss in greater detail a mathematical model of cell surface receptor binding. A comprehensive study of receptors, models for binding, trafficking and signaling was first provided in [LL93] on which we draw heavily. Receptors are most commonly found at the cell surface, where extracellular signaling molecules, the ligand, can bind to them. Signaling proteins include cytokineses, insulin, hormones or growth factors, which could for example be transported through the blood stream. The binding process leads to a transmission of the signal into the cell where it can affect various processes, including the transcription of genes, which in turn can control various important cell functions.

We begin with a basic model for cell surface receptor binding, using the reversible bimolecular reaction

$$L + R \underset{k_d}{\overset{k_a}{\rightleftharpoons}} C$$

where R and L denote the free receptor of a cell and ligand molecules, respectively and C denotes the LR complex, i.e., receptors that are "occupied". k_a is the rate constant at which ligands bind to receptors and k_d is the dissociation constant. We refer to receptor and ligand as *monovalent* to assume that at any time only one ligand molecule and one receptor form a complex. For a single cell, the kinetic model that describes temporal changes in the number of LR complexes is then

monovalent binding

$$\frac{dC}{dt} = k_a R[L] - k_d C \ , \tag{4.16}$$

where $[L]$ gives the free ligand concentration in moles per liter of the medium; R is the number of free receptor per cell (#/cell), C the number of receptor-ligand complexes per cell, k_d and k_a are in sec^{-1} and $M^{-1}\text{sec}^{-1}$, respectively. The number of receptors or complexes per cell can be converted into a concentration (moles per volume solution) or density (#/cell surface area) if necessary. To check the units of (4.16) we have

$$\frac{1}{\text{cell} \cdot \text{sec}} = \frac{1}{\text{M} \cdot \text{sec}} \cdot \frac{1}{\text{cell}} \cdot \text{M} - \frac{1}{\text{sec}} \cdot \frac{1}{\text{cell}} \ .$$

Equation (4.16) has three variables, C, L, and R. A reasonable assumption to simplify the analysis is that the total number of surface receptors, denoted R_T, is constant:

$$R_T = R + C \ , \tag{4.17}$$

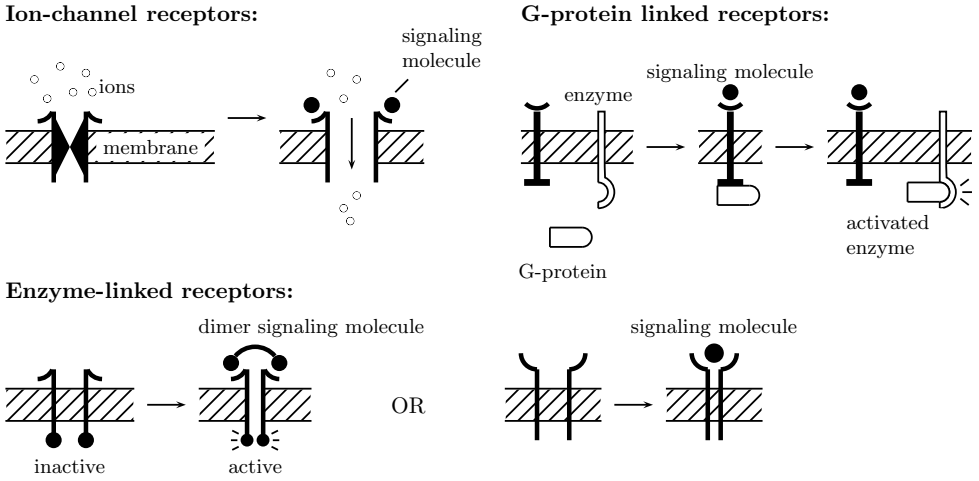


Figure 4.22: The figure illustrate the three most common kinds of cell surface receptors and mechanisms for their activation.

leading to an equation in two variables:

$$\frac{dC}{dt} = k_a(R_T - C)[L] - k_d C \ . \tag{4.18}$$

The ligand concentration is determined by the initial concentration minus the ligands bound in complexes, $L = L_0 - C$. Furthermore, in most experimental set-ups we are going to have n cells in the medium to which also the concentration of the ligand refers. This means that we ought to multiply C (given in #/cell) by n. Furthermore, since C is a count per cell, we turn the concentration of L, given in mol/liter, into a count of numbers by multiplying with N_A the Avogadro constant (#/mol):

$$[L] = L_0 - \left(\frac{n}{N_A}\right) C \ , \tag{4.19}$$

where we write for the initial *concentration* of L, $[L](t = 0) = L_0$. The brackets are therefore left to simplify the notation. Inserting the two conservation assumptions (4.17) and (4.19) into (4.16) gives us a single differential equation to describe the receptor-ligand binding process:

$$\begin{aligned} \frac{dC}{dt} &= k_a(R_T - C)\left(L_0 - \frac{n}{N_A}C\right) - k_d C \\ &= k_a R_T L_0 - k_a R_T \frac{n}{N_A} C - k_a L_0 C + k_a \frac{n}{N_A} C^2 - k_d C \ . \end{aligned} \tag{4.20}$$

Rewriting this equation we recognise it as an inhomogeneous second-order differential equation:

$$\frac{dC}{dt} + \left(k_a R_T \frac{n}{N_A} + k_a L_0 + k_d\right) C - k_a \frac{n}{N_A} C^2 = k_a R_T L_0 \ . \tag{4.21}$$

4.7.1 Negligible ligand depletion

In order to obtain a simpler differential equation than (4.21) we make further assumptions. If we assume that ligand depletion can be neglected and therefore replace L in (4.18) with L_0 the differential equation becomes dependent on only one variable:

$$\frac{dC}{dt} = k_a(R_T - C)L_0 - k_d C \ . \tag{4.22}$$

If we imagine there are initially no ligands bound, the assumption of negligible ligand depletion implies that the initial concentration is much larger than the ligand bound in complexes, i.e., $(n/N_A)C \ll L_0$. Alternatively we could have replaced $L_0 - (n/N_A)C$ by L_0 in (4.20) to arrive at (4.22).

Let us now determine the solution to differential equation (4.22), by recognising it as a linear, inhomogeneous, first-order ordinary differential equation[4]

$$\frac{dC}{dt} + \underbrace{(k_a L_0 + k_d)}_{P(t)} C = \underbrace{k_a R_T L_0}_{Q(t)} \ . \tag{4.23}$$

Such equations are solved by defining an *integrating factor* integrating factor

$$\begin{aligned}\rho(t) &= e^{\int P(t)dt} \\ &= \exp\left\{\int (k_a L_0 + k_d)\ dt\right\} \\ &= \exp\left\{(k_a L_0 + k_d)t\right\} \ .\end{aligned}$$

The next step is to multiply both sides of (4.23) with $\rho(t)$

$$\exp\left\{(k_a L_0 + k_d)t\right\} \frac{dC}{dt} + \exp\left\{(k_a L_0 + k_d)t\right\} (k_a L_0 + k_d)C = \exp\left\{(k_a L_0 + k_d)t\right\} k_a R_T L_0 \ .$$

We notice that the left-hand side is the derivative of the product $C(t) \cdot \rho(t)$, which is in fact the whole idea behind the use of an integrating factor. We can thus write

$$\rho(t) \cdot C(t) = \int \rho(t) \cdot Q(t) dt + c \ ,$$

where c is some arbitrary constant. Insert the expressions for $\rho(t)$ and $Q(t)$,

$$\begin{aligned}\rho(t) \cdot C(t) &= \int \exp\left\{(k_a L_0 + k_d)t\right\} \cdot k_a R_T L_0\ dt + c \\ &= k_a R_T L_0 \int_0^t e^{(k_a L_0 + k_d)t}\ dt + c \\ &= \frac{k_a R_T L_0}{k_a L_0 + k_d}\left(e^{(k_a L_0 + k_d)t} - 1\right) + c \ ,\end{aligned}$$

[4] This is an ordinary differential equation since there is only a single independent variable, t. The equation is linear since no terms such as C^2 appear and it is first-order since only the first derivative dC/dt appears. The homogenous version of (4.22) is obtained for $k_a R_T L_0 = 0$. This special case would correspond to the monomolecular reaction (2.2) on page 98.

leading to

$$C(t) = \frac{k_a R_T L_0}{k_a L_0 + k_d} - \frac{k_a R_T L_0}{k_a L_0 + k_d} e^{-(k_a L_0 + k_d)t} + c \cdot e^{-(k_a L_0 + k_d)t} \ , \tag{4.24}$$

where $1/\rho(t) = \exp\{-(k_a L_0 + k_d)t\}$. From the initial condition, $C(t = 0) = C_0$, we can determine the constant

$$C_0 = \frac{k_a R_T L_0}{k_a L_0 + k_d} - \frac{k_a R_T L_0}{k_a L_0 + k_d} + c \ ,$$

i.e., $c = C_0$, which inserted into (4.24), gives us the solution

$$C(t) = C_0 \exp\{-(k_a L_0 + k_d)t\} + \frac{k_a R_T L_0}{k_a L_0 + k_d} \left(1 - \exp\{-(k_a L_0 + k_d)t\}\right) \ . \tag{4.25}$$

4.7.2 Equilibrium and steady state

equilibrium

At *equilibrium* the reaction rates are equal

$$k_a[L]R = k_d C \ ,$$

dissociation constant

where the *dissociation constant* is defined (cf. page 79) as

$$K_d = \frac{R[L]}{C} = \frac{k_d}{k_a} \ ,$$

where here we have for $R = R_T - C$ and $[L](t = 0) = L_0$, from our assumptions (4.17) and (4.19) above:

$$K_d = \frac{(R_T - C)L_0}{C} \ ,$$

with unit M (mol per liter). Let us denote by C_{eq} the number of ligand-receptor complexes at equilibrium,

$$C_{\text{eq}} = \frac{R_T L_0}{K_d + L_0} \ . \tag{4.26}$$

steady state

At *steady state*, $dC/dt = 0$,

$$0 = k_a(R_T - C)L_0 - k_d C \ ,$$

leading to

$$C = \frac{k_a R_T L_0}{k_d + k_a L_0} = \frac{R_T L_0}{K_d + L_0}.$$

The steady state value is therefore in this case identical to the number of receptor-ligand complexes at equilibrium.

4.7.3 Dimensionless representation

To visualise the solution (4.25) with a plot, we would have to specify the rate constants, the initial ligand concentration, the total number of receptors and the initial number of receptor-ligand complexes. The appearance of the plot may therefore vary considerably for different values of these parameters. This can be avoided by not looking at $C(t)$ but plotting the dimensionless fraction of occupied receptors,

$$y = \frac{C}{R_T} \ , \tag{4.27}$$

where $0 \leq y \leq 1$. For $y = 0$ all receptors are free (no complexes), and for $y = 1$ all receptors are occupied. Let us furthermore introduce a scaled time, τ,

$$\tau = k_d t \ . \tag{4.28}$$

Rewriting (4.22) first by dividing both sides by R_T

$$\begin{aligned} \frac{d}{dt}\left(\frac{C}{R_T}\right) &= k_a \left(1 - \frac{C}{R_T}\right) L_0 - k_d \frac{C}{R_T} \ , \\ \frac{dy}{dt} &= k_a (1-y) L_0 - k_d y \ , \end{aligned}$$

next taking account of (4.28) gives

$$\begin{aligned} \frac{dy}{d\tau} &= \frac{k_a}{k_d}(1-y)L_0 - y \ , \\ &= \frac{L_0}{K_d}(1-y) - y \ . \end{aligned} \tag{4.29}$$

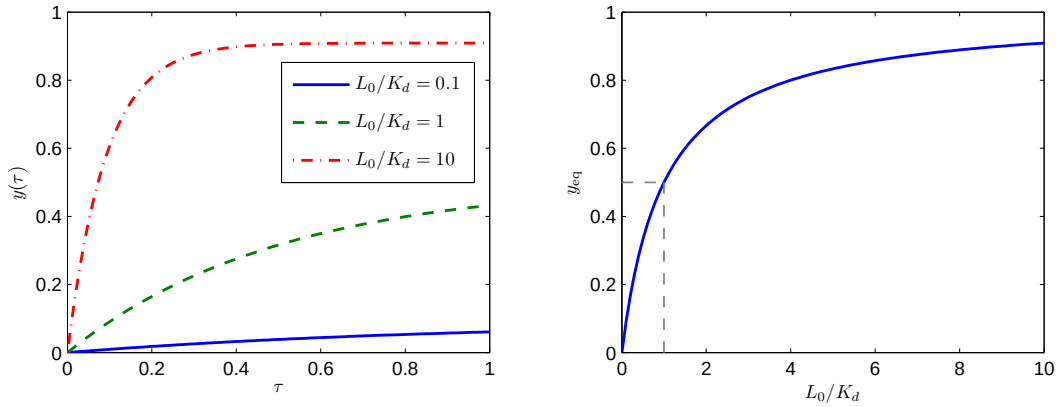
For the fractional occupancy of receptors, the transient changes are therefore described by the following solution of (4.29)

$$y(\tau) = y_0 \exp\left\{-\left(1 + \frac{L_0}{K_d}\right)\tau\right\} + \frac{L_0/K_d}{1 + (L_0/K_d)}\left(1 - \exp\left\{-\left(1 + \frac{L_0}{K_d}\right)\tau\right\}\right) \ , \tag{4.30}$$

and the equilibrium value is determined by the ratio L_0/K_d:

$$y_{\text{eq}} = \frac{L_0/K_d}{1 + (L_0/K_d)} \ . \tag{4.31}$$

See Figure 4.23 for an illustration of the transient binding of ligands to cell receptors. Note that $y_{\text{eq}} = C_{\text{eq}}/R_T$ and $y_{\text{eq}} = 0.5$ when $L_0/K_d = 1$, i.e., half the receptors are bound by ligands at equilibrium when the ligand concentration is equal to the value of K_d.



(a) (b)

Figure 4.23: Left: Transient binding of ligands to cell receptors, where a bimolecular reaction is used as a model, with ligand depletion assumed to be negligible and the total number of receptors remaining unchanged. At $t = 0$, it is assumed that all receptors are free. Right: The fractional occupancy of receptors at equilibrium.

4.7.4 Half-Time

Taking a break before we continue with our model of ligand binding, we complete the previous study by looking at the *half-time* of an experiment. Let us assume an experiment with initial condition $y_0 = C_0/R_T = 0$, i.e., initially no ligands are bound to the receptors. half-time

One definition for a half-time, τ_h, is for the transient solution $y(\tau)$ to reach half of the change from $y_0 = 0$ to y_{eq}, $y = 0.5y_{\text{eq}}$:

$$\frac{1}{2}y_{\text{eq}} = y_{\text{eq}}\left[1 - \exp\left\{-\left(1 + \frac{L_0}{K_d}\right)\tau_h\right\}\right] , \tag{4.32}$$

$$\frac{1}{2} = -\exp\left\{-\left(1 + \frac{L_0}{K_d}\right)\tau_h\right\} , \tag{4.33}$$

$$-\left(1 + \frac{L_0}{K_d}\right)\tau_h = \ln\frac{1}{2} , \tag{4.34}$$

such that

$$\tau_h = \frac{-\ln 1/2}{1 + L_0/K_d} = \frac{\ln 2}{1 + L_0/K_d} \approx \frac{0.69}{1 + L_0/K_d} . \tag{4.35}$$

This half time is shown in the right-hand plot of Figure 4.23.

Alternatively, we may ask for the left-hand plot of $y(\tau)$ in Figure 4.23 for when $y(\tau_h) =$

1/2. Considering again an initial value $y(t=0)=0$,

$$\begin{aligned}
\frac{1}{2} &= y_{\text{eq}}\left[1-\exp\left\{-\left(1+\frac{L_0}{K_d}\right)\tau_h\right\}\right] \ , \\
\frac{1}{2y_{\text{eq}}} &= 1-\exp\left\{-\left(1+\frac{L_0}{K_d}\right)\tau_h\right\} \ , \\
\exp\left\{-\left(1+\frac{L_0}{K_d}\right)\tau_h\right\} &= -\left(\frac{1-2y_{\text{eq}}}{2y_{\text{eq}}}\right) \ , \\
-\left(1+\frac{L_0}{K_d}\right)\tau_h &= \ln\left(\frac{2y_{\text{eq}}-1}{2y_{\text{eq}}}\right) \ .
\end{aligned}$$

Since it must be possible for half of the receptors to be occupied, $y_{\text{eq}} \geq 0.5$.

$$\begin{aligned}
\left(1+\frac{L_0}{K_d}\right)\tau_h &= -\ln\left(\frac{2y_{\text{eq}}-1}{2y_{\text{eq}}}\right) = \ln\left(\frac{2y_{\text{eq}}}{2y_{\text{eq}}-1}\right) \ , \\
\tau_h &= \left(\ln\frac{2y_{\text{eq}}}{2y_{\text{eq}}-1}\right)\cdot\left(1+\frac{L_0}{K_d}\right)^{-1} = \frac{\ln[2y_{\text{eq}}/(2y_{\text{eq}}-1)]}{1+L_0/K_d} \ .
\end{aligned}$$

4.7.5 Stimulus-Response analysis

The previous study of receptor-ligand binding, based on equation (4.22), assumed that the ligand concentration is more or less constant. The solution (4.25), respectively (4.30), can be thus be interpreted as the response to a step-change in the ligand concentration. We now return to equation (4.18) and consider different kinds of stimuli.

$$\begin{aligned}
\frac{dC}{dt} &= k_a(R_T-C)[L]-k_dC \qquad (4.18)\\
&= k_aR_T[L]-k_aC[L]-k_dC \ ,
\end{aligned}$$

where both C and the ligand concentration $[L](t)$ are a function of time. A check of the units is quickly done for the last equation

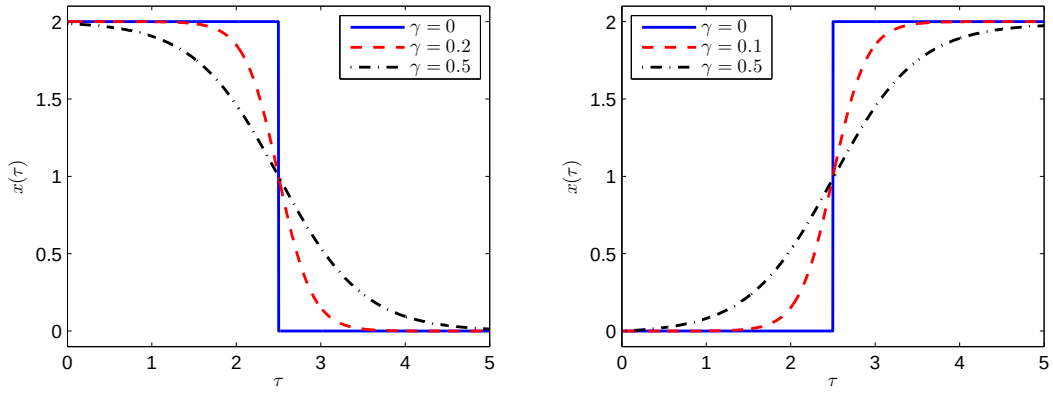
$$\frac{\#}{\text{cell}\cdot\text{sec}} = \frac{1}{\text{M}\cdot\text{sec}}\cdot\frac{\#}{\text{cell}}\cdot\text{M}-\frac{1}{\text{M}\cdot\text{sec}}\cdot\frac{1}{\text{cell}}\cdot\text{M}-\frac{1}{\text{sec}}\cdot\frac{\#}{\text{cell}} \ .$$

To investigate the response in receptor binding to different forms of ligand stimulation we again use the dimensionless variable $y = C/R_T$, which represents the fraction of occupied receptors

$$\frac{dy}{dt} = k_a[L]-k_ay[L]-k_dy \ .$$

Let us further scale time by introducing $\tau = t\cdot k_d$, which means that the right-hand side of the above equation is divided by k_d

$$\frac{dy}{d\tau} = \frac{k_a}{k_d}[L]-\frac{k_a}{k_d}y[L]-y \ .$$



(a) (b)

Figure 4.24: Left: Downward step-change stimulus $x(\tau)$ for different parameter values of γ; $\alpha = 2$, $\beta = 2.5$. Right: Positive step.

For the reversible bimolecular reaction the equilibrium constant and dissociation constants are defined as (cf. page 79)

$$K_{\text{eq}} = \frac{k_a}{k_d} , \qquad K_d = \frac{k_d}{k_a} ,$$

such that we have

$$\frac{dy}{d\tau} = \frac{1}{K_d}[L] - \frac{1}{K_d}y[L] - y . \tag{4.36}$$

In order to make this equation more appealing for the eye, we hereafter use x to denote the stimulus L and replace the equilibrium constant by θ:

$$\frac{dy}{d\tau} = \theta x - \theta xy - y . \tag{4.37}$$

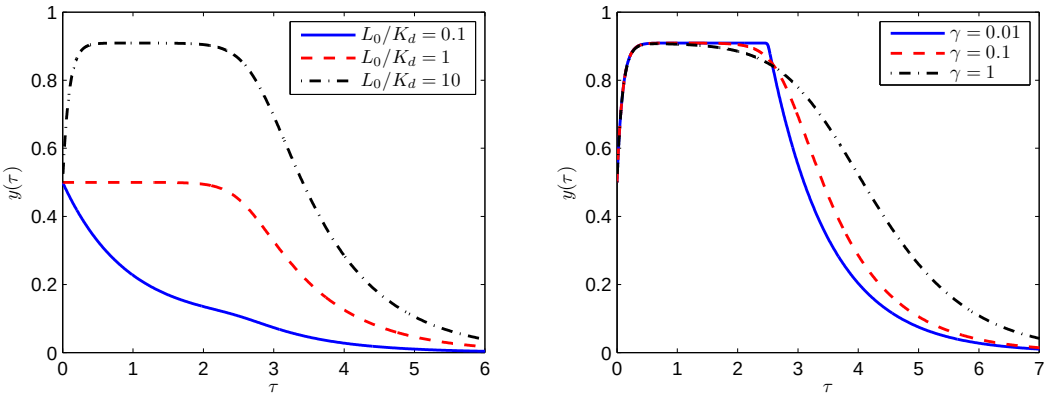
This then is a nonlinear ordinary differential equations with one parameter θ and two temporal variables $x(t)$ and $y(t)$. We begin with a downward step-change

$$x(\tau) = \frac{\alpha}{1 + \exp\left\{\frac{\tau - \beta}{\gamma}\right\}} , \tag{4.38}$$

where the parameters

α: determines the initial height,
β: defines the turning point of the curve,
γ: determines sharpness of the transition.

For $\gamma \to 0$ we obtain the Heaviside step-function as illustrated in Figure 4.24. In Figure 4.25, on the left, the solution of (4.37) is shown for three different ratios of



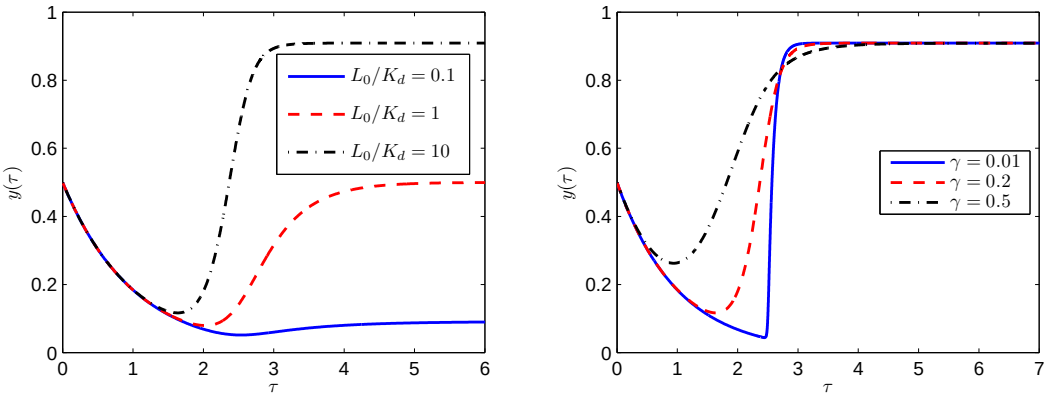
(a) (b)

Figure 4.25: Left: Response to downward step-changes for initial condition $y = C/R_T = 0.5$ and different values for L_0/K_d; $\alpha = 2$, $\beta = 2.5$ and $\gamma = 0.2$. Right: Response to a downward step with the slope of the change, γ, changing; $L_0/K_d = 10$ and α and β as before.

L_0/K_d, $\alpha = 2$, $\beta = 2.5$ and $\gamma = 0.2$. The right-hand plot of Figure 4.25 shows the response to a downward step with the slop of the change, γ, changing; $L_0/K_d = 10$ and α and β as before. The inverse of the "downward" or "negative" step, is the function

$$x(\tau) = \frac{\alpha}{1 + \exp\left\{-\frac{\beta}{\gamma}\right\}} - \frac{\alpha}{1 + \exp\left\{\frac{\tau - \beta}{\gamma}\right\}} .$$

The response pattern to this stimulus are shown in Figure 4.26.



(a) (b)

Figure 4.26: Left: Response to downward step-changes for initial condition $y = C/R_T = 0.5$ and different values for L_0/K_d; $\alpha = 2$, $\beta = 2.5$ and $\gamma = 0.2$. Right: Response to a downward step with the slope of the change, γ, changing; $L_0/K_d = 10$ and α and β as before.

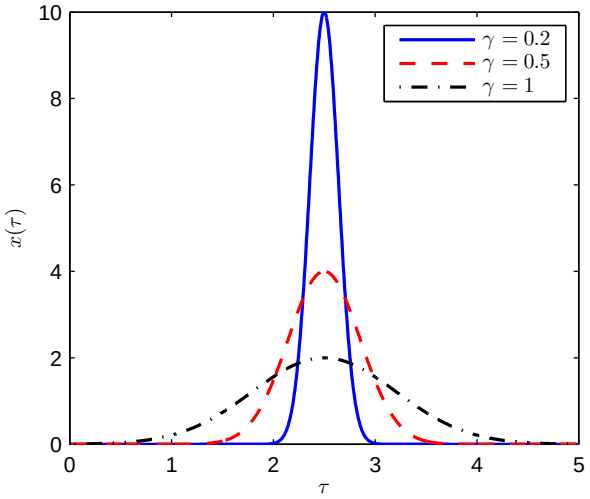
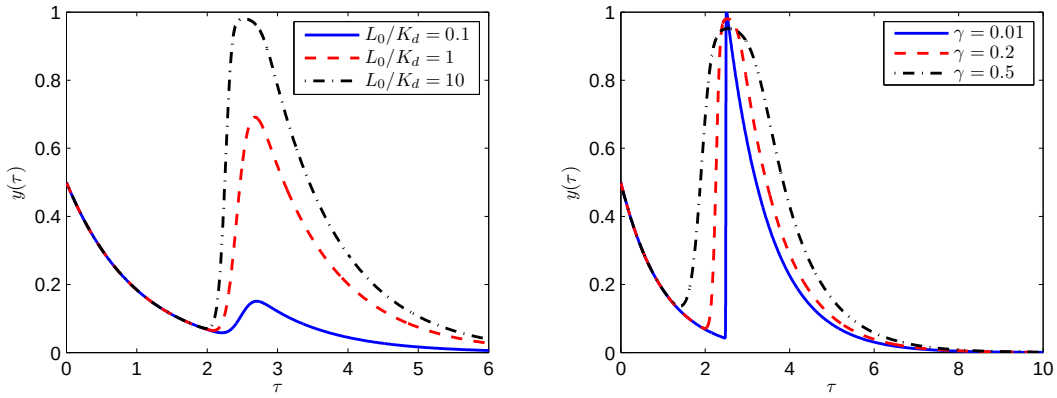


Figure 4.27: Impulse-like stimulus $x(\tau)$, $\alpha = 2$, $\beta = 2.5$ for different width parameter.



(a) (b)

Figure 4.28: Left: Response to impulse stimulus for initial condition $y = C/R_T = 0.5$ and different values for L_0/K_d; $\alpha = 2$, $\beta = 2.5$ and $\gamma = 0.2$. Right: Response to an impulse with width, γ, changing; $L_0/K_d = 10$ and α and β as before.

Next we consider an 'impulse'-like function, which we represent by a gaussian function

$$x(\tau) = \frac{\alpha}{\gamma} \cdot \exp\left\{-\frac{(\tau - \beta)^2}{\gamma^2}\right\} . \tag{4.39}$$

The pre-factor α/γ is chosen in this way as to ensure the integral of the right-hand side (i.e., the intensity) is constant. For $\alpha = 2$, $\beta = 2.5$ and changing width, γ, the function is shown in Figure 4.27 and the response pattern in Figure 4.28.

4.7.6 Conclusions

For a full understanding of the dynamics of a system it is necessary to conduct a series stimulus-response experiments. For most experiments an initial concentration of ligands is depleted and it is not possible to control the exact shape of the stimulus. Here we have considered 'typical' input stimuli that may occur in a system.

We can easily extend the model of extracellular ligands binding to receptors, as discussed above, to *intra*cellular signaling. A common mechanism for receptor regulated signalling is *dimerisation*, a ligand induced monomer-to-dimer transition. As a monomer a single receptor is inactive, dimerisation leads to an activation and intracellular autophosphorylation of the signaling domain as illustrated in Figure 4.29.

dimerisation

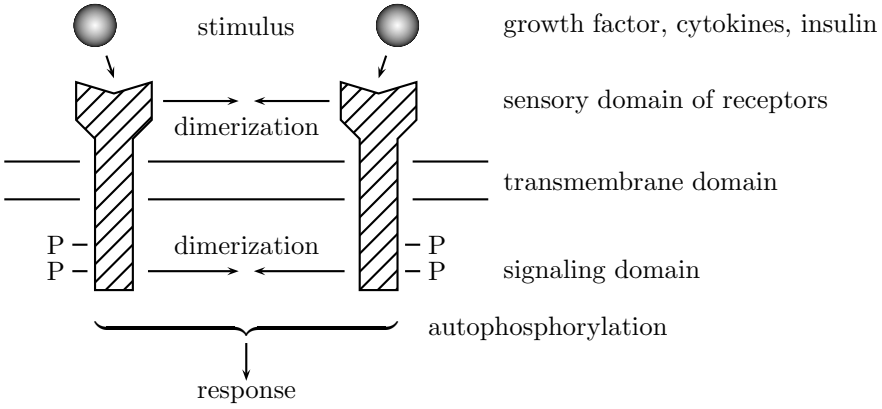


Figure 4.29: Example of a common mechanism for receptor regulated signalling. Dimerisation is a ligand induced monomer-to-dimer transition. As a monomer a single receptor is inactive, dimerisation leads to an activation and intracellular autophosphorylation of the signaling domain.

Let us denote by $\tilde{S}$ the non-phosphorylated form of a molecular species or substrate S. The model for a signaling step suggested here is a phosphorylation

$$\tilde{S}_i + S_{i-1} \underset{k_{di}}{\overset{k_{ai}}{\rightleftharpoons}} S_i$$

where analog to (4.16)

$$\frac{d[S_i]}{dt} = k_{ai}[\tilde{S}_i][S_{i-1}] - k_{di}[S_i] \ .$$

We assume the total concentration of kinase S_i is constant

$$[S_{Ti}] \doteq [\tilde{S}_i] + [S_i] \ ,$$

which inserted into the previous equation gives

$$\begin{aligned} \frac{d[S_i]}{dt} &= k_{ai}[S_{i-1}]([S_{Ti}] - [S_i]) - k_{di}[S_i] \ , \\ &= k_{ai}[S_{Ti}][S_{i-1}] - k_{ai}[S_{i-1}][S_i] - k_{di}[S_i] \ . \end{aligned}$$

In support of our eyesight we introduce x_i to denote the phosphorylation of S_i and write $c_i \doteq [S_{Ti}]$, $\alpha_i \doteq k_{ai}c_i$, $\beta_i \doteq k_{di}$, leading to

$$\begin{aligned} \frac{dx_i}{dt} &= \alpha_i x_{i-1} - \frac{1}{c_i} x_{i-1}x_i - \beta_i x_i \ , \\ \frac{dx_i}{dt} &= \alpha_i x_{i-1}\left(1 - \frac{x_i}{\alpha_i c_i}\right) - \beta_i x_i \ . \end{aligned} \tag{4.40}$$

This last equation is the basis for an interesting study of properties of a signaling pathway, first introduced by Heinrich et al. [HNR02] and which we are going to discuss in greater detail in Section 4.8.5.

4.8 Dynamic Modelling of Biochemical Networks

Pathways are networks of biochemical reactions, most of which are facilitated by highly specialised enzymes. The enzyme kinetic reaction can therefore serve as a template to construct more complex pathway models. In the present section we therefore first look at the equations that represent an enzyme kinetic reaction and hint at the use in dynamic modelling of signal transduction pathways.

We are going to consider a compartment or region of the cell with volume V for which we assume that diffusion is fast compared to the time scales of the reactions and hence concentrations within this volume are homogenous. In many cases it is possible to decompose more complex reaction networks into a set of uni- or monomolecular (first-order) reactions and bimolecular (second order reactions), depicted

$$S \xrightarrow{k_m} \cdots \quad \text{respectively} \quad S + E \xrightarrow{k_b} \cdots \ ,$$

where the arrow denotes a conversion according to the law of mass action. Concentrations are specified in mol per liter (M). When it is clear from the context, the square brackets which denote concentrations, are often left away to have a simpler notation. The letters for the variables are chosen arbitrarily and depending on the context. The

reaction rate

rate of the reaction or *reaction rate* v is, in case of the monomolecular reaction defined by the product k_m and $[S]$ and in case of the bimolecular reaction defined by the product of k_b with $[S]$ and $[E]$:

$$\begin{aligned} S &\xrightarrow{k_m} \cdots \quad \text{where } v = k_m[S] \\ S + E &\xrightarrow{k_b} \cdots \quad \text{where } v = k_b[S][E] \end{aligned}$$

The linear (monomolecular reaction), respectively bilinear relationship (bimolecular reaction) of the reaction rate on the concentrations is in essence the law of mass action. Note that the units of the rate constant k is per second (sec^{-1}) for the monomolecular reaction and in moles per second ($\text{M}^{-1}\text{sec}^{-1}$) for bimolecular reactions.

One approach to model more complex signal transduction pathways is to model each step of the pathway on a template of an enzyme kinetic reaction (see also Section 2.7)

$$E + S \underset{k_1}{\overset{k_2}{\rightleftharpoons}} ES \xrightarrow{k_3} P + E \tag{4.41}$$

with k_1 denoting the rate at which the complex ES is formed; k_2 at which ES dissociates into enzyme E and substrate S; k_3, the rate at which ES dissociates into product P and E. The reaction diagram (4.41) can be decomposed into a set of mono- and bi-molecular reactions

$$\begin{aligned} E + S &\xrightarrow{k_1} ES\ , \quad v_1 = k_1[E][S] \\ ES &\xrightarrow{k_2} E + S\ , \quad v_2 = k_2[ES] \\ ES &\xrightarrow{k_3} P + E\ , \quad v_3 = k_3[ES] \end{aligned}$$

The ordinary differential equation model is directly derived from these reactions

$$\left.\begin{aligned} \frac{d[E]}{dt} &= -k_1[E][S] + k_2[ES] + k_3[ES] \\ \frac{d[S]}{dt} &= -k_1[E][S] + k_2[ES] \\ \frac{d[ES]}{dt} &= k_1[E][S] - k_2[ES] - k_3[ES] \\ \frac{d[P]}{dt} &= k_3[ES]\ . \end{aligned}\right\} \tag{4.42}$$

Figure 4.30 gives a graphical representation of these equations. If these graphical representation have a one-to-one mapping to the equations, they are an important communication tool in interdisciplinary collaborations. Biologists naturally draw cartoons to represent the relationship between variables in a pathway. There is however no standard about the meaning of symbols and it is usually not obvious how to translate this into equations.

Using the rate of reactions, an alternative compact representation of (4.42) is

$$\left.\begin{aligned} \frac{d[E]}{dt} &= -v_1 + v_2 + v_3 \\ \frac{d[S]}{dt} &= -v_1 + v_2 \\ \frac{d[ES]}{dt} &= v_1 - v_2 - v_3 \\ \frac{d[P]}{dt} &= v_3\ . \end{aligned}\right\} \tag{4.43}$$

The enzyme is considered a catalyst, which facilitates the reaction without loss, i.e., the total enzyme concentration, i.e. the sum of free enzyme $[E]$ and enzyme in the complex

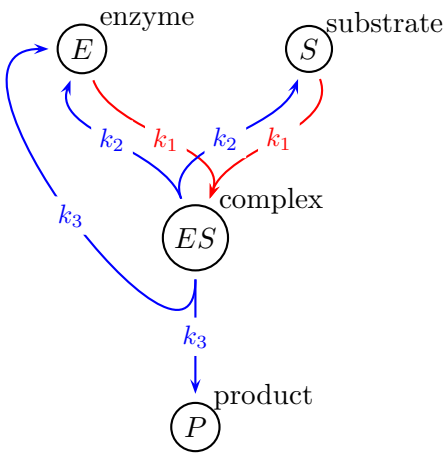


Figure 4.30: Graphical representation of the enzyme kinetic reaction. More complex signal transduction pathways can be constructed using this template.

$[ES]$, is constant. This is also apparent from adding the equations for $[\dot{E}]$ and $[\dot{ES}]$:

$$[\dot{E}] + [\dot{ES}] = 0 \quad \text{which implies} \quad [E](t) + [ES](t) = c_1 \ .$$

Assuming there is initially no complex, $[ES](0) = 0$, the constant equals the initial enzyme concentration $c_1 = [E](0)$. Inserting $[E](t) = [E](0) - [ES](t)$ into the equation for substrate $[S]$, and complex $[ES]$, the system of ODEs reduces to two equations:

$$[\dot{S}] = -k_1[E](0)[S] + (k_1[S] + k_2)[ES] \qquad \text{(substrate)}$$
$$[\dot{ES}] = k_1[E](0)[S] - (k_1[S] + k_2 + k_3)[ES] \qquad \text{(complex)}$$

with initial conditions $[S](0)$ and $[ES](0) = 0$.

The structure of the equations (i.e., the signs, the number of terms, and the variables involved) are obtained as a direct translation of the biologist's knowledge of a pathway. This knowledge is usually not firm and the purpose of modelling is not only to fit experimental data to an assumed model, but to identify an appropriate model structure, validating or updating the knowledge we have of the proteins in a pathway and how they interact. To decide whether a model structure is realistic, a simulation could reveal whether the concentration profiles match experimental experience. For a simulation we do however need to know the values of the parameters. Ideally, we would like to derive some general properties of the system, without knowing the exact parameter values. For the set of equations (4.42) this can be done easily by looking at the equations. We know that the enzyme $[E]$ turns the substrate $[S]$ into the product $[P]$ and thus we would expect $[P]$ to increase steadily while $[S]$ decreases. The last equation $[\dot{P}] = k_3[ES]$ makes the product increase so long as $[ES]$ is positive. Since we deal with concentrations, all x_i can only be positive. Looking at the equation for the substrate, $[S]$ will decrease so long as the right-hand side is negative, i.e., $k_1[E][S] > k_2[ES]$. Thus from any initial condition $[E]$, $[S]$, and $[ES]$ would adjust themselves steadily until $k_1[E][S] > k_2[ES]$ and from then on $[S]$ would decrease steadily.

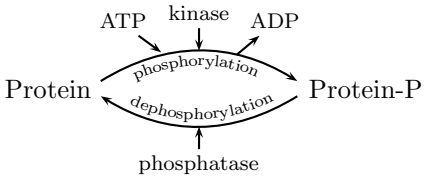


Figure 4.31: A common means of signal transduction is through sequential phosphorylation and dephosphorylation. Phosphorylation is facilitated by means of a kinase and dephosphorylation is realised by a phosphatase. The phosphorylated state is denoted by adding -P to the name of the protein. More complex pathway diagrams can be built by connecting these basic components into cascades and networks.

Symbolically, we can summarise the enzyme kinetic reaction model (2.137), (4.41) as follows:

$$S \overset{E}{\Rightarrow} P$$

We read this as "the conversion of S into P, facilitated (or catalyzed) by E". For example, signal transduction pathways are commonly considered as a series or cascade of modules, each of which can be modelled using the enzyme kinetic reaction as a template. The signal in these pathways is transmitted through facilitated phosphorylation of proteins referred to as 'kinases':

1. $P_3 \overset{P_4}{\Rightarrow} P_3^*$
2. $P_2 \overset{P_3^*}{\Rightarrow} P_2^* \overset{P_3^*}{\Rightarrow} P_2^{**}$
3. $P_1 \overset{P_2^{**}}{\Rightarrow} P_1^* \overset{P_2^{**}}{\Rightarrow} P_1^{**}$

where the * and ** denote phosphorylation and double phosphorylation, respectively. Here P_4 facilitates the phosphorylation of P_3 and so forth. More generally we use the * to denote an activation, which can but must not be achieved by phosphorylation. A phosphorylation, e.g. of MEK, is often also denoted by adding -P to the name of the protein, MEK-P, or ERK-PP, for phosphorylation and doublephosphorylation, respectively. Figure 4.31 shows another common way to illustrate signaling steps in diagrams.

4.8.1 Simulation example

As an illustration, we here describe the simulation of the enzyme-kinetic reaction (4.42). The four differential equations of (4.42) are an example for the mass action representation (1.30) consisting of N ordinary differential rate equations

$$\frac{d}{dt}[S_i] = \sum_{\mu=1}^{M} \nu_{\mu i} k_\mu \prod_{j=1}^{L_\mu} [S_{p(\mu j)}]^{l_{\mu j}} \qquad i = 1, 2, \ldots, N \tag{4.44}$$

where the k_μ's are rate constants and ν_μ denotes the change in molecules of S_i resulting from a single R_μ reaction. For more complex reaction networks one first has to divide

reversible reactions up into basic reaction channels

$$R_\mu \colon l_{\mu 1} S_{p(\mu 1)} + l_{\mu 2} S_{p(\mu 2)} + \cdots + l_{\mu L_\mu} S_{p(\mu L_\mu)} \xrightarrow{k_\mu} \cdots$$

where L_μ is the number of reactant species in channel R_μ, $l_{\mu j}$ is the stoichiometric coefficient of reactant species $S_{p(\mu j)}$, $K_\mu = \sum_{j=1}^{L_\mu} l_{\mu j}$ denotes the molecularity of reaction channel R_μ, and the index $p(\mu j)$ selects those S_i participating in R_μ. For the enzyme-kinetic reaction there are $M = 3$ reaction channels

$$\begin{aligned} R_1 &: E + S \xrightarrow{k_1} ES && \text{(bimolecular reaction)} \\ R_2 &: ES \xrightarrow{k_2} E + S && \text{(monomolecular reaction)} \\ R_3 &: ES \xrightarrow{k_3} E + P && \text{(monomolecular reaction)} \end{aligned}$$

For $i = 1, \ldots, N = 4$, we translate the names 'Enzyme', 'Substrate', 'Enzyme/Substrate complex', and 'Product' into the notation of Chapter 2:

$$S_1 \doteq E \quad S_2 \doteq S \ , \quad S_3 \doteq ES \ , \quad S_4 \doteq P \ .$$

Subsequently, we have the matrix $\nu = [\nu_{\mu i}]$

$$\begin{aligned} &\nu_{11} = -1 \ , && \nu_{12} = -1 \ , && \nu_{13} = +1 \ , && \nu_{14} = 0 \\ &\nu_{21} = +1 \ , && \nu_{22} = +1 \ , && \nu_{23} = -1 \ , && \nu_{24} = 0 \\ &\nu_{31} = +1 \ , && \nu_{32} = 0 \ , && \nu_{33} = -1 \ , && \nu_{34} = 1 \end{aligned}$$

The indices for participating species are collected in terms of vectors p_μ

$$p_1 = (1, 2) \ , \quad p_2 = 3 \ , \quad p_3 = 3$$

Similar, to facilitate the software implementation of the equations, the stoichiometry is defined by

$$l_1 = (1, 1) \ , \quad l_2 = 1 \ , \quad l_3 = 1$$

such that the molecularity $K_\mu = \sum_{j=1}^{L_\mu} l_{\mu j}$ is encoded as follows

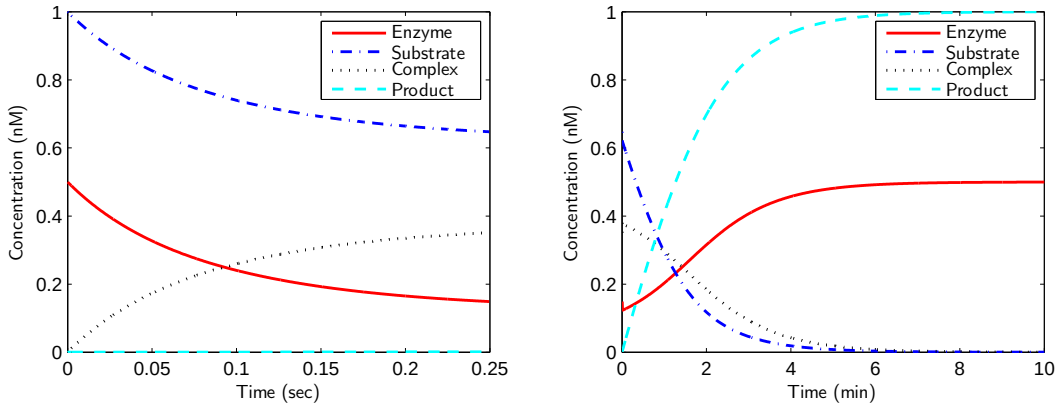
$$\begin{aligned} &L_1 = 2 \ , && L_2 = 1 \ , && L_3 = 1 \\ &K_1 = 2 \ , && K_2 = 1 \ , && K_3 = 1 \ . \end{aligned}$$

This leads us to a representation of the enzyme-kinetic reaction (4.41) in the form of (4.44)

$$\begin{aligned} \frac{dS_1}{dt} &= \nu_{11} k_1 S_1^1 S_2^1 + \nu_{21} k_2 S_3^1 + \nu_{31} k_3 S_3^1 \\ \frac{dS_2}{dt} &= \nu_{12} k_1 S_1^1 S_2^1 + \nu_{22} k_2 S_3^1 + \nu_{32} k_3 S_3^1 \\ \frac{dS_3}{dt} &= \nu_{13} k_1 S_1^1 S_2^1 + \nu_{23} k_2 S_3^1 + \nu_{33} k_3 S_3^1 \\ \frac{dS_4}{dt} &= \nu_{14} k_1 S_1^1 S_2^1 + \nu_{24} k_2 S_3^1 + \nu_{34} k_3 S_3^1 \end{aligned}$$

M-code 4.3: Matlab code to simulate an enzyme-kinetic reaction as an example of the mass action model. The solutions of the differential equations are split into two parts, which are shown in Figure 4.32.

```
S0 = [0.5; 1; 0; 0]; % initial concentrations (nM)
Nu = [-1 1 1;-1 1 0;1 -1 -1;0 0 1]; % stoichiometric matrix
L = -D.*(D<0);
k = 60*[10 2 0.02]; % rate constants
M1s = ones(size(k));
V = @(t,S) Nu*(k.*prod(S(:,M1s).^L)).'; % rates dS/dt
sol = ode15s(V, [0 10], S0); % solution over t = 10 min
Si = deval(sol,ti); plot(60*ti,Si); % inner solution
So = deval(sol,to); plot(to,So); % outer solution
```



(a) (b)

Figure 4.32: Temporal evolution of substrate, enzyme, complex and product for the enzyme kinetic reaction. The plot on figure (a) shows the inner solution, the plot on figure (b) the outer solution.

The Matlab listing 4.3 computes the inner and outer solution of (4.44) for the enzyme kinetic reaction plotted in Figure 4.32. For most enzyme kinetic reactions, k_1 is usually magnitudes larger than k_2, which means that there is initially a rapid drop in the enzyme and substrate concentrations. It is for this reason that we have split the solution into an early phase or inner solution and a later phase, called outer solution. For most practical cases, it is very difficult to take measurements at very short time intervals. See [Rub75] for a comprehensive discussion.

4.8.2 Michaelis-Menten modelling

Phosphorylation steps in signaling cascades are enzyme kinetic reactions, the kinase facilitating the phosphorylation of a substrate. However, even for the relatively simple

system of an enzyme kinetic reaction

$$E + S \longleftrightarrow ES \to E + P$$

it is already rather difficult to obtain an analytical solution[5] to the set of differential equations (4.43). Furthermore, not all variables may be observable, i.e., measurable, or identifiable from experimental data. Here we are going to discuss commonly used assumptions and simplifications.

Above we realised that the total enzyme is constant, $[E] + [ES] = c_1$. Ignoring degradation and reconstitution of the enzyme, the constant can be evaluated from the initial conditions:

$$[E](t) + [ES](t) = [E](0) + [ES](0) \ . \tag{4.45}$$

Using (4.45) to eliminate $[E]$, we obtain

$$[\dot{S}] = -k_1\big([E](0) + [ES](0) - [ES]\big)[S] + k_2[ES] \tag{4.46}$$

$$[\dot{ES}] = k_1\big([E](0) + [ES](0) - [ES]\big)[S] - (k_2 + k_3)[ES] \ . \tag{4.47}$$

which, together with initial conditions $[S](0)$ and $[ES](0)$ can be numerically solved. The solution for $[P]$ can be derived easily from the solution of $[ES]$. The solution of $[E]$ is given by (4.45).

From our discussion in the previous section, we know that the substrate concentration $[S]$ steadily decreases. However, if it is the case that the available amount of substrate is relatively large, i.e., we might consider it as unchanged for a suitable period of time, we would have $[S](t) = [S](0)$, so that (4.46) is not required. For this *steady state assumption* (w.r.t. $[S]$), we are left with only (4.47)

$$[\dot{ES}] = k_1\big([E](0) + [ES](0) - [ES](t)\big)[S](0) - (k_2 + k_3)[ES](t) \ . \tag{4.48}$$

Denoting the steady state of $[ES]$ by the constant $[\tilde{ES}]$, and inserting this into (4.48)

$$k_1\big([E](0) + [ES](0) - [\tilde{ES}]\big)[S](0) - (k_2 + k_3)[\tilde{ES}] = 0 \ ,$$

from which we obtain an expression for $[\tilde{ES}]$

$$[\tilde{ES}] = \frac{\big([E](0) + [ES](0)\big)[S](0)}{K_\mathrm{M} + [S](0)} \ ,$$

where

$$K_\mathrm{M} = \frac{k_2 + k_3}{k_1} \tag{4.49}$$

is called the Michaelis- or Michaelis-Menten constant. Denoting the deviation of the complex from its steady state by $[ES]'$

$$[ES]' = [ES](t) - [\tilde{ES}]$$

[5] Schnell [SM97] describes a closed form solution employing the omega function.

and substituting this expression into (4.48)

$$[ES](t) = [ES]'(t) + [\tilde{ES}] \ , \tag{4.50}$$

and

$$\frac{d[ES]}{dt} = \frac{d[ES]'}{dt} \ .$$

It follows

$$\frac{d[ES]'}{dt} = -\lambda[ES]' \quad \text{where} \quad \lambda = k_1[S](0) + k_2 + k_3 \ ,$$

the solution of which is

$$[ES]'(t) = c \cdot e^{-\lambda t} \ ,$$

where c is a constant we obtain from initial conditions. From (4.50)

$$[ES](t) = c \cdot e^{-\lambda t} + [\tilde{ES}] \ .$$

The initial condition for $[ES]$ gives us then an expression for the constant c:

$$[ES](0) = c + [\tilde{ES}] \ , \quad \text{or} \quad c = [ES](0) - [\tilde{ES}] \ ,$$

leading to a solution for the temporal evolution of the complex concentration

$$[ES](t) = ([ES](0) - [\tilde{ES}]) \cdot e^{-\lambda t} + [\tilde{ES}] \ . \tag{4.51}$$

In conclusion, if substrate $[S]$ can be considered constant, then the complex concentration $[ES](t)$ approaches asymptotically the steady state $[\tilde{ES}]$, regardless of its initial conditions.

If we therefore let $[\dot{ES}] \approx 0$, this *quasi steady-state assumption* (w.r.t. $[ES]$) applied to (4.47), gives us the following expression

quasi steady-state

$$[ES](t) = \frac{(E[0] + [ES](0))\,[S](t)}{[S](t) + K_{\mathrm{M}}} \ .$$

Substituting this into (4.46) gives an expression that forms the basis for many experimental textbook settings:

$$[\dot{S}] = -\frac{k_3\,(E[0] + [ES](0))\,[S](t)}{[S](t) + K_{\mathrm{M}}}$$

The value $V = |[\dot{S}]|$ is called the (initial) velocity of the reaction. Assuming $[ES](0) = 0$, the equation above is commonly written as

$$V = \frac{k_3[E](0)[S]}{K_{\mathrm{M}} + [S]} \ .$$

Because $\partial V/\partial[S] > 0$, the reaction velocity is an increasing function of the substrate concentration. The maximum value of V, i.e., the maximum rate by which a product can be formed, is approached for very large values of $[S]$

$$V_{\max} = \lim_{[S]\to\infty} V = k_3[E](0)$$

and is called the *limiting rate*. Dividing this maximum rate by the enzyme concentration is called the *turnover number*. Typical turnover numbers are 1000 substrate molecules processed per second per enzyme molecule [AJL+02].

limiting rat

If the assumptions made above are realistic, the equation for V can be written as follows

$$V = \frac{V_{\max} \cdot [S]}{K_{\mathrm{M}} + [S]} \ . \tag{4.52}$$

The Michaelis-Menten constant K_{M} gives the initial substrate concentration at which the reaction velocity is half maximal (since for $[S](0) = K_{\mathrm{M}}$ substituted above gives $V = V_{\max}/2$). K_{M} is therefore an approximate measure of substrate affinity for the enzyme. A low K_{M} value means that the enzyme reaches its maximum catalytic rate at a low substrate concentration, which generally indices a tighter substrate binding.

Since V can be measured as a function of $[S](0)$, equation (4.52) allows us to estimate $V_{\max}$ and K_{M} from curve fitting. To this end, we rearrange (4.52)

$$\frac{1}{V} = \frac{1}{V_{\max}} \left(1 + \frac{K_m}{[S]} \right) \ .$$

This represents a straight line with slope $K_{\mathrm{M}}/V_{\max}$ and intercept $1/V_{\max}$. Plotting values of $1/V$ against $1/[S](0)$ is referred to the Lineweaver-Burk plot.

4.8.3 Multinomial systems

In our discussion of stochastic modeling and simulation, equation (1.30) represented GMA models. The general structure of these differential equations is of the form[6]

$$\dot{x}_i = \sum_{k=1}^{n} \theta_{ik} \prod_{j=1}^{n} x_j^{l_{ijk}} \ . \tag{4.53}$$

Including m independent variables, we write (see also page 94)

$$\dot{x}_i = \sum_{k=1}^{n+m} \theta_{ik} \prod_{j=1}^{n+m} x_j^{l_{ijk}} \ . \tag{4.54}$$

The mathematical structure was introduced by M.Peschel and W.Mende and is referred to as *multinomial systems*. Applied to biochemical reaction networks, these equations are the generalised mass action models (1.30). The set of equations for the enzyme kinetic reaction (4.41) is an example for a GMA system. We obtain the matrix of rate coefficients

$$[\theta_{ik}] = \begin{bmatrix} -k_1 & k_2 & k_3 & 0 \\ -k_1 & k_2 & 0 & 0 \\ k_1 & -k_2 & -k_3 & 0 \\ 0 & 0 & k_3 & 0 \end{bmatrix}$$

[6] Dealing with differential equations we use n not as the state-vector but as an integer and limit to sums. Similar, rather than referring to molecular species S we denote all variables with x, a notation that is most frequently used in systems theory.

and the set of matrices $L_i = [l_{jk}]_i$ for the kinetic orders l_{jk} in equation i:

$$L_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad L_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad L_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad L_4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

While the class of models, which is defined by (4.54) seems very general and it may seem difficult to conduct a formal analysis of these equations without turning to numerical solutions, we ought to remember that the values x_i cannot be negative and that although the coefficients can be positive or negative, which sign applies is predefined by the model structure or the reaction diagram respectively.

The ability to derive general properties of the dynamic behavior of a system, independent of specific parameter values, is the most attractive aspect of a formal mathematical analysis. If we consider the sparse data sets we obtain from experiments, such analysis would benefit parameter estimation and experimental design. If we are able to establish an order relationship between parameters, e.g. "$k_1 \gg k_3$", this would very useful in guiding parameter estimation, or providing confidence in using parameter values from literature. Due to experimental uncertainties, absolute values have little value, and an analysis in terms of basic temporal profiles (e.g. "x_1 decays exponentially", "x_2 peaks before $[P]$", "x_1 is pulled down", "x_2 is delayed") is at the heart of the biologists reasoning.

An example for a model of the ERK pathway is the following set of equations:

$$\begin{aligned}
\dot{x}_1 &= -k_1 x_1 x_2 + k_2 x_3 - k_{10} x_1 x_{12} + k_{11} x_{13} + k_6 x_7 + k_{14} x_{14} \\
\dot{x}_2 &= -k_1 x_1 x_2 + k_2 x_3 \\
\dot{x}_3 &= k_1 x_1 x_2 - (k_2 + k_3) x_3 \\
\dot{x}_4 &= k_3 x_3 \\
\dot{x}_5 &= k_3 x_3 - k_4 x_5 x_6 + k_5 x_7 \\
\dot{x}_6 &= -k_4 x_5 x_6 + k_5 x_7 + k_9 x_{10} \\
\dot{x}_7 &= k_4 x_5 x_6 - (k_5 + k_6) x_7 \\
\dot{x}_8 &= k_6 x_7 - k_7 x_8 x_9 + k_8 x_{10} \\
\dot{x}_9 &= k_{14} x_{14} - k_7 x_8 x_9 + k_8 x_{10} \\
\dot{x}_{10} &= k_7 x_8 x_9 - (k_8 + k_9) x_{10} \\
\dot{x}_{11} &= k_9 x_{10} - k_{12} x_{11} x_{13} + k_{13} x_{14} \\
\dot{x}_{12} &= -k_{10} x_1 x_{12} + k_{11} x_{13} + k_{15} x_{15} \\
\dot{x}_{13} &= k_{10} x_1 x_{12} - k_{11} x_{13} - k_{12} x_{11} x_{13} + k_{13} x_{14} \\
\dot{x}_{14} &= k_{12} x_{11} x_{13} - (k_{13} + k_{14}) x_{14} \\
\dot{x}_{15} &= k_{14} x_{14} - k_{15} x_{15} \ .
\end{aligned}$$

The structure of these equations is determined from knowledge of the proteins (x_i) involved. For some of these proteins we can obtained experimental time course data but

not for all. The question is then whether we could extract from this system of equation basic relationships between the k's, considering that we are looking for basic temporal profiles (all of which either converge to zero or some steady state)?

Since the structure of these equations is fairly well defined (sums of simple products of variables, the signs of terms are given, the parameters are always positive), one might imagine some 'qualitative' analysis of the kind described above: Given time course data for some of the variables, we first of all wish to validate the model structure (e.g. test for feedback through an additional negative term in one of the equations). The model structure is in this sense more important than knowing exact parameter values. We elaborate on these issues further in Chapter 4.

4.8.4 S-Systems

The mathematical structure (4.53) suggests a form in which we separate positive terms (complex formation, production) from negative terms (dissociation, degradation, depletion):

$$\dot{x}_i = V_i^+(x_1, \ldots, x_n) - V_i^-(x_1, \ldots, x_n) \ .$$

If we are to include m independent variables we write

$$\dot{x}_i = V_i^+(x_1, \ldots, x_n, x_{n+1}, \ldots, x_{n+m}) - V_i^-(x_1, \ldots, x_n, x_{n+1}, \ldots, x_{n+m}) \ . \tag{4.55}$$

where $i = 1, \ldots, n$. This general format allows for different classes of representations, one of which are S-systems.

The S-systems approach, mostly developed by M.Savageau [Sav69b, Sav69a, Sav70] and E.O.Voit [Voi00], starts with the general description (4.55)

$$\dot{x}_i = V_i^+ - V_i^- \quad \text{for} \quad i = 1, 2, \ldots, n \ ,$$

where the general functions are reduced to simple products of the variables involved. Such power law representation has some attractive mathematical features, but is implicitly based on a Taylor series approximation around a steady state value. This is alright for studying metabolic fluxes but does not work for transient phenomena in, for instance, signal transduction pathways.

For example, considering the conversion of x_1 into x_2, catalyzed by x_3. We assume a constant influx α to replenish x_1. The degradation of x_1 depends on the concentration or pool size of x_1 itself and also on the enzyme x_3:

$$\dot{x}_1 = \alpha - V_1^-(x_1, x_3) \ .$$

The production of x_2 is described in the same way as the degradation of x_1, $V_2^+ = V_1^-$. Finally, the degradation of x_2 depends only on its current concentration or pool size:

$$\dot{x}_2 = V_1^-(x_1, x_3) - V_2^-(x_2) \ .$$

The S-systems approach would then choose the following power-law representations for V_1^- and V_2^-:

$$V_1^-(x_1, x_3) = \beta x_1^a x_3^b \ ,$$
$$V_2^-(x_2) = \gamma x_2^c \ .$$

For a general S-system we write

$$\dot{x}_i = \alpha_i \prod_{j=1}^{n+m} x_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} x_j^{h_{ij}} \quad \text{for} \quad i = 1, 2, \ldots, n \ ,$$

and α denoting the rate constant for the production of each pool and β for its degradation or loss. These values can be positive or zero but not negative. There are numerous publications, exploring the theoretical properties of S-systems and applying them to metabolic pathways.

4.8.5 The Heinrich model

In [HNR02], Heinrich and colleagues demonstrated that even without experimental data to estimate parameter values, mathematical modelling allows an interesting study of protein kinase signal transduction pathways. Amongst other things they concluded from their study that:

- The simplest model pathways allow amplified signalling only at the expense of slow signal transduction.
- Phosphatases have a more pronounced effect than kinases on the rate and duration of signalling, whereas signal amplification is controlled primarily by kinases.

Heinrich first considers a linear signalling cascade with the stimulation of a receptor, forming the upper part of the pathway and subsequent sequential activation of downstream proteins through phosphorylation via kinases. The output of the pathway is the phosphorylation of a protein, which is assumed to have further consequences, e.g. on the activation of transcription factors and thereby influencing transcription of genes. Deactivation of proteins is realised through dephosphorylation via phosphatases in case of intermediate pathway elements. Receptors can thought of being deactivated by receptor dephosphorylation, internalisation of the receptor-ligand complex, and/or degradation of the receptor or ligand.

Considering the enzyme-kinetic reaction as a template for activation and deactivation, and assuming that the concentration of each kinase-substrate complex in the pathway is small compared to the total concentration of the reaction partners, and assuming that the concentration of active phosphatase is constant, each phosphorylation step or activation is described as a second-order or bi-molecular reaction, where the phosphorylated form x_{i-1} of protein $i-1$ one step up in the pathway takes the role of the kinase which facilitates the activation of the unphosphorylated form $\tilde{x}_i$ of the next downstream protein,

referred to as the substrate of the reaction.

$$\tilde{x}_i + x_{i-1} \xrightarrow{\tilde{\alpha}_i} x_i \qquad \text{: phosphorylation}$$
$$x_i \xrightarrow{\beta_i} \tilde{x}_i + x_{i-1} \qquad \text{: dephosphorylation}$$

The phosphorylation rate is given by the expression

$$v_{p,i} = \tilde{\alpha}_i x_{i-1} \tilde{x}_i \ ,$$

where $\tilde{\alpha}_i$ is the phosphorylation second-order rate constant for phosphorylation of the ith kinase. Assuming that the concentration of active phosphatase is constant, dephosphorylation is modelled as a first order reaction with dephosphorylation rate

$$v_{d,i} = \beta_i x_i \ ,$$

where β_i is the rate constant for dephosphorylation by the ith phosphatase. Defining

$$c_i = \tilde{x}_i + x_i$$

as the total concentration of protein kinase i. The differential equation for the phosphorylation or activation of x_i is given by

$$\begin{aligned} \frac{dx_i}{dt} &= v_{p,i} - v_{d,i} \\ &= \tilde{\alpha}_i x_{i-1} \tilde{x}_i - \beta_i x_i \end{aligned}$$

Let

$$\alpha_i = \tilde{\alpha}_i c_i$$

be a pseudo-first-order rate constant, so that we can write

$$\frac{dx_i}{dt} = \alpha_i x_{i-1} \left(1 - \frac{x_i}{c_i}\right) - \beta_i x_i \ , \tag{4.56}$$

which we also introduced in Section 4.7. The first step of the pathway, receptor stimulation, is modelled as

$$\frac{dx_1}{dt} = \alpha_1 u(t) \left(1 - \frac{x_1}{c_1}\right) - \beta_1 x_1 \ , \tag{4.57}$$

where $u(t)$ is the concentration profile of the activated receptor. For example, the inactivation of the receptor may be modelled as $u(t) = \exp(-\lambda t)$, where $1/\alpha$ is the time constant of the receptor. For $\lambda \to 0$, the pathway is permanently activated. Heinrich *et al.* then introduce the *signalling time* as the average time to activate protein i

signalling time

$$\tau_i = \frac{T_i}{l_i} \quad \text{where} \quad l_i = \int_0^\infty x_i(t)\ dt \quad \text{and} \quad T_i = \int_0^\infty t x_i(t)\ dt \ . \tag{4.58}$$

l_i denotes the total amount of active protein i, generated during the signaling period. The *signal duration* is defined by

signal duration

$$\theta_i = \sqrt{\frac{Q_i}{l_i} - \tau_i^2} \ , \quad \text{where} \quad Q_i = \int_0^\infty t^2 x_i(t) dt \ . \tag{4.59}$$

k activation

For a *weakly activated* pathway all of its proteins are phosphorylated to a low degree such that $x_i \ll c_i$. As a consequence, Equation (4.56) becomes

$$\frac{dx_i}{dt} = \alpha_i x_{i-1} - \beta_i x_i \ . \tag{4.60}$$

Heinrich *et al.* showed that the signaling time and signal duration can for this case be calculated explicitly:

$$\text{signalling time: } \tau = \frac{1}{\lambda} + \sum_{i=1}^{n} \frac{1}{\beta_i} \ , \tag{4.61}$$

$$\text{signal duration: } \theta = \sqrt{\frac{1}{\lambda^2} + \sum_{i=1}^{n} \frac{1}{\beta_i^2}} \ . \tag{4.62}$$

4.8.6 The MAP Kinase (MAPK) pathway

This section is to introduce an important class of signaling pathways. For unfamiliar biochemical expressions the reader is referred to the glossary on page 395.

input ⟶ MAPKKK ⟶ MAPKK ⟶ MAPK ⟶ output

Figure 4.33: Compact representation of the MAPK pathway.

The mitogen-activated protein kinase (MAPK) cascade, is part of the growth-factor/Ras pathway in eucaryotic cells. The cascade[7] is highly *conserved*, which means that the same principles can be observed in a variety of organisms and cell types.

The core of this pathway is formed by a module which is defined by three protein kinases: MAPKKK (e.g. RAS/Raf), MAPKK (e.g. MEK) and MAPK. This module is activated by a collection of proteins, some of which have to occur in combination. The first element of the module to be activated is the MAPKKK. The activated MAPKKK phosphorylates MAPKK at two sites. This double phosphorylated MAPKK, denoted MAPKK**, acts as a threonine/tyrosine kinase and phosphorylates MAPK at two sites of the protein structure. MAPK can then act as a kinase for transcription factors, but may also have a feedback effect on the activity of kinases like the MAPKKK further *upstream*.

The ERK (extracellular-signal-regulated kinase) pathway is an example for a MAPK cascade, which features Ras as the G-protein, Raf as MAPKKK, MEK (MAPK/ERK kinase) as MAPKK and ERK as MAPK. Ras and Raf are proto-oncogenes which explains

[7] Biologists refer to an unbranched sequence of modules combined in a cascade as a *linear* cascade. The term 'linear' has not relationship with the question whether the biochemical reactions and their mathematical model are linear or nonlinear. See page 27 for a definition a linear model. The upper - membrane near parts of the pathway are also referred to 'upstream regions'.

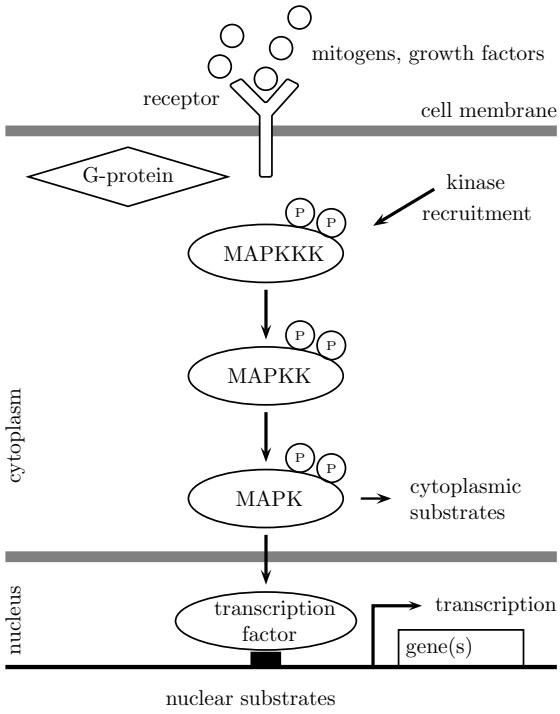


Figure 4.34: The MAP kinase pathway. The core of this pathway is formed by a module which is defined by three protein kinases: MAPKKK, MAPKK and MAPK. Drawing adapted from [Kol00].

the interest in this pathway [Kol00]. Most stimuli to the receptor leads to an activation of the G-protein Ras by inducing the exchange of GDP with GTP. GDP and GTP are therefore also referred to as *exchange factors*. This exchange will convert the Ras molecule into its active conformation. Ras resides near the cell membrane and one says that the exchange factors are *recruited*. The protein SOS (son of sevenless) is another Ras exchange factor, which can terminate Ras activation. This termination proceeds by phosphorylation of SOS, which leads to the disassembly of the complex. The phosphorylation of SOS is also feedback regulated by the activated ERK pathway [Kol00].

Activated Ras functions as an adaptor that binds to Raf kinases with high affinity and causes their translocation to the cell membrane, where Raf activation takes place.[8] Mammals posses three Raf proteins which are also referred to as *isoforms*: Raf-2, A-Raf, and B-Raf. For all three, MEK acts as a substrate further 'downstream' of the pathway.

MEK is activated by phosphorylation of two serine residues in the activation loop. The most predominant activator of MEK in most cell types are Raf kinases. It is believed that Raf-1 can inhibit itself through some negative feedback. Raf seems to be suspended in a balance between activation and auto-inhibition [Kol00]. Raf can activate both MEK

[8]The spatial dimension, translocation of molecules, is an important aspect that is ignored by conventional models. One idea is to divide a space up into regions, model each region separately and allow for an exchange between them. We are thus in need for a multi-model concept.

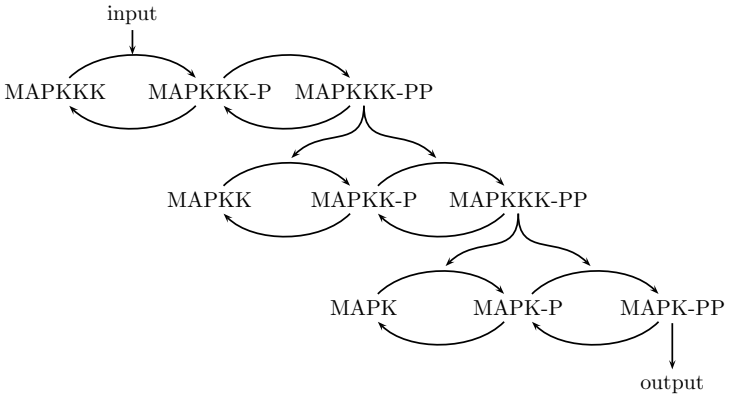


Figure 4.35: General outline of the MAPK pathway.

isoforms, MEK-1 and MEK-2, and both of them can activate the downstream ERK kinases. MEK is a 'dual-specificity' kinase which can phosphorylate both.

Finally, ERK is a serine/threosine kinase with more than 50 substrates [Kol00]. All components of the Ras/Raf/MEK/ERK pathway can interact with each other physically: Ras-GTP binds to Raf; Raf can bind to MEK; and MEK can bind to ERK.

4.8.7 The Ras/Raf/MEK/ERK pathway

signal transduction

This section describes a mathematical model of the ERK module, as an example of a MAPK signal transduction pathway. The model is represented by a set of nonlinear differential equations. We show how this representation can be generalised to capture a large class of dynamic pathway models. In this framework, a pathway diagram corresponds to the state space of a dynamic system, while the entirety of dynamic processes that can occur in a particular pathway is defined by a one-parameter group of transformations in the manifold that is the state space. We are thus providing a conceptual framework in which to describe not only pathway diagrams but also the spatial-temporal interactions within and between cells.

Experimental data show that the inhibition of MEK phosphorylation by RKIP is not linear. There is a threshold of RKIP expression that steeply reduces MEK phosphorylation. This is consistent with a positive feedback mechanism. If not all Raf is bound to RKIP, then there is Raf-1 available for interacting with and activating MEK. MEK then activates ERK. Consequently, the positive feedback phosphorylation of RKIP by ERK will ensure that RKIP is phosphorylated and dissociates from Raf-1. In this situation there will be little or no inhibition of Raf-1 by RKIP. If the level of RKIP expression exceeds a certain threshold, all of the Raf-1 will be bound to RKIP. In this situation there is no phosphorylation of RKIP and no dissociation. Hence MEK phosphorylation remains inhibited.

To analyze the dynamics of the ERK signaling pathway, including the positive feedback mechanism, in both qualitative and quantitative manner, we constructed a mathematical model based on the mass action law and represented by nonlinear ordinary differential

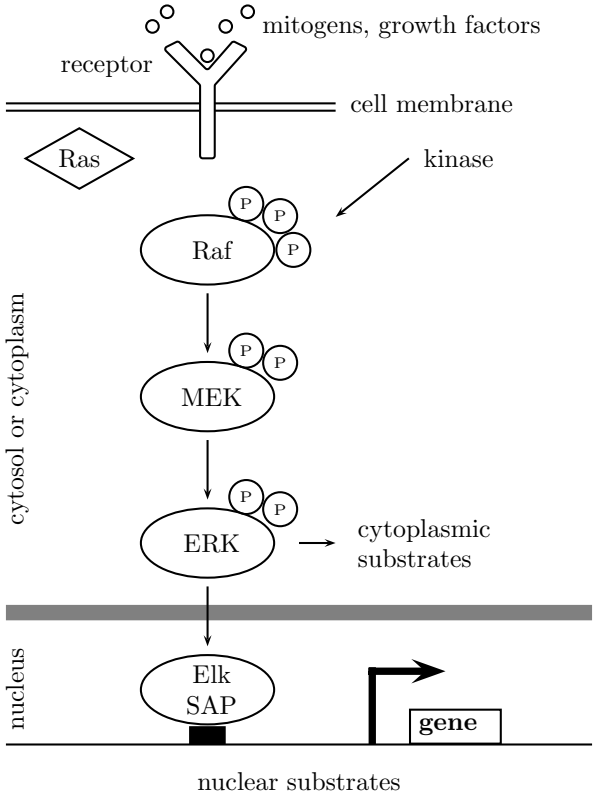


Figure 4.36: Basic outline of the Ras/Raf/MEK/ERK pathway. Drawing adapted from [Kol00].

equations (see Figure 4.39):

$$
\left.
\begin{aligned}
\dot{x}_1 &= -k_1 x_1 x_2 - k_9 x_1 x_{12} + k_{10} x_{13} + k_5 x_7 + k_{13} x_{14} \\
\dot{x}_2 &= -k_1 x_1 x_2 \\
\dot{x}_3 &= k_1 x_1 x_2 - k_2 x_3 \\
\dot{x}_4 &= k_2 x_3 \\
\dot{x}_5 &= k_2 x_3 - k_3 x_5 x_6 + k_4 x_7 \\
\dot{x}_6 &= -k_3 x_5 x_6 + k_4 x_7 + k_8 x_{10} \\
\dot{x}_7 &= k_3 x_5 x_6 - (k_4 + k_5) x_7 \\
\dot{x}_8 &= k_5 x_7 - k_6 x_8 x_9 + k_7 x_{10} \\
\dot{x}_9 &= k_{13} x_{14} - k_6 x_8 x_9 + k_7 x_{10} \\
\dot{x}_{10} &= k_6 x_8 x_9 - (k_7 + k_8) x_{10} \\
\dot{x}_{11} &= k_8 x_{10} - k_{11} x_{11} x_{13} + k_{12} x_{14} \\
\dot{x}_{12} &= -k_9 x_1 x_{12} + k_{10} x_{13} + k_{14} x_{15} \\
\dot{x}_{13} &= k_9 x_1 x_{12} - k_{10} x_{13} - k_{11} x_{11} x_{13} + k_{12} x_{14} \\
\dot{x}_{14} &= k_{11} x_{11} x_{13} - (k_{12} + k_{13}) x_{14} \\
\dot{x}_{15} &= k_{13} x_{14} - k_{14} x_{15} \ .
\end{aligned}
\right\} \tag{4.63}
$$

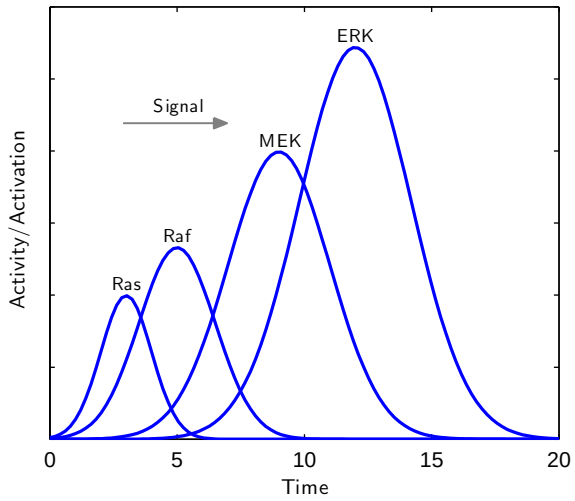


Figure 4.37: Illustration of how a signal is relayed through sequential activation of proteins in the Ras/Raf/MEK/ERK pathway. Note that this picture is an idealised picture and actual amplitudes and base levels may differ significantly when considering experimental data. See also Figure 4.36.

The mathematical structure is one of multinomial systems, concentrations x_i as well as the values for parameters k_i can only be positive. Each step in the pathway is modelled in analogy to an enzyme kinetic reaction (4.41). Possible algebraic simplifications of the model and parameter estimation are not discussed here as this is not the focus of the present section.

The Ras/Raf-1/MEK/ERK module is an ubiquitously expressed signaling pathway that conveys mitogenic and differential signals from the cell membrane to the nucleus [YJM+00]-[Kol00]. This kinase cascade appears to be spatially organised in a signaling complex nucleated by Ras proteins. The small G protein Ras is activated by many growth factor receptors and binds to the Raf-1 kinase with high affinity when activated.

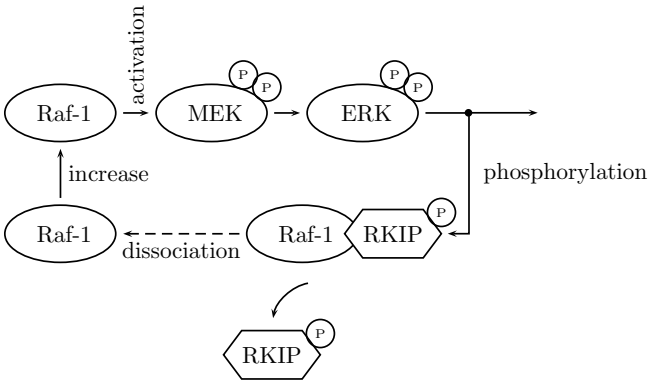


Figure 4.38: Illustration of the interactions of RKIP with the Ras/Raf/MEK/ERK pathway.

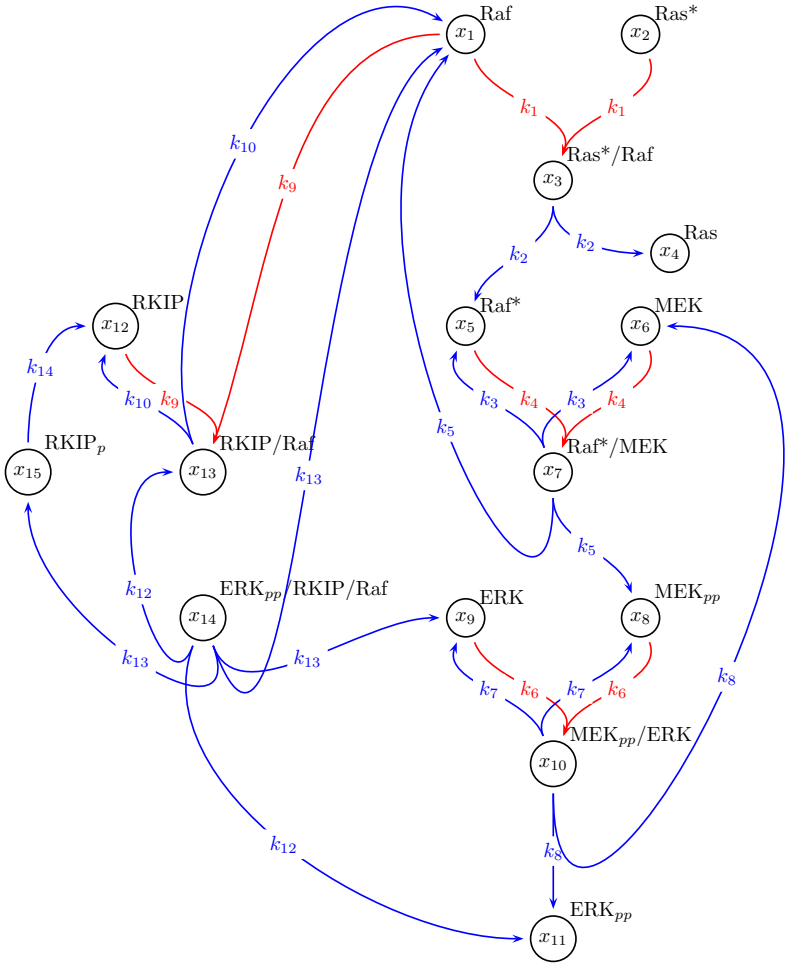


Figure 4.39: Graphical representation of the model for Ras/Raf-1/MEK/ERK signaling module, describing a positive feedback loop between RKIP and double phosphorylated ERK (ERK_{pp}).

This induces the recruitment of Raf-1 from the cytosol to the cell membrane. Activated Raf-1 then phosphorylates and activates MAPK/ERK kinase (MEK), a kinase that in turn phosphorylates and activates Extracellular Kinase (ERK), the prototypic Mitogen-Activated Protein Kinase (MAPK). Activated ERKs can translocate to the nucleus and regulate gene expression by phosphorylation of transcription factors. This kinase cascade controls the proliferation and differentiation of different cell types. The specific biological effects of the kinase cascade are crucially regulated by the Raf-1 kinase inhibitor protein (RKIP) [YRD+01]. RKIP binds to Raf-1 thereby disrupting the interaction between Raf-1 and MEK. As a consequence RKIP inhibits the phosphorylation and activation of MEK by Raf-1. RKIP overexpression interferes with the activation of MEK and ERK, induction of AP-1 dependent receptor genes and transformation elicited by an oncogenically activated Raf-1 kinase [YSL+99].

Figure 4.40 shows a simulation of the ERK model for varying initial concentrations

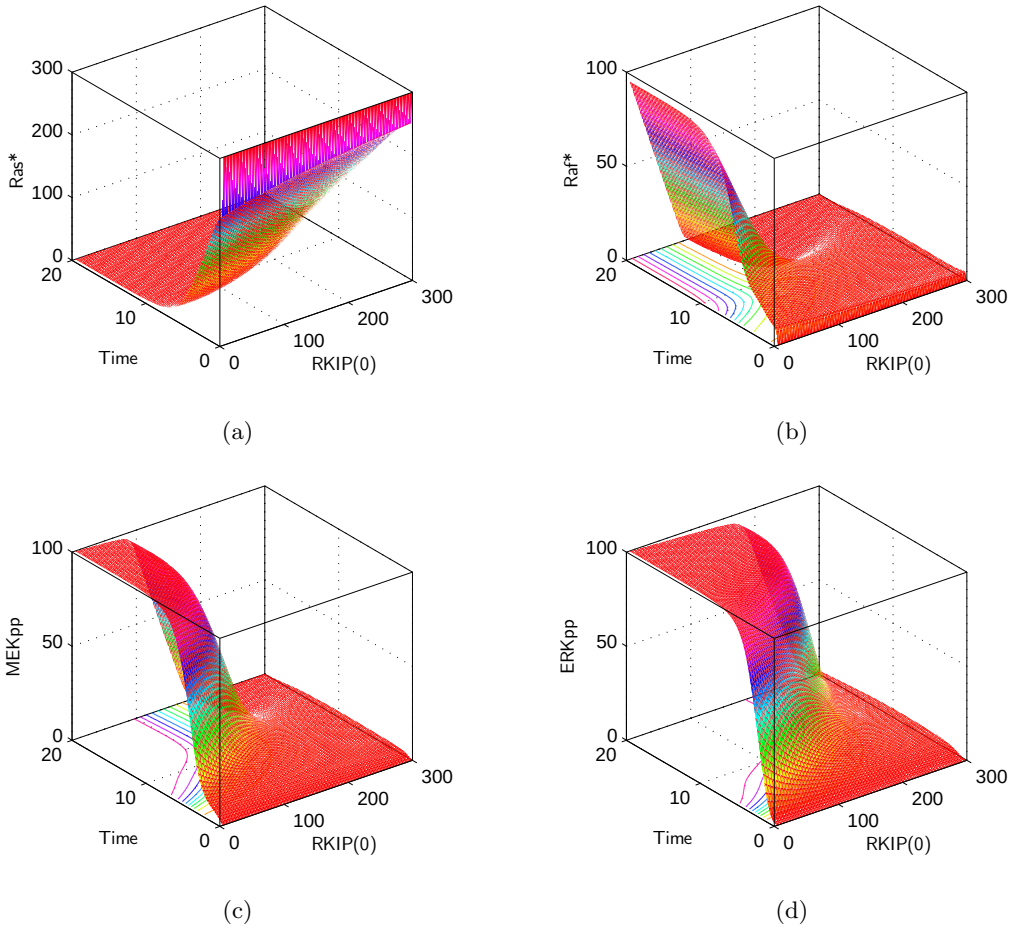


Figure 4.40: Simulation of the ERK model for varying initial concentrations of RKIP.

of the Raf-1 kinase inhibitor protein RKIP. The simulations show the rate of active Ras (x_2) binding to Raf-1 (x_1) linearly decreasing along with the initial value of RKIP. The plots demonstrate how the initial signal transduction through active Ras is interrupted by RKIP. The variation profile of active Raf (x_5) as a function of variations of the initial value for RKIP is similar to active MEK (x_8) and active ERK (x_{11}). The dynamics of these proteins in Figure 4.40 also exhibit the nonlinear relationships encapsulated by the model. At low initial concentration of RKIP all signal proteins are completely activated although with different time lags. At high concentrations of RKIP, the activation ratio is about zero. These simulation results show that there is a threshold of concentration of RKIP that steeply reduces the phosphorylation of each protein, which is in accordance with experimental data.

The concept of a pathway is the framework in which a molecular- or cell biologist captures her/his knowledge of pathway variables, their states and relationships. A pathway model - whether a simple diagram or a mathematical representation like the one

described here is an abstraction. In [RS02] the authors discuss the importance of abstractions in science and suggest an abstract computer language like π-calculus for pathway modelling. This language provides a means to formalise the knowledge of components and their interactions. In order to simulate a model in this framework, the process algebra requires the use of, for instance, the Gillespie algorithm in order to compile the model. Here we are seeking an abstract (and thus generally applicable) algebraic framework in which to discuss the dynamic properties of a pathway but also the entirety of dynamic processes the cell can realise as well as relationships between cells. Instead of a computer language, we begin with chemical rate equations and work 'upwards' through generalisations of the models that can be constructed to represent the dynamic interactions in pathways. The main argument for a π-calculus is its "linguistic structure" and "operational semantics" from which causal relationships can be derived. The motto of the approach presented in this section could be that we wish to realise what the biologist *could see* rather than what he *says*.

4.8.8 Feedback and oscillations in signalling pathways

In Figure 4.34 a prototypical mitogen-activated protein kinase (MAPK) cascade is shown [FJ96, Kol00]. MAP kinase pathways have been found in many organisms and cell types. They are also an important system for cancer research studies [Kol00, YSL+99, YJM+00] and have been considered in various modelling exercises (e.g. [FJ96, HFJ96, BI99, Kho00, BF00, AL01, K+02]).

Let us model phosphorylation and dephosphorylation as a reversible bimolecular reaction, where the phosphorylation of X into X^* is facilitated by the kinase U, and dephosphorylation by phosphatase P:

$$X + U \xrightarrow{k_1} X^* + U \ ,$$
$$X^* + P \xrightarrow{k_2'} X + P \ .$$

We write $\tilde{x}$ for the unphosphorylated form of a molecular species or protein X, u for the kinase U, p for phosphatase P and x to denote the activated, i.e., phosphorylated protein X^*. Referring to the law of mass action, we obtain the following set of ordinary differential equations

$$\begin{aligned} \frac{d}{dt}x &= k_1 u(t)\tilde{x}(t) - k_2' p(t)x(t) && : \text{phosphorylation}, \\ \frac{d}{dt}\tilde{x} &= -k_1 u(t)\tilde{x}(t) + k_2' p(t)x(t) && : \text{dephosphorylation} \ . \end{aligned}$$

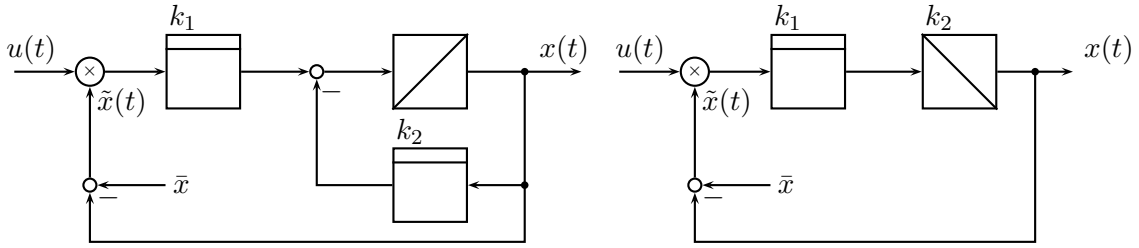
To simplify the mathematical model we assume that the phosphatase is constant. This means we can merge p and k_2' into k_2. Together with the conservation for a constant total of x,

$$\bar{x} = x(t) + \tilde{x}(t) \ ,$$

we obtain the following differential equation, in one variable, to describe the phosphorylation of protein X:

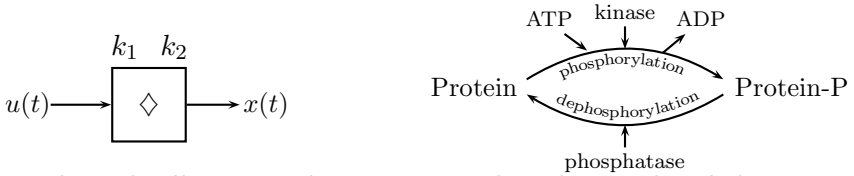
$$\frac{d}{dt}x = k_1 u(t)\big(\bar{x} - x(t)\big) - k_2 x(t) \ .$$

The block diagram for a signaling step, (de-)phosphorylation is readily obtained:



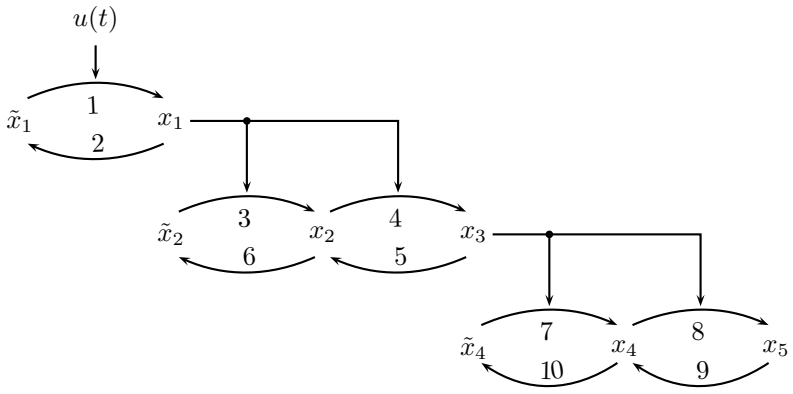
In this model for (de-)phosphorylation, it is assumed that kinase-substrate concentrations are low compared to the total concentration of the reactant species. Furthermore, for dephosphorylation to be considered a first-order reaction, it is assumed that the active phosphatase concentration is constant.

For more complex systems, let us now collapse the previous diagram into a single block to represent the (de-)phosphorylation of $\tilde{x}$ through kinase u:



The diagram on the right illustrates the conventional, and more detailed, representation. Phosphorylation is facilitated by means of a kinase and dephosphorylation is realised by a phosphatase. The phosphorylated state is commonly denoted by adding -P to the name of the protein.

Given such a module, we are now in a position to construct more complex pathway diagrams by connecting these basic components into cascades and networks. For example, for the pathway in Figure 4.34 the map is [Kho00]:



In the diagram x_1 corresponds to activated MKKK-P, x_2 to MKK-P, x_3 to the

double-phosphorylated MKK-PP, x_4 to MAPK-P, x_5 to MAPK-PP. Inactivated forms are denoted with a tilde $\tilde{x}$. In signalling activation/inactivation of proteins corresponds to phosphorylation/dephosphorylation, while in some cases one considers double-phosphorylations:

$$\begin{aligned} v_1 &= k_1 u(t)\tilde{x}_1(t), & v_6 &= k_6 x_2(t) \\ v_2 &= k_2 x_1(t), & v_7 &= k_7 x_3(t)\tilde{x}_4(t) \\ v_3 &= k_3 x_1(t)\tilde{x}_2(t), & v_8 &= k_8 x_3(t) x_4(t) \\ v_4 &= k_4 x_1(t) x_2(t), & v_9 &= k_9 x_5(t) \\ v_5 &= k_5 x_3(t), & v_{10} &= k_{10} x_4(t) \end{aligned}$$

The model is the derived from the fact that

$$\begin{aligned} \dot{x}_1 &= v_1 - v_2 & \dot{x}_2 &= v_3 - v_6 - v_4 + v_5 \\ \dot{x}_3 &= v_4 - v_5 & \dot{x}_4 &= v_7 - v_{10} + v_9 - v_8 \\ \dot{x}_5 &= v_8 - v_9 \end{aligned}$$

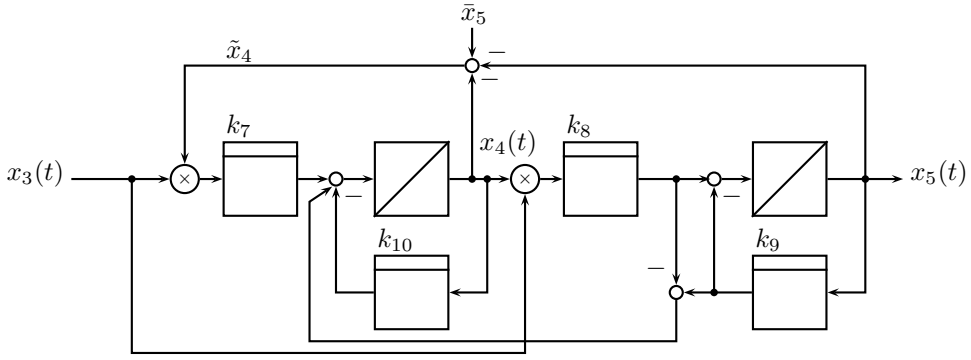
Inserting the reaction rates into these equations we obtain the following set of equations for the activated proteins in the pathway:

$$\begin{aligned} \dot{x}_1 &= k_1 u(t)\tilde{x}_1(t) - k_2 x_1(t) \\ \dot{x}_2 &= k_3 x_1(t)\tilde{x}_2(t) - k_6 x_2(t) \underbrace{-k_4 x_1(t) x_2(t) + k_5 x_3(t)}_{-\dot{x}_3} \\ \dot{x}_3 &= k_4 x_1(t) x_2(t) - k_5 x_3(t) \\ \dot{x}_4 &= k_7 x_3(t)\tilde{x}_4(t) \underbrace{+k_9 x_5(t) - k_8 x_3(t) x_4(t)}_{-\dot{x}_5} - k_{10} x_4(t) \\ \dot{x}_5 &= k_8 x_3(t) x_4(t) - k_9 x_5(t) \ . \end{aligned}$$

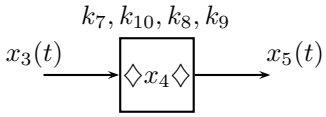
In addition, the following conservation relationships hold:

$$\begin{aligned} \bar{x}_1 &= \tilde{x}_1(t) + x_1(t) \\ \bar{x}_3 &= \tilde{x}_3(t) + x_2(t) + x_3(t) \\ \bar{x}_5 &= \tilde{x}_4(t) + x_4(t) + x_5(t) \end{aligned}$$

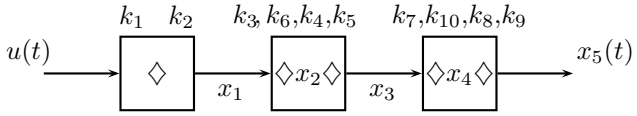
Using the mathematical model for (de-)phosphorylation from above, the double phosphorylation of x_4 into x_5 is described by the following block diagram:



This again we can collapse into a single block, without loss of information:



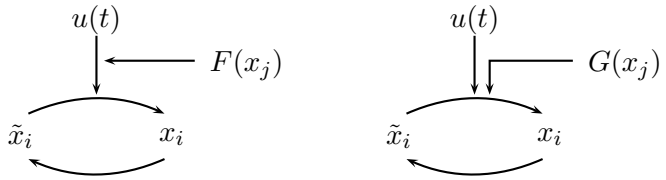
The MAP kinase pathway can then be represented by the compact block diagram:



This system is then another example of the state-space representation, where for $u(t)$ we might assume a negative exponential function $u(t) = e^{-\lambda t}$, where an initial concentration of ligands is depleted through binding to the receptors on the cell surface. The greater the value of λ, the faster the ligands bind with receptors to form complexes. In [HNR02], these series-connected submodels of (de-)phosphorylation, have been used to analyze pathways for their dynamic properties. In particular the authors derived expression for the *signaling time*, defined as the average time to activate a protein in the pathway and the *signal duration*, characterised by an integral of the concentration profile. In [Kho00] a very similar model as the one above is modified by introducing a negative feedback loop between the end product MAPK-PP and a Ras/MAPKKKK complex at the top of the pathway. Kholodenko showed how ultrasensitivity, leading to switch-like behavior, combined with negative feedback can bring about sustained oscillations in this pathway. Considering populations of cells, this may be of particular interest in the context of the synchronisation of coupled oscillators, which has been observed in a range of biological and physical systems [Str00b]. Approaches to test for feedback loops have been presented in [K+02, AFJS04]. In [BI99, BF00, AL01] computational studies of feedback effects on signal dynamics in a detailed MAPK pathway model are presented.

4.8.8.1 Feedback in signalling pathways

We continue the previous section, introducing feedback loops from a protein x_j further down the pathway, up to x_i. We have two options indicated in the following diagrams:



On the left-hand side, feedback manifests itself as a multiplicative term in the differential equation for x_i:

$$\frac{d}{dt}x_i(t) = k_1 u(t) F(x_j)\big(\bar{x}_i - x_i(t)\big) - k_2 x_i(t) \ ,$$

where for the function $F(x_j)$ we can choose from the following commonly used versions:

$$F(x_j) = \frac{1}{1 + \left(\frac{x_j}{K_I}\right)^n} \ , \qquad F(x_j) = \frac{c}{d^n + x_j^n} \ , \qquad F(x_j) = \frac{c x_j^n}{d^n + x_j^n} \ ,$$

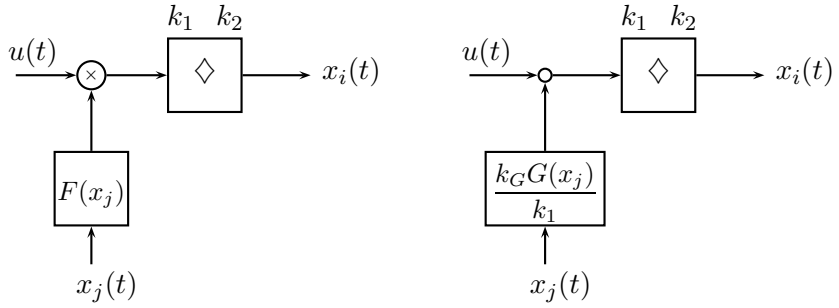
where $n \geq 1$ defines the steepness of the feedback function and the subscript I of K_I stands for 'inhibition'. The main requirement for the choice of a function $F(x_j)$ is that at $x_j = 0$, we should have $F(x_j) = 1$. Mechanistic interpretations and experimental evidence for these functions are discussed in [Gol96, FJ96, HFJ96, LQ03]. Note the distinction between a *mechanistic* (or physical) and a *operational* (or phenomenological) definition for an interaction. An operational definition is based on observations, not necessarily requiring an interpretation/understanding of the physical interactions of the molecules involved, as would be the case for a mechanistic definition of kinetic behavior [CB04, pg. 116]. For the feedback indicated on the right-hand side, and represented by $G(x_j)$, there is an additional contribution to the activation of X:

$$X + U \xrightarrow{k_1} X^* + U \ , \qquad X + G \xrightarrow{k_G} X^* + G \ ,$$

leading to the following, modified ODE model:

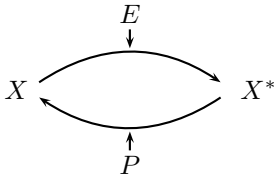
$$\begin{aligned}\frac{d}{dt}x(t) &= k_1 u(t)\big(\bar{x} - x(t)\big) + k_G G(x_j)\big(\bar{x} - x(t)\big) - k_2 x(t) \\ &= k_1\big(\bar{x} - x(t)\big)\left[u(t) + \frac{k_G G(x_j)}{k_1}\right] - k_2 x(t) \ .\end{aligned}$$

If $G(x_j)$ is monotonically increasing as x_j increases, we are dealing with positive feedback and, vice versa, if $G(x_j)$ monotonically decreasing with x_j, we are dealing with negative feedback. While for conventional pathway maps the kind of feedback employed remains unclear, if not explicitly stated. In the block diagram , however in our scheme we recognise the two situations as follows:



4.8.8.2 Michaelis-Menten kinetics

The mass-action models for (de-)phosphorylation, as introduced above, could be criticised in that if the activation of a protein X is seen as an enzyme catalyzed reaction, the rate of activation in experiments is limited. Some authors might argue that a more realistic model would be to consider Michaelis-Menten kinetics. We here show how the Michaelis-Menten model can be derived from the mass-action model. In the context of signalling pathways an argument against Michaelis-Menten kinetics is the fact that in its derivation, the enzyme concentrations should be much smaller than the substrate concentration. In the context of cell signalling the relationships between kinase concentrations and those of inactivated proteins may however not satisfy this assumption. Let us consider the activation (phosphorylation) of protein X by means of an enzyme (kinase) E:



The dephosphorylation of the activated protein X^* is realised by some phosphatase P, which we here assume to have a constant concentration. The mass-action model of these two processes is

$$X + E \xrightarrow{k_1} X^* + E \ , \qquad X^* + P \xrightarrow{k_2'} X + P \ .$$

Using again the previously introduced notation, the corresponding differential equation for activation is

$$\frac{d}{dt}x = k_1 e(t)\tilde{x}(t) - k_2' p(t) x(t) \ .$$

Assuming a constant phosphatase P, let us redefine $k_2 \doteq k_2' p(t)$. With the conservation relation $\tilde{x}(t) + x(t) = \bar{x}$, we thus have

$$\frac{d}{dt}x = k_1 e(t)\big(\bar{x} - x(t)\big) - k_2 x(t) \ .$$

Note that only one differential equation is needed since changes in the non-phosphorylated form can always be derived from the conservation relation. A Michaelis-Menten model of

a signalling module considers an intermediate complex in the enzyme catalyzed reaction

$$X + E \underset{d_1}{\overset{a_1}{\leftrightarrows}} XE \xrightarrow{k_1} X^* + E \ , \qquad X^* + P \underset{d_2}{\overset{a_2}{\leftrightarrows}} X^*P \xrightarrow{k_2} X + P \ .$$

It is assumed that the first forward reaction is very fast, i.e., for phosphorylation $a_1 \gg d_1$. Focussing on steady-states, one finds for the Michaelis-Menten constants

$$K_{m1} = \frac{d_1 + k_1}{a_1} \ , \qquad K_{m2} = \frac{d_2 + k_2}{a_2} \ .$$

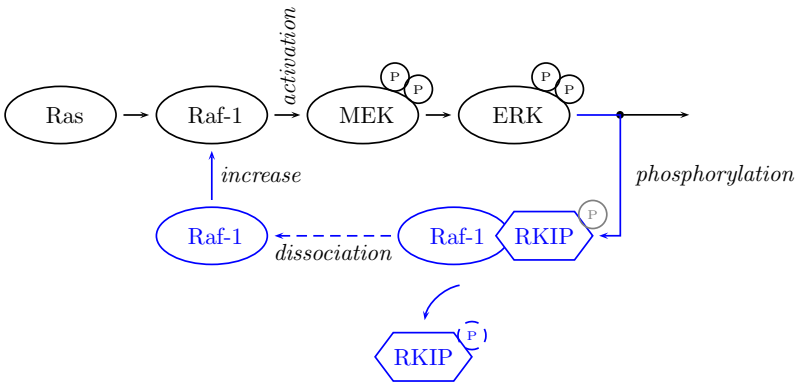
The equations for a Michaelis-Menten model are then

$$\frac{d}{dt}x = \frac{k_1 e(t)\big(\bar{x} - x(t)\big)}{K_{m1} + \bar{x} - x(t)} - \frac{k_2 x}{K_{m2} + \bar{x} - x(t)} \ .$$

If E can be assumed to be constant, one would frequently define $V_{\max} \doteq k_1 E$ as the maximum velocity of the reaction, that is achieved for large concentrations of the non-activated form $x = \bar{x} - x$. Michaelis-Menten models of the MAP-kinase signaling cascade are considered in [HFJ96, Kho00, K+02].

4.8.8.3 The Ras/Raf/MEK/ERK pathway

In this section we work out an example for the MAP-kinase pathway in Figure 4.34. In the Ras/Raf/MEK/ERK pathway Ras is the G-protein, Raf the MAPKKK, MEK the MAPKK and ERK the MAPK [Kol00]. While the linear cascade of Figure 4.34 is a textbook illustration, the research literature suggests the existence of various feedback loops such that we are dealing with a network rather than a linear cascade. As an example, for the Ras/Raf/MEK/ERK pathway a positive feedback mechanism can be illustrate with the following cartoon [YSL+99, YJM+00]:



The MAP kinase module is realised by the sequential activation of Raf-1, upstream near the cell membrane, followed by activation of the proteins MEK and ERK through structural modifications in the form of phosphorylations indicated by the P's. ERK translocates into the nucleus of the cell, where it effects the transcription of genes. Double phosphorylated ERK-PP also phosphorylates RKIP and thereby releases Raf

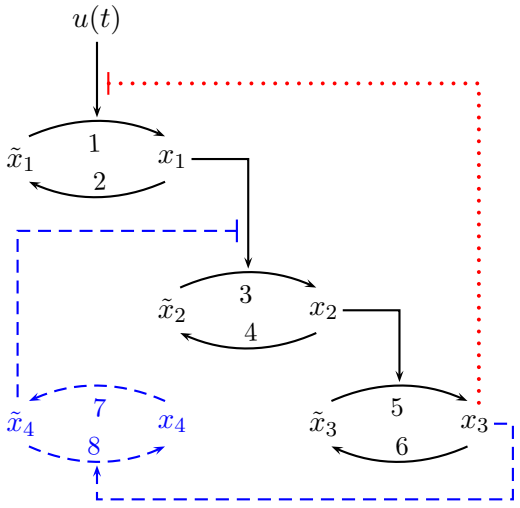


Figure 4.41: Pathway map for the Ras/Raf/MEK/ERK signal transduction pathway with two speculated feedback loops. The dashed parts describe a positive feedback loop. The bar at the end of a line denotes an inhibition. Note that x_3 is acting on the phosphorylation of $\tilde{x}_4$. The dotted line describes a negative feedback loop. The variables are $x_1 \doteq$ Raf, $x_2 \doteq$ MEK, $x_3 \doteq$ ERK, $x_4 \doteq$ RKIP. $\tilde{x}$ denotes the non-activated or non-phosphorylated form.

from the Raf-1/RKIP complex, and Raf in turn activates MEK. This positive feedback loop leads to switch-like behavior of the pathway.

We first translate the cartoon into a pathway map to reduce ambiguity. The pathway map is shown in Figure 4.41. The variables are $x_1 \doteq$ Raf, $x_2 \doteq$ MEK, $x_3 \doteq$ ERK, $x_4 \doteq$ RKIP. To simplify the example we ignored the double-phosphorylations of the previous Section. We first consider the pathway without any feedback loop. The mathematical of model of this simple three module cascade is specified by the following set of equations.

$$\begin{aligned}
\frac{d}{dt}x_1 &= \underbrace{\frac{k_1 u(t)(\bar{x}_1 - x_1(t))}{K_{m1} + (\bar{x}_1 - x_1(t))}}_{\text{phosphorylation}} - \underbrace{\frac{k_2 x_1(t)}{K_{m2} + x_1(t)}}_{\text{dephosphorylation}} \\
\frac{d}{dt}x_2 &= \frac{k_3 x_1(t)(\bar{x}_2 - x_2(t))}{K_{m3} + (\bar{x}_2 - x_2(t))} - \frac{k_4 x_2(t)}{K_{m4} + x_2(t)} \\
\frac{d}{dt}x_3 &= \frac{k_5 x_2(t)(\bar{x}_3 - x_3(t))}{K_{m5} + (\bar{x}_3 - x_3(t))} - \frac{k_6 x_3(t)}{K_{m6} + x_3(t)} \ ,
\end{aligned}$$

where for the conservation relations $\bar{x}_1 = \tilde{x}_1(t) + x_1(t)$, $\bar{x}_2 = \tilde{x}_2(t) + x_2(t)$, $\bar{x}_3 = \tilde{x}_3(t) + x_3(t)$ hold. Next we consider the positive feedback loop introduced by RKIP and which is denoted by x_4. First phosphorylation and dephosphorylation are described as before,

$$\frac{d}{dt}x_4 = \frac{k_7 x_3(t)(\bar{x}_4 - x_4(t))}{K_{m7} + (\bar{x}_4 - x_4(t))} - \frac{k_8 x_4(t)}{K_{m8} + x_4(t)} \ ,$$

where $\bar{x}_4 = \tilde{x}_4(t) + x_4(t)$. Note that x_3, activated ERK-PP, is acting on the phosphorylation of x_4 (RKIP). The inhibitory effect of RKIP on the phosphorylation of x_2 (MEK)

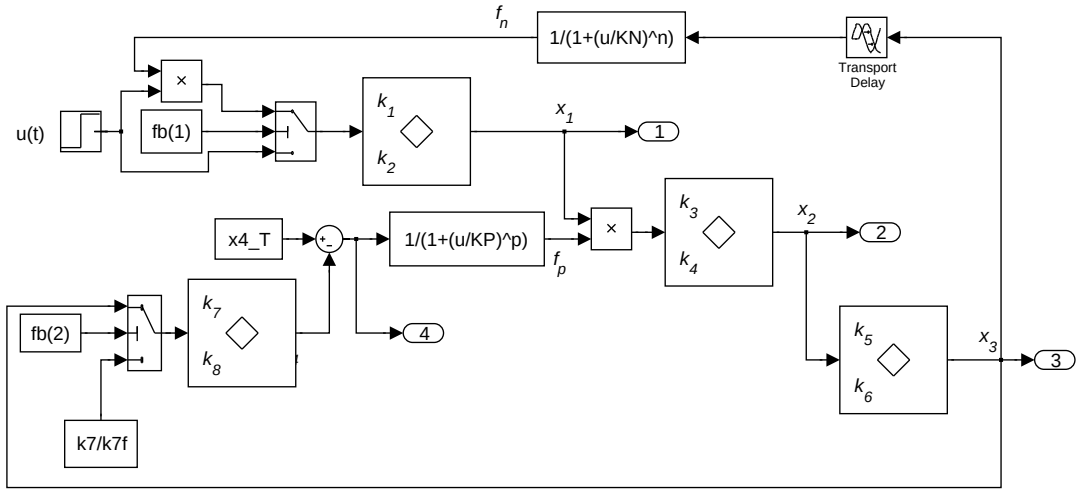


Figure 4.42: Simulink block diagram of the Ras/Raf/MEK/ERK signal transduction pathway with RKIP regulation. Individual blocks can be "unmasked" to reveal their inside. The inside elements of these blocks are identical to those introduced before. The switches in the diagram are used to introduce or remove feedback loops in the simulation.

is reflected by a change to the rate equation of x_2:

$$\frac{d}{dt}x_2 = \frac{k_3 x_1(t)\left[1/\left(1+\left[\frac{\bar{x}_4 - x_4(t)}{K_P}\right]^p\right)\right](\bar{x}_2 - x_2(t))}{K_{m3} + (\bar{x}_2 - x_2(t))} - \frac{k_4 x_2(t)}{K_{m4} + x_2(t)} \ ,$$

where K_P is a constant that defines the strength of the feedback and n defines the steepness of the response curve. The negative feedback from x_3 (ERK-PP) to x_1 (Raf) leads to an insertion in the equation for dx_1/dt:

$$\frac{d}{dt}x_1 = \frac{k_1 u(t)\left[1/\left(1+\left[\frac{x_3(t)}{K_N}\right]^n\right)\right](\bar{x}_1 - x_1(t))}{K_{m1} + (\bar{x}_1 - x_1(t))} - \frac{k_2 x_1(t)}{K_{m2} + x_1(t)} \ .$$

For all proteins involved, conservation relationships hold for a constant total of the activated and non-phosphorylated form.

Figure 4.42 shows a Simulink[9] block diagram of the Ras/Raf/MEK/ERK signal transduction pathway with RKIP regulation. Simulink is a graphical simulation environment, using the mathematical programming language Matlab. Block diagrams such as those introduced in previous sections can be drawn and simulated directly from within an interactive graphical editor. The effect of changes to parameters, the removal or introduction of feedback loops can be simulated conveniently. As illustrated Figure 4.43, with only positive feedback added to the pathway and no transport delay, the pathway

[9]Matlab and Simulink files for all models and simulations shown here are available from `www.sbi.uni-rostock.de`.

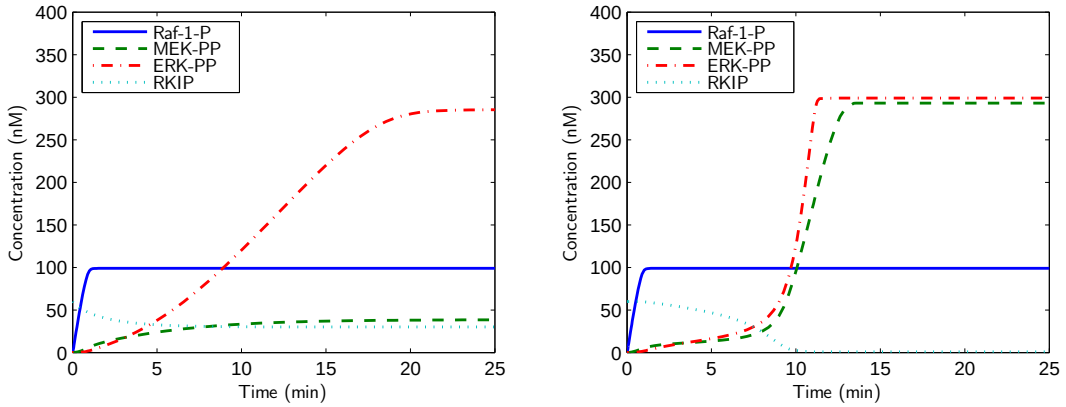


Figure 4.43: Simulation of the Ras/Raf/MEK/ERK pathway in Figure 4.41. The plots show the response to a unit-step input signal. Left: without feedback loops. Right: as before but with a positive feedback loop. Note the sharpening of the ERK response from the introduction of positive feedback loop that is realised by RKIP.

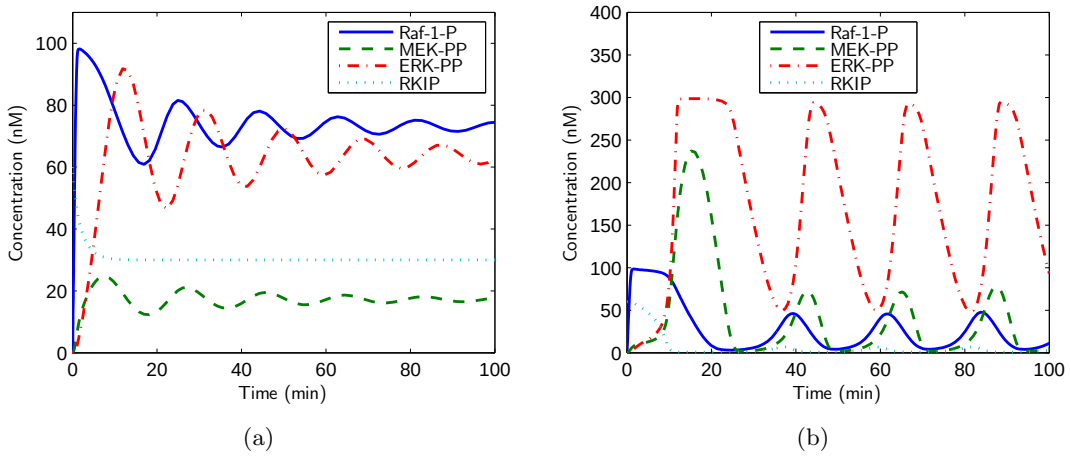
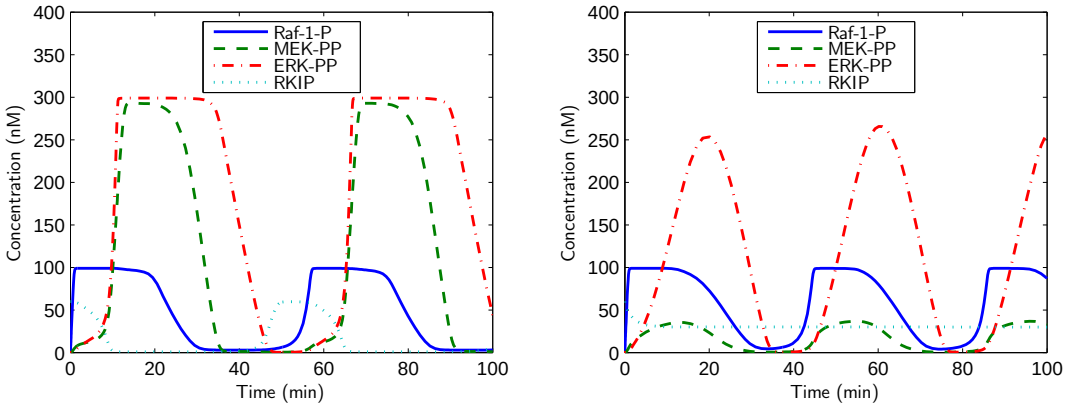


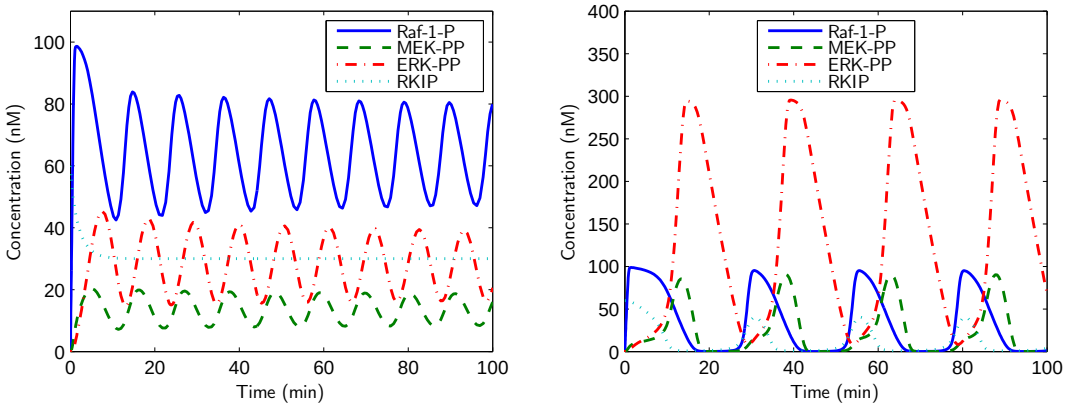
Figure 4.44: Simulation of the Ras/Raf/MEK/ERK pathway in Figure 4.41. Left: with negative feedback loop. Right: with both, positive and negative feedback loops.

displays switch-like behavior. Switching dynamics have been found in various intracellular systems (e.g. [FJ96, TCN03]). Our previous observation that a positive feedback loop sharpens the response, making it *ultrasensitive*, remains true in this more complex system. Because the positive feedback loop effects only proteins from MEK downward, the Raf concentration profile has not been changed. Considering a negative feedback loop, no transport delay and without the positive feedback loop in the system, we observe that negative feedback can destabilise the response. What can also be observed are lower steady-state values for Raf and ERK.



(a) (b)

Figure 4.45: Simulation of the Ras/Raf/MEK/ERK pathway in Figure 4.41. The plots show the response to a unit-step input signal. Left: Negative feedback loop with transport delay $T_d = 10$min. Right: As before but with additional positive feedback loop. In both cases $n = p = 1$.



(a) (b)

Figure 4.46: Simulation of the Ras/Raf/MEK/ERK pathway in Figure 4.41. The plots show the response to a unit-step input signal. Left: Negative feedback loop with $T_d = 0$, $n = 2$, $p = 1$. Right: As before but with additional positive feedback loop.

Once a model is established, simulation allows quick studies of changes to the elements and parameters. For example, one way to make the model more realistic is to introduce a time delay between ERK near or inside the nucleus and its feedback effect on Raf further up the pathway. In Figure 4.45, we introduce a transport delay in the negative feedback loop with $T_d = 10$min. We observed that transport delays lead to increased oscillatory behavior, turns the damped oscillations into sustained oscillations. In [SMT+03] this consideration for nucleocytoplasmic transport was crucial in obtaining a predictive mathematical model for the JAK-STAT pathway. Our next experiment is to

change the feedback indices n and p, that were also introduced above and which define the sharpness or sensitivity of the feedback effect. In Figure 4.46 we find that without transport delay an increase from $n = 1$ to $n = 2$ in the negative feedback loop also leads to sustained oscillations.

Our study demonstrates various sources of sustained oscillations: negative feedback combined with ultrasensitivity, combined negative and positive feedback and transport delays in negative feedback loops. Oscillations have been investigated in various systems (e.g. [Gol96, WH00, Kho00]) and have been of interest in mathematical modelling for some time (e.g. [GH83, Str00b]). An interesting question is to ask whether our model applies to a single cell or a population of cells. If a single cell is an oscillator, one would have to consider a population of coupled oscillators. Oscillations are not easy to detect and require complex experimental set-ups. Simulation studies demonstrate the usefulness of mathematical model in generating hypotheses of phenomena that have yet to be observed in experiments. On the other hand, a simulation can also be used to support the design of experiments, helping to decide which variables to measure and why. The role of feedback in intracellular dynamics has been investigated for some time in the literature (e.g. [Gri68, Tys79, BPM82, Tys83, TD90]) and will, no doubt, play an important role in (re)emerging area of systems biology.

$k_1 = 2.5$	$K_{m1} = 10$	$k_2 = 0.25$	$K_{m2} = 8$	$[\text{Raf-1}]_T = 100$
$k_3 = 0.025$	$K_{m3} = 15$	$k_4 = 0.75$	$K_{m4} = 15$	$[\text{MEK}]_T = 300$
$k_5 = 0.025$	$K_{m_5} = 15$	$k_6 = 0.5$	$K_{m6} = 15$	$[\text{ERK}]_T = 300$
$k_7 = 0.5$	$K_{m7} = 15$	$k_8 = 0.5$	$K_{m8} = 15$	$[\text{RKIP}]_T = 60$
$k_{7_f} = 0.025$	$K_N = 9$	$K_P = 9$	n, p, T_d	

Table 4.2: Parameter values for the Ras/Raf/MEK/ERK pathway model. Concentrations are in nM; k_1,k_3,k_5,k_{7_f} are in s^{-1}; k_2,k_4,k_6,k_7,k_8 in $nM{\cdot}s^{-1}$; K_{m1} to K_{m8} in nM. The right column specifies total concentrations. Note that the purpose of this model is to illustrate the role of feedback loops on protein concentration profiles only.

4.9 Modules and Control Mechanisms

In the present chapter we are going to investigate a class of *modules* with particular dynamics behaviors, such as switches, buzzers, oscillators etc. Rather than investigating dynamic properties in the time domain and with the help of phase-plane analysis, the present chapter introduces *rate curves* and *stimulus-response curves* for steady-states as a tool for the analysis of dynamic modules and control mechanisms. The presentation here is an extended version of the paper by Tyson et al. [TCN03]. The graphical representation of the modules follows the description in [TCN03].

4.9.1 Linear module

To start with, we consider a linear system in which the synthesis of the response molecule R is facilitated by a stimulus S; illustrated by the following diagram:

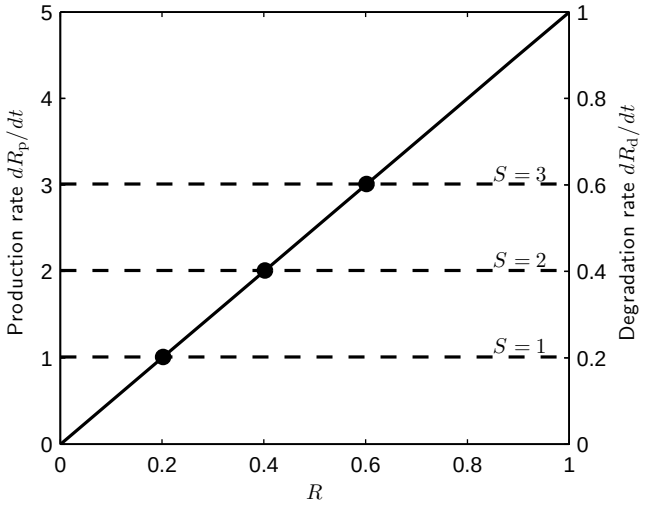
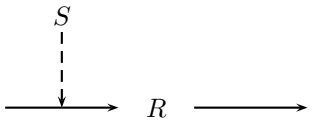


Figure 4.47: Rate curve, comparing the response rate (dashed lines) and the degradation rate (solid line) for the linear system (4.64). The intersections of both rates (black dots) are the steady states for the given stimulus. The parameters for the system are: $k_0 = 0.01$, $k_1 = 1$, $k_2 = 5$.



The mathematical model, the rate equation is given as

$$\frac{dR}{dt} = k_0 + k_1 S - k_2 R \ . \tag{4.64}$$

The term k_0 describes a constant base level synthesis of the *response component* R. What we describe here as a response could also be referred to as the *production* of R. Assuming monomolecular reactions for conversion and degradation, the second term $k_1 S$ is the conversion of the *stimulus component* S into the response component. The degradation of R is given by the last term $-k_2 R$. The *rate of response* is then the sum of the base level flux and the conversion of S:

rate of response

$$\text{rate of response} = k_0 + k_1 S \ .$$

This allows us to write for the change of the response component

$$\frac{dR}{dt} = \text{rate of response} - \text{rate of degradation.}$$

We can then compare the degradation rate and the response rate by plotting the *rate curve* as function of R. This is shown in Figure 4.47. Note that hereafter we are going to suppress units in plots for easier viewing. Depending on the number of participating components the rate coefficient k_i has the unit 'per second' divided by the unit of the

rate curve

component to the power $(n-1)$. Michaelis-Menten constants have the same units as the corresponding component. The solid line represents the degradation or removal of R, which is $k_2 R$. The rate is a straight line with a slope equal to k_2. The rate of response $k_0 + k_1 S$ (dashed lines) is independent of R and thus a horizontal line. The points, where the rate of response and the rate of degradation are equal are of particular interest:

$$\text{rate of response} = \text{rate of degradation} \ . \tag{4.65}$$

At these points the system is in a steady-state such that no net change of concentrations is measurable. This means response and degradation are in balance, although this does not mean that no reaction occurs. This state is mathematically determined by the balance equation

$$\frac{dR}{dt} = \text{rate of response} - \text{rate of degradation} = 0 \ . \tag{4.66}$$

Note that a system fulfilling condition (4.65) is either in equilibrium or in a steady-state. There is no difference between both states from the mathematical point of view. If the system is in one of these states, all sub-systems have to fulfil (4.65) too. The difference between both states lies in the considered type of system. An equilibrium is defined for a *closed system*, where there is no transport of matter and energy in and out of the system. A closed system will relax to the equilibrium state and will not leave it by itself. Following a small perturbation the system returns to the equilibrium. In an *open system*, such as a cell, the transport of energy is possible. If we further consider the flow of matter, we get a *fully open system*.

closed/open system

Here we want to describe the response of a system to an external stimulus, without assumptions about flow of energy or matter, and therefore assume a fully open system, that are characteristic of living systems. In such a system we usually reach a steady-state dependent from the external stimulus. The system state is held by the signal and can be far away from the equilibrium state. Only, if we close the system, for instance we choose $S = 0$, the system relaxes into its equilibrium state. In this sense, the equilibrium is a particular steady-state but if there is no flow of molecules in and out of the cell this could mean that they die. As we will see later for the sigmoidal module, the conditions (4.65) and (4.66) are not fully equivalent. Strictly speaking, if condition (4.65) is fulfilled, the system is in a steady-state. It can but has not to be in a steady-state if (4.66) is fulfilled.

Note that the dR_r/dt of the ordinate in Figure 4.47 is not the *net-rate* dR/dt on the left-hand side of (4.64). Figure 4.48 shows the net-rate as a function of R. Is the net rate equal to zero, shown as horizontal dashed line in the plot, the production and degradation rates are balanced. The system is in a steady-state.

Next we investigate the steady-state as a function of the stimulus. According to (4.64) and (4.66) this state is obtained as a solution to the equation

$$0 = k_0 + k_1 S - k_2 R \ .$$

The steady-state value of the response component is an increasing straight line with slope k_1/k_2

$$R_{\text{ss}} = \frac{k_0}{k_2} + \frac{k_1}{k_2} S \ . \tag{4.67}$$

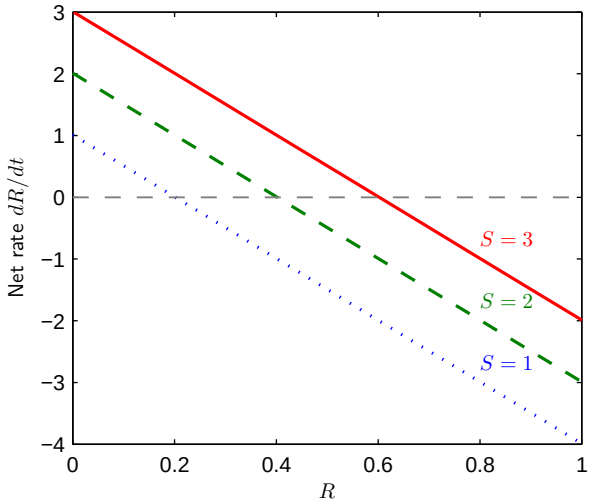


Figure 4.48: The net rate for a linear system as function of the response component R for three values of the signal S. If the net rate is greater than zero, R increases and if the net rate is smaller than zero, R decreases. The value $dR/dt = 0$ (dashed horizontal line) corresponds to the steady-state. Parameters: $k_0 = 0.01$, $k_1 = 1$, $k_2 = 5$.

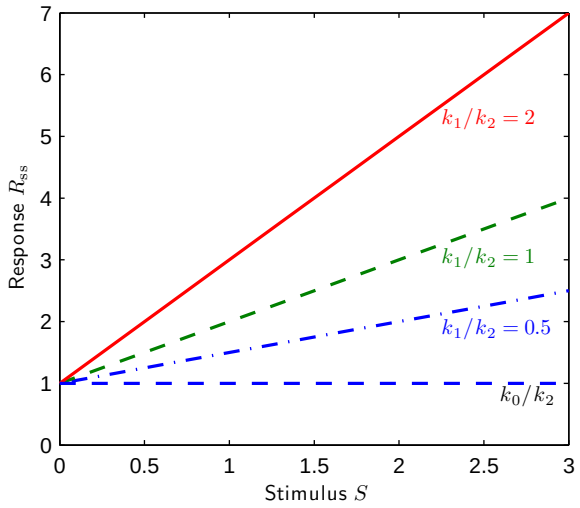


Figure 4.49: Stimulus-response curve for a linear module. Three ratios of k_1/k_2 are compared, whereby the decay rate k_2 is constant. The response R_{ss} is the stronger the stronger the signal S is. The horizontal line is the limit for a vanishing stimulus. Parameters: $k_0 = 1$, $k_2 = 1$.

If the stimulus is increasing than the response is increasing too. For a finite constant external source of R the response is $R_{ss} = k_0/k_2$ at a signal strength $S = 0$. The representation of the steady-state R_{ss} as function of the signal S is shown in Figure 4.49. This kind of plot is called *stimulus-response curve*. Figure 4.49 shows the response of the linear module for different ratios of k_1/k_2.

stimulus-response curve

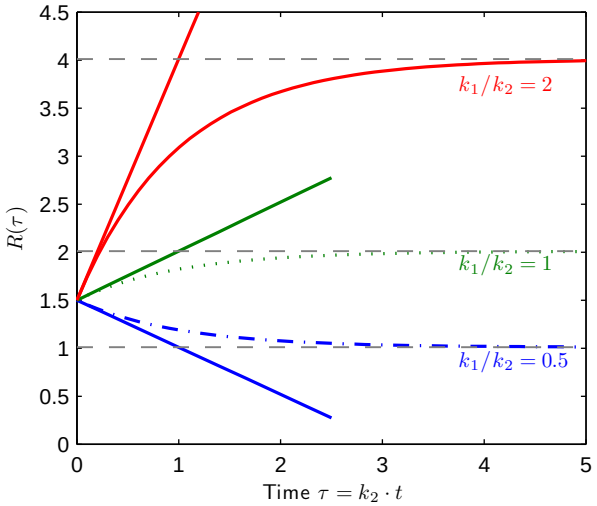


Figure 4.50: The response of a linear module for a step change to $S = 2$ as function of the dimensionless time $\tau = k_2 \cdot t$. Three ratios of k_1/k_2 are shown. Parameters: $k_0 = 0.01$, $k_2 = 1$, $R_0 = 1.5$.

To this point, we have investigated static properties of the system. Investigating the temporal evolution of R, we now assume that the signal S does not change over time. This may also be interpreted as the response to an abrupt change of the stimulus, at $t_0 = 0$, to the given value, which is then constant ("step-response"). With this assumption, (4.64) is a linear first-order differential equation for which the solution can be found analytically:

$$R(t) = \frac{k_0 + k_1 S}{k_2}\big(1 - \exp\{-k_2 t\}\big) + R_0 \exp\{-k_2 t\} \ , \tag{4.68}$$

where R_0 is the initial value of R at t_0. The first term on the right-hand-side describes the relaxation to the stationary state R_{ss} (4.67) and the second term is the degradation of the initial state R_0. The relaxation time

$$\tau = \frac{1}{k_2}$$

is independent from external stimulus S. This quantity is an approximation to the time needed to relax into the steady state. It can also be interpreted as the time the system is able to react to perturbations. For $t \ll 1/k_2$ we can expand the exponential function

$$e^{-x} \approx 1 - x + O(x^2) \ ,$$

where $O(x^2)$ denotes terms of the order of x^2. For a short initial period of time that leads to

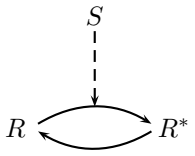
$$R(t) \approx (k_0 + k_1 S - k_2 R_0)\, t + R_0 \ . \tag{4.69}$$

The response $R(t)$ as function of the dimensionless time $\tau = k_2 \cdot t$ is shown in Figure 4.50. We chose a fixed value for the decay rate coefficient k_2 and vary the rate coefficient k_1.

According to (4.68) the response component $R(t)$ exponentially relaxes to the steady-states (4.67). The solid straight lines are the asymptotic solutions for initial times (4.69). The horizontal dashed lines correspond to the steady state. Analogue to Figure 4.49 the steady-state depends on the ratio k_1/k_2. If the production dominates, $k_1 > k_2$, the steady-state value is greater than the signal strength[10] S.

4.9.2 Hyperbolic module

A hyperbolic module is slightly more complicated than the linear module discussed above. The response or activation is now understood as a bimolecular reaction while degradation is still considered monomolecular.



This formulation is a popular model for activation or phosphorylation in signalling pathways. In this situation we denote by R^* the activated or phosphorylated form and use R to denote the non-activated form. Such a reaction system we describe as

$$\frac{dR^*}{dt} = k_1 SR - k_2 R^* \ ,$$

where in the case of a signalling pathway, stimulus S corresponds to a kinase that facilitates the phosphorylation of R. If we assume a constant total concentration,

$$R_T = R^* + R \ , \tag{4.70}$$

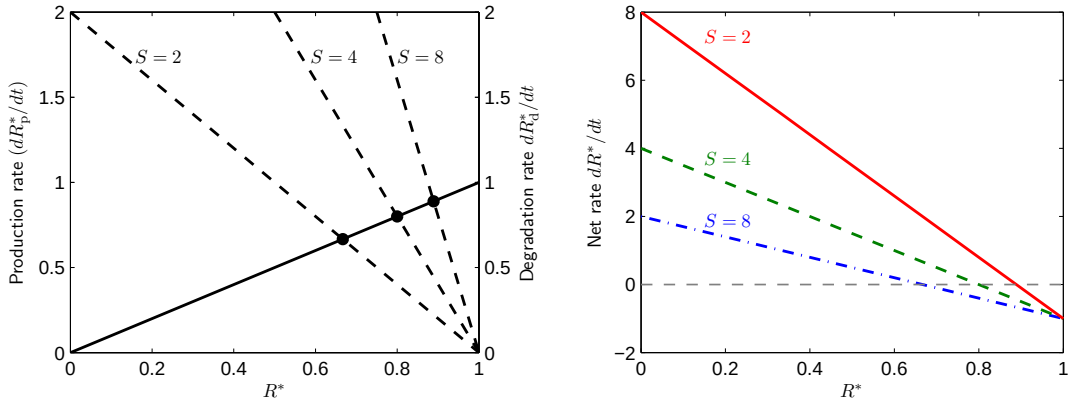
we can rewrite the differential equation of the hyperbolic module:

$$\frac{dR^*}{dt} = k_1 S(R_T - R^*) - k_2 R^* \ , \tag{4.71}$$

where the first term on the right-hand-side is the rate of response or activation and the second the rate of deactivation. The corresponding rate curve is shown in Figure 4.51. The rate of deactivation (solid line) has a slope of k_2. According to the conservation law (4.70) the response component is restricted to the range $[0, R_T]$. If all R molecules are phosphorylated the response must be zero. The net rate of (4.71) is a straight line with a slope $-(k_1 S + k_2)$. The conservation law restricts the response to $[0, 1]$. If the net rate is greater than zero, activation dominates. The intersections with the zero line are the steady-states values of R^* for this system. From (4.71) we obtain for the steady-states a hyperbolic function, which gives this module its name:

$$R^*_{\text{ss}} = \frac{SR_T}{k_2/k_1 + S} \ . \tag{4.72}$$

[10] The first term k_0/k_2 in (4.67) is negligible for $k_0 = 0.01$ and $S = 2$.



(a) (b)

Figure 4.51: Left: The rate curve for the hyperbolic module, assuming bimolecular activation and monomolecular deactivation, for different signal strengths. The deactivation rate is the solid line and the rate of response for three values of S are shown by dashed lines. The intersections marked by the black dots are the steady states for the shown parameters $k_1 = 1$, $k_2 = 1$, $R_T = 1$. Right: Net rate as a function of R^* for three different stimuli.

For the limits $S \to 0$ and $S \to \infty$ we can expand (4.72). For a signal strength $S \ll k_2/k_1$ the contribution of S in the denominator is negligible and the steady-states govern the linear function

$$R^*_{\text{ss}} \approx \frac{k_1}{k_2} R_T S \ . \tag{4.73}$$

The ratio of the rate coefficients k_1/k_2 determines the slope of the asymptote. If for the signal strength $S \gg k_1/k_2$, the signal S dominates the denominator. In this case all proteins are phosphorylated and $R^*_{\text{ss}} \approx R_T$. The stimulus-response curve for a hyperbolic module is shown in Figure 4.52. The straight lines are the asymptotic expansion (4.73). If the rate coefficient of activation k_1 greater than the rate coefficient of deactivation k_2, the hyperbolic properties of the system is in evidence. For small values of this ratio the hyperbolic system looks like a linear system within the presented signal range.

The temporal evolution of the hyperbolic module is described by the differential equation (4.71). This is again a first-order linear differential equation that we can solve analytically:

$$R^*(t) = \frac{R_T S}{k_2/k_1 + S} \left(1 - \exp\left\{-(k_2 + k_1 S)t\right\}\right) + R_0 \exp\left\{-(k_2 + k_1 S)t\right\} \ . \tag{4.74}$$

The first term describes the relaxation to the steady-state while the second term corresponds to the degradation from the initial state. In contrast to the linear system the relaxation time

$$\tau = \frac{1}{k_2 + k_1 S}$$

is now dependent on the signal strength. The greater the signal strength, the faster the steady-state is reached. For times $t \ll \tau$ we can expand the exponential function and

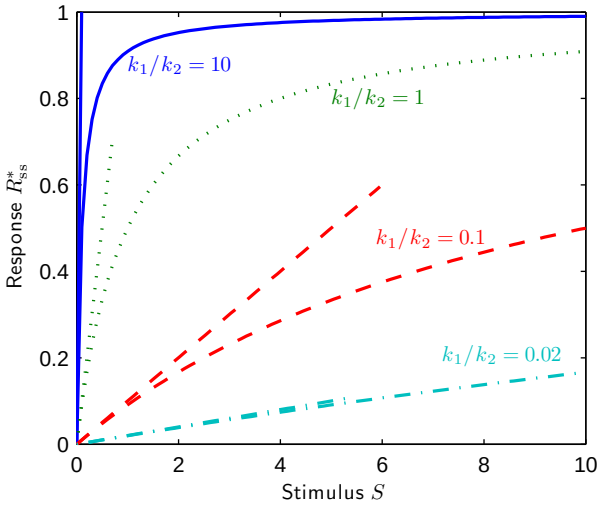


Figure 4.52: The stimulus-response curve of a hyperbolic module (4.71) for different ratios of k_1/k_2. The straight lines are the asymptotes for small signal strength (4.73). Parameters: $k_1 = 1$, $R_T = 1$.

obtain for small initial times the asymptote

$$R^*(t) \approx \Big[k_1(R_T - R_0)S - k_2 R_0\Big]\, t + R_0 \ . \tag{4.75}$$

The temporal evolution of the hyperbolic module is shown in Figure 4.53. We compare solutions for three different ratios of the rate coefficients. For comparison, the small times asymptotes (4.75) are also drawn.

4.9.3 Sigmoidal module

The next module is an extension of the hyperbolic module. Both activation and deactivation are here considered as Michaelis-Menten type kinetic reactions, leading to the differential equation

$$\begin{aligned} \frac{dR^*}{dt} &= \frac{k_1 S R}{K_{\mathrm{M1}} + R} - \frac{k_2 R^*}{K_{\mathrm{M2}} + R^*} \\ &= \frac{k_1 S (R_T - R^*)}{K_{\mathrm{M1}} + R_T - R^*} - \frac{k_2 R^*}{K_{\mathrm{M2}} + R^*} \ , \end{aligned} \tag{4.76}$$

where on the second line, the conservation law (4.70) is used to relate the unphosphorylated form R of the response component to the phosphorylated form R^*. The parameter $K_{\mathrm{M1}}, K_{\mathrm{M2}}$ are Michaelis-Menten constants. The first term on the ride-hand-side describes the activation, while the second term corresponds to the deactivation or dephosphorylation. A comparison of response and deactivation rate is shown by the rate curve in Figure 4.54. The conservation law (4.70) limits the possible values of R^* to the interval $[0, R_T]$. If R^* reaches the value R_T, the activation rate has to be zero, independent of the signal strength, since there is no more unphosphorylated R available.

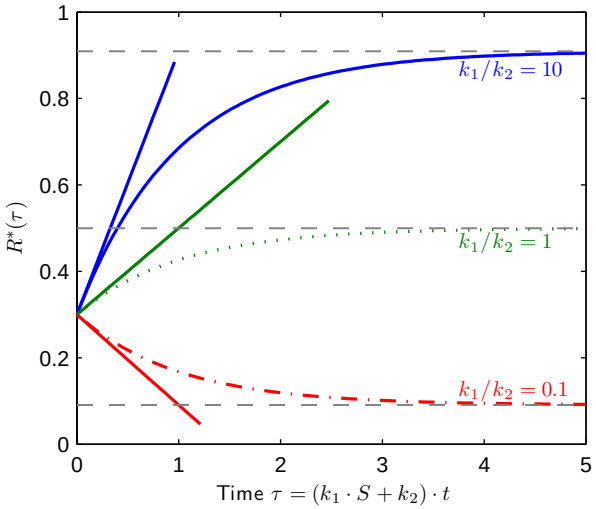


Figure 4.53: Temporal evolution of the hyperbolic module (4.71). The full solution of the differential equation (4.74) is drawn for three different ratios of k_1/k_2. Parameters: $k_1 = 1$, $S = 1$, $R_T = 1$, $R_0 = 0.3$.

The steady-state of the sigmoidal module (4.76) is determined by the quadratic equation

$$0 = \frac{k_1 S(R_T - R^*)}{K_{\mathrm{M1}} + R_T - R^*} - \frac{k_2}{K_{\mathrm{M2}} + R^*} \ . \tag{4.77}$$

For $0 < R^* < R_T$, the solution is given by what is known as the *Goldbeter-Koshland*

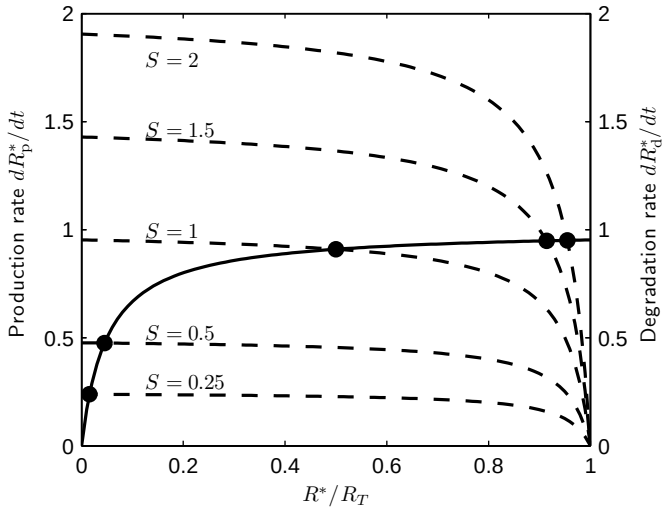


Figure 4.54: The activation and deactivation rate of the sigmoidal module (4.76). The solid line is the deactivation rate. The dashed lines are the rate of activation for different signal strength. The black dots mark the points where both rates are in balance. Parameters: $k_1 = 1$, $k_2 = 1$, $K_{\mathrm{M1}} = 0.05$, $K_{\mathrm{M2}} = 0.05$, $R_T = 1$.

function:

$$\frac{R^*_{\rm ss}}{R_T} = \frac{k_2\left(1+\frac{K_{\rm M1}}{R_T}\right)+k_1S\left(\frac{K_{\rm M2}}{R_T}\right)}{2(k_2-k_1S)} + \frac{\sqrt{\left[k_2\left(1+\frac{K_{\rm M1}}{R_T}\right)+k_1S\left(\frac{K_{\rm M2}}{R_T}\right)\right]^2+4\frac{K_{\rm M2}}{R_T}(k_2-k_1S)k_1S}}{2(k_2-k_1S)} \ . \quad (4.78)$$

Since this expression is somewhat complicated we replace it by the short form

$$\frac{R^*_{\rm ss}}{R_T} = G(k_1S, k_2, K_{\rm M1}/R_T, K_{\rm M2}/R_T) \ . \quad (4.79)$$

The solutions are shown in Figure 4.55. Additionally, the allowed range of R^* is shown as a grey-shaded region. Only one solution of (4.77) lies inside this region. The others do not fulfil the physical restrictions on the response component and are therefore irrelevant for a biological system. This system is our first example, where a mathematical solution of (4.66) is not a steady-state. This demonstrates that the conditions (4.65) and (4.66) are not fully equivalent. Strictly speaking, if condition (4.65) is fulfilled, the system is in a steady-state. It can but must not be in a steady-state if (4.66) is fulfilled. In case of the sigmoidal module we cannot derive the solution $R^*(t)$ to differential equation (4.76) in an analytical form.

The stimulus-response curve of the sigmoidal module is shown in Figure 4.56. The sigmoidal shape of the curve is determined by the ratio of the Michaelis-Menten constants. If the Michaelis constant of the activation term is much smaller than the constant

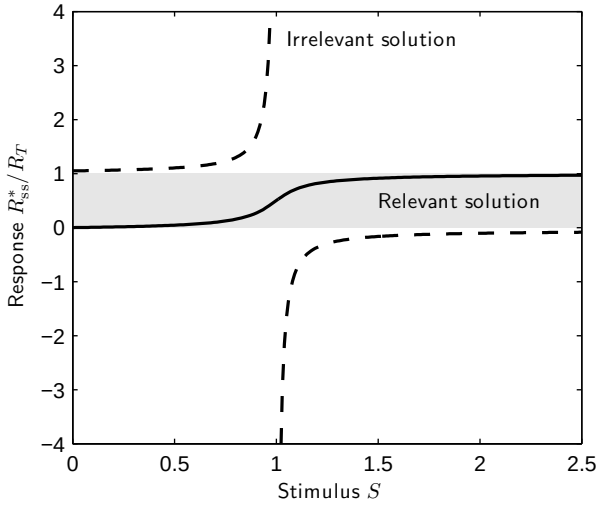


Figure 4.55: Solutions of (4.77) as a function of the signal strength. Only the relevant solution ($+\sqrt{\ldots}$) fulfils the conservation law (4.70). The irrelevant solution ($-\sqrt{\ldots}$) lies in physically unreachable state-space. The grey shaded region shows the reachable states that follow from the conservation law.

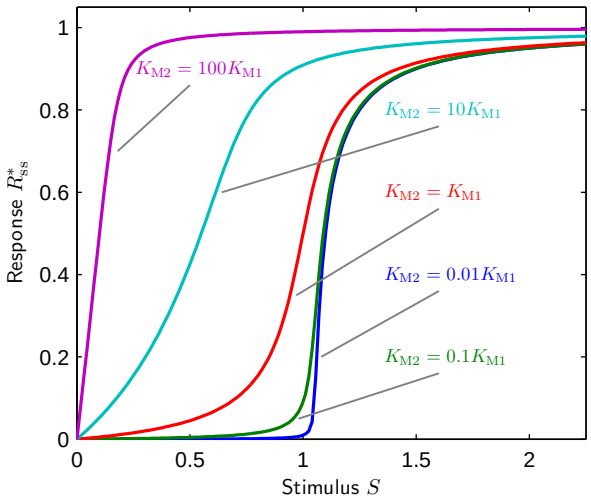
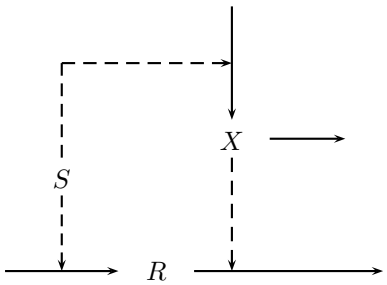


Figure 4.56: The stimulus-response curve for the sigmoidal system (4.76) for different ratios of the Michaelis-Menten constants K_{M1} and K_{M2}. Parameters: $k_1 = 1$, $k_2 = 1$, $K_{M1} = 0.05$, $R_T = 1$.

for the deactivation term, the typical sigmoidal shape vanishes. On the other hand, if the activation constant K_{M1} is much larger than K_{M2}, we get a switch-like behavior for the response function. It is for this reason that the system (4.76) is sometimes called Goldbeter-Koshland switch [GKJ81].

4.9.4 Robust or adaptive module

In the next dynamic module we consider, $R(t)$ is robust to changes in the stimulus. Looked at from the perspective of $X(t)$ the system is adaptive in that it tracks the stimulus. While there is a transient response of $R(t)$ to step-changes in $S(t)$, the response returns eventually to its steady state value R_{ss}. Such a behavior can be realised by two parallel reaction paths pushing the response in opposite directions:



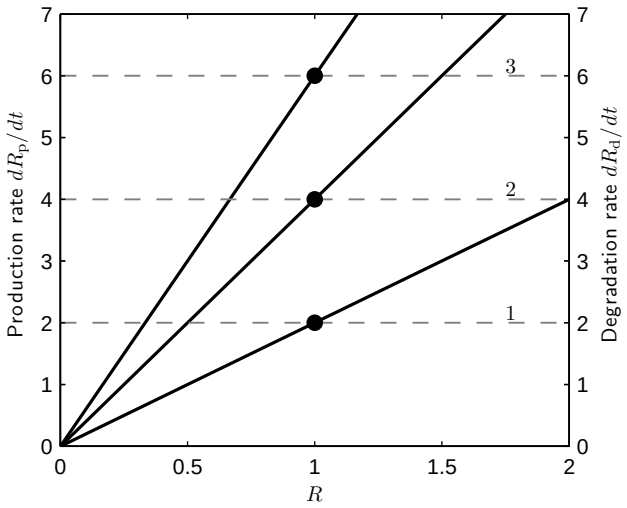


Figure 4.57: Rate curve of a perfectly-adapted or robust system (4.80) for different external stimuli S and amounts of X. The rate of response (dashed lines) is proportional to the external signal and the rate coefficient k_1, but constant with respect to R. The solid lines are the rates of degradation for three different X. These linear functions have the slope k_2X. The steady states, marked by dots, have a constant R-value. Parameters: $k_1 = k_2 = 2$, $k_3 = k_4 = 1$.

For the simplest case we combine two linear systems

$$\begin{aligned} \frac{dR}{dt} &= k_1 S - k_2 X R \ , \\ \frac{dX}{dt} &= k_3 S - k_4 X \ , \end{aligned} \tag{4.80}$$

where both systems are coupled through the parallel stimulus by S. The degradation or deactivation of the response component R depends on the amount of X. Because an increasing signal will also increase the amount of X, this simultaneously increases the degradation rate of the response component. The rate curves for this module are shown in Figure 4.57. For the response component R we have:

$$\begin{aligned} \text{rate of response} &= k_1 S \ , \\ \text{rate of degradation} &= k_2 X\, R \ . \end{aligned}$$

The response rate is a linear function of the stimulus S, with slope k_1. The degradation rate is independent of S but dependent on X. For this reason we vary the signal strength for the rate of production and the amount X in Figure 4.57. The intersections of corresponding curves are again the steady states of the system. In the representation of the net rate (Figure 4.58) this is more visible. According to the underlying linear system, we obtain again linear functions. The slope is given by k_2X and the ordinates by k_1S.

For the steady-state R_{ss} we have to solve the system of balance equations

$$\begin{aligned} 0 &= k_1 S - k_2 X\, R \ , \\ 0 &= k_3 S - k_4 X \ , \end{aligned}$$

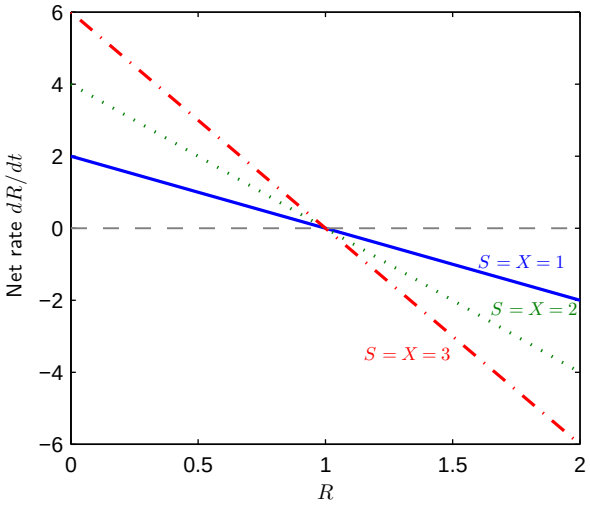


Figure 4.58: Net rate for system (4.80) for various external signals S and amounts of X. According to the underlying linear system, the net rate is a linear function, too. The common intersection of all three curves approves the behavior of perfect adaptation. Parameters: $k_1 = k_2 = 2$, $k_3 = k_4 = 1$.

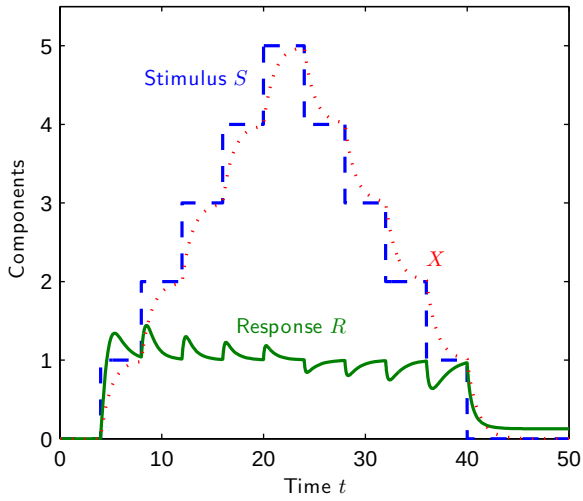


Figure 4.59: Temporal evolution of robust-adaptive module (4.80) under the influence of a stepwise change in stimulus. Parameters: $k_1 = k_2 = 2$, $k_3 = k_4 = 1$, $\Delta t = 4$, $\Delta S = 1$, $S_0 = X_0 = R_0 = 0$.

leading to a constant steady state response

$$R_{\text{ss}} = \frac{k_1 k_4}{k_2 k_3} \ .$$

The steady state response is therefore determined by the ratios of the rate coefficients for both channels. In Figure 4.59 we plot the temporal evolution of stimulus S, regulating

component X, and response component R. We see that the response is only transient and returns eventually to its steady state. The height of the transient peak depends on X. The larger X, the smaller the peak. The duration for which S is kept constant is denoted by Δt and the height of the step-change by ΔS. The return of $R(t)$ to its steady state is due to the monomolecular decay that after some time is faster than the bimolecular degradation of R. Looking at R from S, the system may be considered *robust* to changes in S. Looked at from the perspective of X, the system shows perfect *adaptation* the changes.

4.9.5 Feedback systems

In previous sections we considered systems without feedback. In the context of this chapter, feedback means, that the response component influences its own production or degradation. We refer to the feedback as positive or negative, depending on whether the response amplified or suppressed [MVHA99]. Positive feedback can be realised through:

1. Acceleration of the production, for instance in an autocatalytic process,
2. Inhibition of the degradation.

Feedback is negative, if it weakens the response signal through:

1. Inhibition of the production,
2. Acceleration of the degradation.

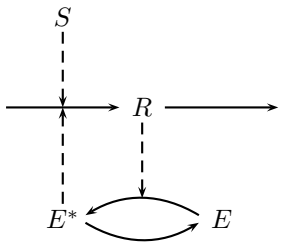
In subsequent sections we discuss different feedback mechanisms, starting with positive feedback, negative feedback followed by mixed negative and positive feedback.

4.9.5.1 Positive feedback/feedforward - Switches

The present section is to discuss the two ways by which positive feedback/feedforward control mechanisms can be realised. The acceleration of production of the response component R is related to *mutual activation*, while the inhibition of degradation is related to *mutual inhibition*. In both cases, the positive feedback/feedforward can create a switch-like behavior, in which a certain level of stimulus can lead to a sudden change of the response.

Mutual activation - positive feedback

We add to the linear system (4.64) a backward directed loop, i.e., feedback is realised through an intermediate enzyme E. The response component activates the enzyme E, for instance through phosphorylation. The activated enzyme E^* on the other hand enhances the synthesis of R. In this sense, both components S and R mutually activate R:



To study the properties of this system we assume that the activation of the enzyme E can be described by means of the Goldbeter-Koshland function (4.79) such that the differential equation model of the mutually activated system is

$$\frac{dR}{dt} = k_0 E^*(R) + k_1 S - k_2 R \tag{4.81}$$

where

$$E^*(R) = G(k_3 R, k_4, J_3, J_4) \ .$$

The rate of degradation is a linear function of the response component,

$$\text{rate of degradation} = k_2 R \ ,$$

and depends only on R. The proportionality coefficient is again the rate coefficient k_2. In contrast to systems discussed in previous sections, the rate of production,

$$\text{rate of production} = k_0 E^*(R) + k_1 S \ ,$$

is now a function of the response component itself as well as signal S. Thereby, we neglect a signal- and enzyme-independent production of the response component and assume that the enzyme-catalyzed production is dominant. The comparison of both rates, shown in Figure 4.60, illustrates new properties. Dependent on the signal strength S, the number of intersections between the curves varies between one and three. Let us, for the time being, consider these as steady states, although we have to refine the meaning of steady-states.

Examining the net-rate of the mutually activated system, with the same parameters as in Figure 4.60, the resulting curves are shown in Figure 4.61. There is a signal strength, for which there are cases with a net rate equal to zero. We also notice that there is a region for which the net-rate increases with R. For this region the lines are dashed segments. Mathematically this corresponds to the condition

$$\frac{d}{dR}\frac{dR}{dt} > 0 \ ,$$

where the derivations are carried out in the given order from left to right. For the considered system (4.81) we have

$$\frac{d}{dR}\frac{dR}{dt} = k_0 \frac{dE^*}{dR} - k_2 \ . \tag{4.82}$$

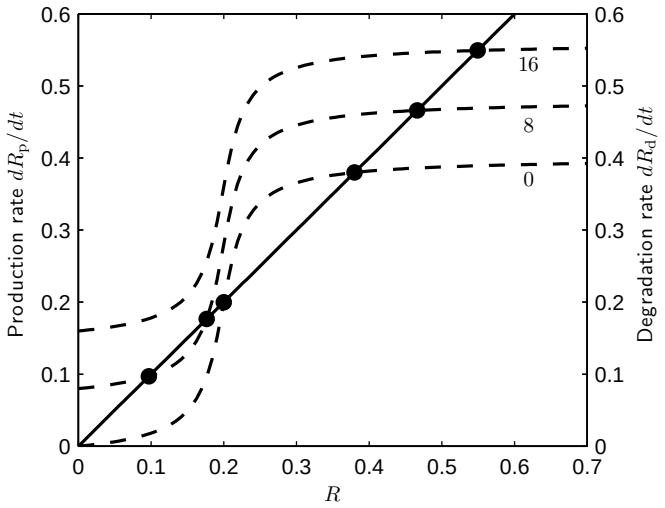


Figure 4.60: Comparison of the rate of degradation (solid line) and the rate of production for different signal strengths (dashed lines), for a mutually activated system (4.81). Parameters: $k_0 = 0.4$, $k_1 = 0.01$, $k_2 = 1$, $k_3 = 1$, $k_4 = 0.2$, $J_3 = 0.05$, $J_4 = 0.05$.

This is an expression that is independent of the signal strength. The derivative of dE^*/dR can be derived analytically, although this is cumbersome and therefore avoided here. To see what happens, if a zero net-rate is located within this range, let us assume, that we are in steady state. If we perturb the system a little by increasing the response component, the rate of degradation will decrease. The system cannot return to its former state. On the other hand, if we decrease the response component the rate of production also decrease and the system cannot return. This state is unstable and hence not a steady or equilibrium state. The necessary condition for a stable (bio)chemical system is thus

$$\frac{d}{dR}\frac{dR}{dt} < 0 \ .$$

Only, if this condition is fulfilled, the system returns to the steady state for small perturbations to R. In Figure 4.62 we illustrate the differences between stable and unstable states, using the analogy of a ball in a landscape of changing altitudes. The only stable equilibrium[11] states are the minima of the height function. As we have seen above, the ball is moving back to the deepest point of the valley after a small perturbation. The position on the top of the mountain is very unstable. For a comprehensive discussion of these issues see [GS94]

Returning to our mutually activated module (4.81), in Figure 4.61 a stimulus independent region occurred, for which no stable solution exits. For the chosen parameter values it lies within the interval $0.147 < R < 0.249$. On the other hand, for values of R outside this region we obtain stable solutions. Dependent on the signal strength, the net

[11] We assume, that the motion of the ball depends from the height only and that there are no further macroscopic forces. The system is then closed and can reach a state with minimal potential energy, an equilibrium state. Nevertheless, there can be more than one equilibrium states.

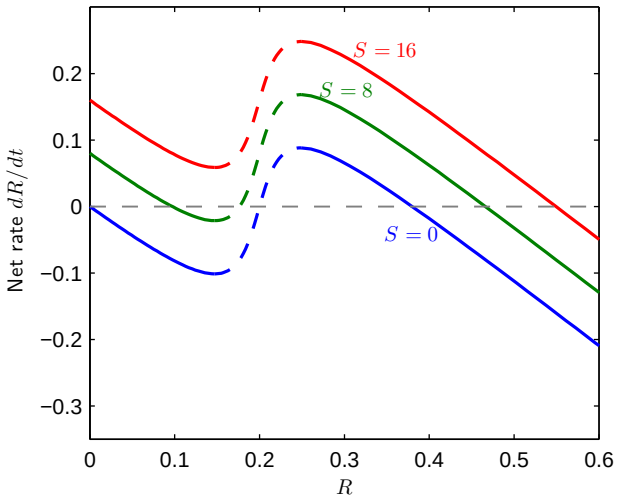


Figure 4.61: The net-rate as function of the response component R for the system (4.81). The results for three different signal strengths are plotted. The curves are separated into two parts. The solid line represents stable solutions and the dashed lines unstable solutions. Parameters: $k_0 = 0.4$, $k_1 = 0.01$, $k_2 = 1$, $k_3 = 1$, $k_4 = 0.2$, $J_3 = 0.05$, $J_4 = 0.05$.

rate has now one or two stable zeros. Let us analyze the net-rate for a zero stimulus of Figure 4.61 in more detail. As S increases, the location of the zero moves to the right. If the value of the minima is positive, no steady state exists in the lower branch. The steady states of the mutually activated system are obtained from the balance equation

$$0 = k_0 E^*(R) + k_1 S - k_2 R \; . \tag{4.83}$$

The analytic solution is lengthy and complicated so that we restrict our discussion to few important features. Equation (4.83) has three solutions. One is a full real function. The others are partly complex. Because the response component must be a real measurable

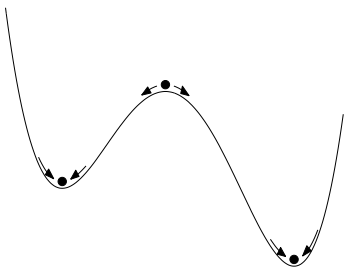


Figure 4.62: Illustration of stable and unstable states for a ball in a landscape of changing altitudes. The maximum and the minima of this schematic mountain scenery fulfil the necessary condition $dh/dx = 0$, where h denotes the height and x the horizontal position. But only the minima are stable equilibrium states, because the ball will return back from alone to his position after a small perturbation. The maximum is a unstable state; a small perturbation means the ball drops down to one of the equilibrium states.

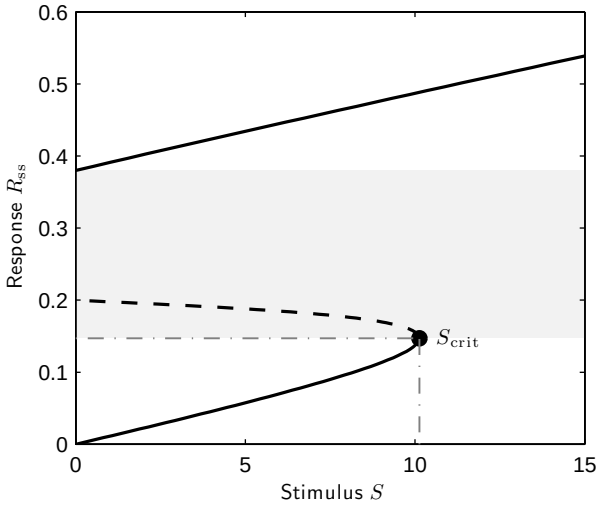


Figure 4.63: Signal-response curve for the mutually activated system (4.81). The balance equation (4.83) has three branches. The upper one is the full real solution. The others are only in parts real. The steady-states (stable solutions) are shown as solid lines. The unstable solution is drawn as a dashed line. Because of the two stable branches, the system is called bistable. The grey region of the plots denotes the gap between the upper and the lower stable branches. Parameters: $k_0 = 0.4$, $k_1 = 0.01$, $k_2 = 1$, $k_3 = 1$, $k_4 = 0.2$, $J_3 = 0.05$, $J_4 = 0.05$.

quantity, such as a concentration, only the real parts are relevant for the analysis. All three solutions are shown in the signal-response curve of Figure 4.63. The mutually activated system (4.81) has two stable branches plotted as solid lines with an unstable branch, shown as a dashed line. Such a system is called *bistable*. As the stimulus S increases, the response is low until a critical level of stimulus is reached. At this point the module's behavior changes with a sudden increase of response R. If thereafter S decreases, the response remains high, the switch is irreversible. Because of its irreversible behaviour this system is also called a *toggle switch*. Note that the sigmoidal module can act as a reversible switch. The plot in Figure 4.63 as a (one-parameter) *bifurcation diagram*. The critical point is in this context called a *bifurcation point*.

bistable switch

toggle switch

Note that so far we have looked at steady values of S and R and whenever we spoke of a change in S, the switch-like behavior was discussed in the stimulus-response plane, not in the time domain. The transition that occurs with changes to S were not explicitly included in our discussion. Figure 4.64 shows numerical solutions to the differential equations (4.81) for different initial values of R. Figure 4.64 displays solutions to the system with mutual activation through positive feedback. The range of the initial response component R_0 can be separated into two parts. The lower part relaxes to the steady state given by the lower branch of the bifurcation diagram, while for values higher than R_{crit}, the system moves to the upper branch in Figure 4.63. Both parts are separated by the unstable solution of the balance equation (4.83) not by the critical response R_{crit}. The unstable solution is a separatrix in the phase space. At this point it is a good opportunity to point out why stimulus-response curves are valuable. As can be seen from

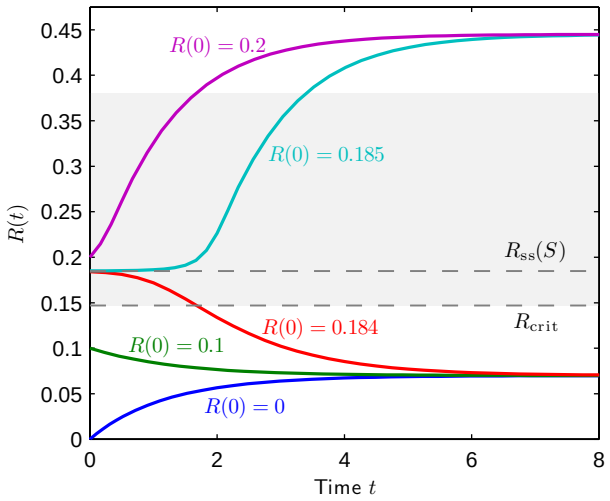


Figure 4.64: Relaxation of the mutual-activation module into the steady state for different initial response signals R_0. The unstable solution of the balance equation (4.83) acts as a separatrix, which separates trajectories. For $R_0 < R_{\mathrm{ss}}(S)$ the system achieves a steady state on the lower branch. Outside this range the upper branch becomes the final state. For comparison, the unstable solution and the critical response signal R_{crit} are shown. Parameters: $k_0 = 0.4$, $k_1 = 0.01$, $k_2 = 1$, $k_3 = 1$, $k_4 = 0.2$, $J_3 = 0.05$, $J_4 = 0.05$, $S = 6$.

the time-plots, the behavior of the system displayed in those plots depends critically on the stimulus and initial conditions. The bifurcation plot on the other hand summarises the behavior in an efficient way.

Figure 4.65 illustrates the irreversibility of the considered system. We compare the response of the system on a sub- and a supercritical signal step change at $t = 0$. The initial response signal is $R_0 = 0$. The signal remains constant until the system reaches a steady state. After a certain time, $t = 15$, the signal is switched and the response relaxes to a new steady state. The sub-critically stimulated system goes back to a zero response signal, while the supercritical stimulus shows the expected behavior of a continued high level. In the second case the nonlinearity of the system is visible. Looking at the response with the subcritical stimulus, one might interpret the temporal evolution $R(t)$ as the consequence of a linear system. Figure 4.66 illustrates the fact that in addition to a critical value of S, the stimulus must persist for a sufficient period of time if the full bistable behavior of the system is to be observed.

Mutual inhibition - positive feedforward

From the above definition of positive feedback/feedforward control, there is another possibility to increase the response signal. In the previous section we increased the rate of production via an intermediate enzyme. Now, we use a similar model to inhibit degradation. Here the response component is acting 'forward' via E:

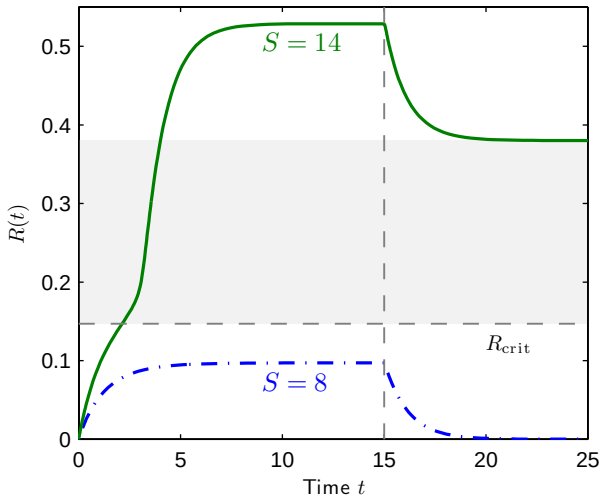
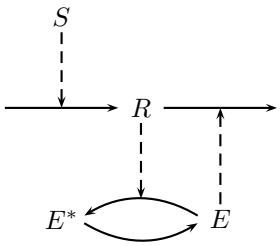


Figure 4.65: The time evolution for two different stimuli, the first subcritical and the second supercritical. Once the response relaxes into the S-dependent steady state, the signal is switched off. The critical response signal R_{crit} is important for the change of the behavior of the system. At this point the activation/deactivation strongly increase from a lower to a high level. Parameters: $k_0 = 0.4$, $k_1 = 0.01$, $k_2 = 1$, $k_3 = 1$, $k_4 = 0.2$, $J_3 = 0.05$, $J_4 = 0.05$, $S = 14$, $R_0 = 0$.



For mutual inhibition, the response component R facilitates the activation of enzyme E. The activated E^* in turn increases R. The corresponding system of differential equations is

$$\begin{aligned} \frac{dR}{dt} &= k_0 + k_1 S - \left[k_2 + k_2' E(R)\right] R \\ E(R) &= G(k_3, k_4 R, J_3, J_4) \ , \end{aligned} \tag{4.84}$$

where again we assume that the enzyme reaction is much faster than the signalling reaction. Therefore we can use the steady state solution of this reaction given by the Goldbeter-Koshland function. The term $k_2' R$ describes the direct decay of R and $k_2' E(R) R$ the enzyme catalyzed degradation. The rate curve of this system is shown in Figure 4.67. The solid line is the rate of degradation and the dashed lines are the production rates for different signal strengths. For the mutually inhibited system, the rate of production is independent of the response component. Again, there are more than one intersection for some values of the stimulus. The corresponding balance equation is

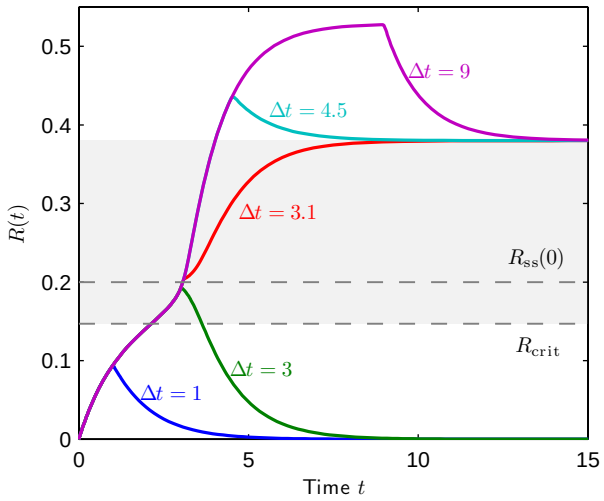


Figure 4.66: The temporal evolution of the mutually activated system (4.81) for a supercritical signal $S = 14$ of different durations. The activated state is reached only and only if the separatrix $R_{ss}(0)$ is exceeded. Parameters: $k_0 = 0.4$, $k_1 = 0.01$, $k_2 = 1$, $k_3 = 1$, $k_4 = 0.2$, $J_3 = 0.05$, $J_4 = 0.05$, $S = 14$, $R_0 = 0$.

the necessary condition for a steady state:

$$0 = k_0 + k_1 S - \left[k_2 + k_2' E(R)\right] R \tag{4.85}$$

A more detailed study the properties of the system are obtained from the net rate curve shown in Figure 4.68. We plotted (4.84) as function of the response component R. The stability criterion

$$\frac{d}{dR}\frac{dR}{dt} < 0$$

is fulfilled by the solid lines. For the dashed lines we have $d/dR\ dR/dt > 0$. Analog to the discussion in the previous section, the system exhibits instabilities. The derivative

$$\frac{d}{dR}\frac{dR}{dt} = -\left[k_2 + k_2' E(R) + k_2' R \frac{dE(R)}{dR}\right] \tag{4.86}$$

is independent of the stimulus S. As described above, the stability criterion has to be fulfilled for a stable steady state. Otherwise, the state is referred to as unstable.

Up to now, we only discussed the difference between the stable and the unstable steady state solutions. We pointed out, that the extremum,

$$\frac{d}{dR}\frac{dR}{dt} = 0\ ,$$

of the net rate, limits the unstable range. But, for the critical point we gave no further conditions. We want to do this now but before that we repeat a restriction to our solution. It has to be physically relevant and thus the solution has to be positive,

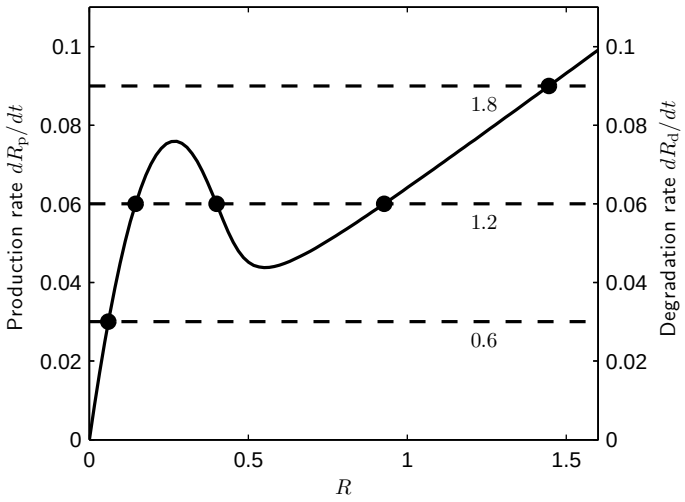


Figure 4.67: The rate curve of a mutually inhibited system (4.84). The solid line is the rate of degradation. The dashed lines are the R-independent production rates for different signal strength S. Note, the change of the number of intersections (black dots) as a function of the signal.

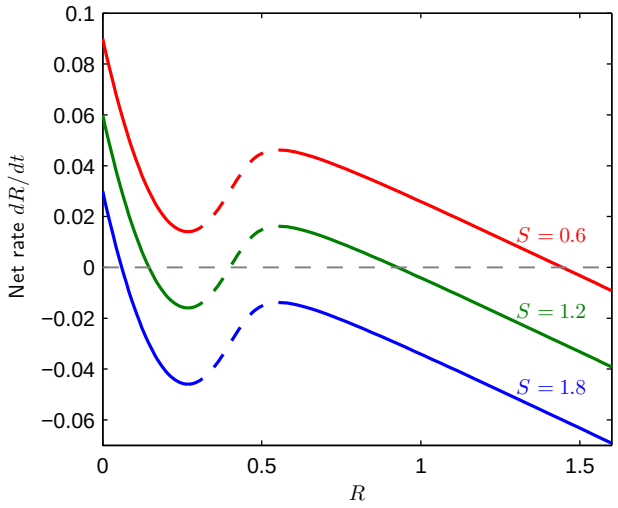


Figure 4.68: The net rate curve of a mutually inhibited system (4.84) as a function of the response component for different signal strengths. The solid line are stable ranges of the net rate, the dashed lines correspond to unstable regions. The intersection between the net rate and the zero line are solutions of the balance equation (4.85).

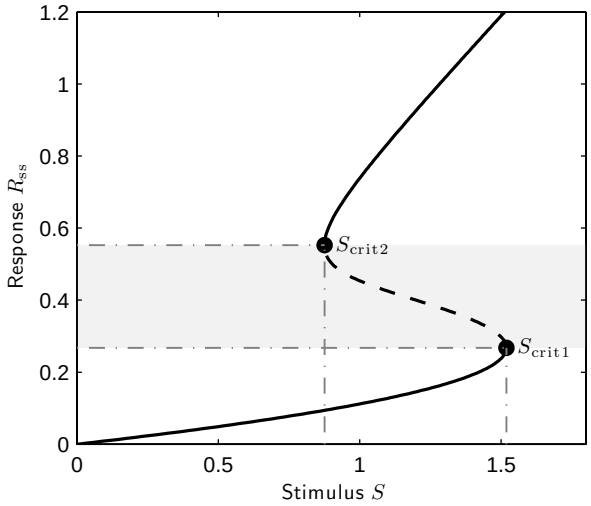


Figure 4.69: The stimulus-response curve for a mutually inhibited system (4.84). The system is partial bistable. The stable branches are drawn as solid lines and the unstable as dashed lines. The critical points S_{crit1} and S_{crit2} limit the unstable solution. The horizontal and vertical dashed lines corresponds to the signal and response strength of the points. The forbidden range of the steady state response is denoted as grey box. Parameter: $k_0 = 0$, $k_1 = 0.05$, $k_2 = 0.06$, $k_2' = 0.5$, $k_3 = 1$, $k_4 = 0.2$, $J_3 = 10$, $J_4 = 0.01$.

i.e., some quantity proportional to the molecule number (e.g. concentration, density, ...). From the net rate plot we obtain the critical point as zeros of the equation that describes the reaction rate as function of the signal strength. From the balance equation (4.85), it is this point that is limiting the real solution to the equation. The condition is fulfilled for two points of the mutually inhibited system (4.84), one minimum and one maximum. The bifurcation points of the mutually inhibited system both satisfy the physical restrictions. We could expect, that the stimulus response curve for this system has two critical points. The stimulus-response curve is shown in Figure 4.69. The stables branches of the balance equation (4.85) are drawn as solid lines. The system is bistable for $S_{crit2} < S < S_{crit1}$ and monostable for all other stimuli. The dashed lines are the unstable solutions limits, described by the critical points S_{crit1} and S_{crit2}. If we increase the strength of the stimulus, starting from the upper level, the steady state response jumps at the first critical point to a high level output. If the stimulus decreases later, the response decreases accordingly. In contrast to the one-way switch in the previous section, the response now goes back to low level if the signal strength is smaller than the second critical point. It is for this reason that we call this system a *buzzer*. The monostable solutions are reversible, the response component is uniquely determined by the signal strength.

buzzer

Next, we discuss temporal properties of system (4.84). We assume that the activation of the enzyme is a much faster reaction than the conversion of the signal component S such that we can describe it with the Goldbeter-Koshland function. We first investigate the behavior of the response to a stimulus between the critical points as a function of the

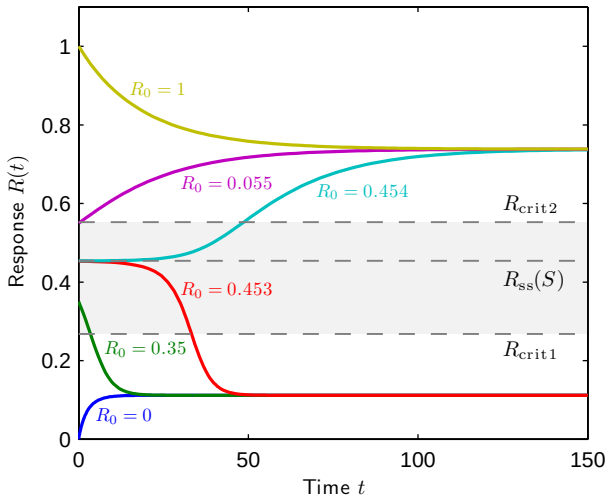


Figure 4.70: The temporal evolution of the mutually inhibited module (4.84), dependent on the initial response signal R_0. We choose a signal strength between the two critical points and hence the system is bistable. The unstable solution of the balance equation (4.85) is a separatrix (dashed horizontal line). For initial states greater than this value the system tends to the upper steady state. In the other case the lower steady state is reached. At the corresponding response value of the critical points (dashed horizontal lines) the response signal changes from unstable to stable behavior. Parameters: $k_0 = 0$, $k_1 = 0.05$, $k_2 = 0.06$, $k_2' = 0.5$, $k_3 = 1$, $k_4 = 0.2$, $J_3 = 10$, $J_4 = 0.01$ and $S = 1$.

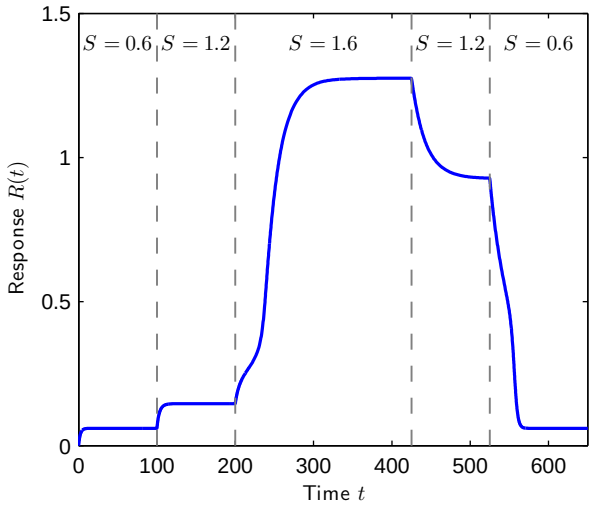


Figure 4.71: Temporal response for the mutually inhibited system (4.84), given a step-like time dependent stimulus S. The steady state for critical signals (here $S = 1.2$) is dependent on the previous state. In the case of subcritical and supercritical stimuli the steady state is uniquely determined by the signal strength. Parameters: $k_0 = 0$, $k_1 = 0.05$, $k_2 = 0.06$, $k_2' = 0.5$, $k_3 = 1$, $k_4 = 0.2$, $J_3 = 10$, $J_4 = 0.01$.

initial response component R_0. For subcritical values $S < S_{\text{crit2}}$ and supercritical $S > S_{\text{crit1}}$ the system is monostable. The response signal moves to the unique steady state for a constant stimulus. The numerical results are plotted in Figure 4.70. Dependent on the initial value, the response signal evolves to the lower or upper steady state. The unstable stimulus dependent solution is again a separatrix. If $R_0 > R_{\text{ss}}(S)$ the upper branch is reached, otherwise the lower branch. Figure 4.71 shows the temporal evolution of the system for successive step-like stimuli of different strengths. Again, we let relax the system to the corresponding steady state and change the stimulus thereafter. In Figure 4.73 the influence of the duration of stimulus is investigated. We start with an initial value $R_0 = 0$ and an external signal $S = 0.6$. This is a subcritical stimulus and the response signal keep on the lower branch. We then increase the stimulus to a critical strength of $S = 1.2$. The response signal remains on the lower level. With the following supercritical signal we force the system to the upper branch of the steady state response. After some time we switch back to the critical signal. As expected, the system now settles to a steady state on the upper branch. The system remembers its previous state. The sub critical stimulus brings the system back on the lower branch. Finally, we switched off the stimulus and the response signal returns to zero.

The temporal evolution of the response component R for different signal strength is shown in Figure 4.72. We choose an example for each range of the stimulus strength. Further on we investigate the behavior for two initial values of the response signal $R_0 = 0$ (solid lines) and $R_0 = 0.7$ (dashed lines). The second initial value is above the unstable region of the steady state response shown as grey box. The limiting horizontal lines denotes R_{crit1} and R_{crit2}. In both situations the response component settles to the same steady state if we apply a sub- and a supercritical signal. For a critical stimulus the steady state depends on the initial state. The response component changes its behavior within the response gap. In this region it is unstable and does never settle to a limiting value. The value of the inflection points is given by the critical points.

Finally, in Figure 4.73 we vary the duration of a step-like signal. We start with the high level response signal $R_0 = 1.1$ and the subcritical external stimulus $S = 0.5$. After the time Δt we switch back to the critical stimulus $S = 1.4$. It follows our previous discussion, that the response component R is decreasing to its lower level steady state if the stimulus is constant. This is clearly seen for long signal durations, for instance $\Delta t = 45$. If we switch to a critical stimulus the system will go to a steady state. Which one, depends on Δt. Only, if the response signal falls below the separatrix, the lower branch of the bistable system (Figure 4.69) defines the new steady state. As long as the response function has not enough time to do this, the system returns to the high level response. We found a similar behavior for the mutually activated system (4.81) in Figure 4.66. A stimulus greater than the critical stimulus value is not enough to change the properties of the system. The duration of the signal must be long enough. The separatrix works like a filter. Fluctuations are suppressed and do not lead to a change in the behavior of the system.

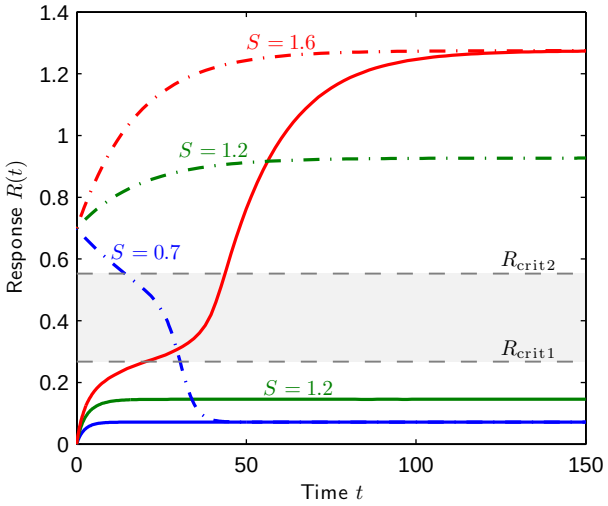


Figure 4.72: Temporal evolution of the response signal for different signal strengths and initial values. The smallest signal is subcritical, the next is critical and the last is supercritical. Independent on the initial state, the system settles to the same steady state for the subcritical and analogue for the supercritical stimulus. For the critical signal the system shows again memory and reaches two different steady states. Parameters: $k_0 = 0$, $k_1 = 0.05$, $k_2 = 0.06$, $k_2' = 0.5$, $k_3 = 1$, $k_4 = 0.2$, $J_3 = 10$, $J_4 = 0.01$.

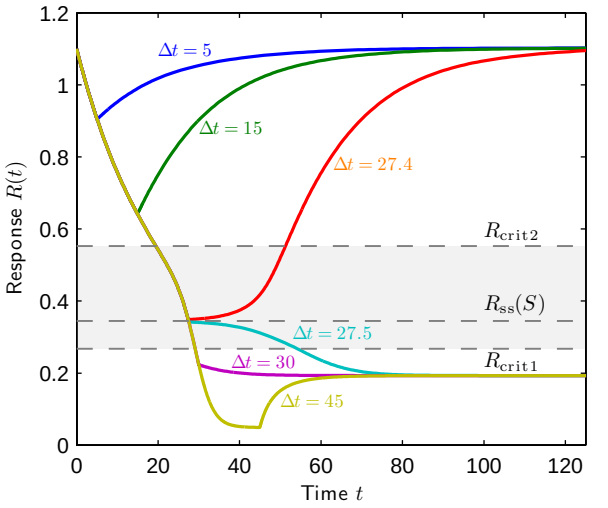


Figure 4.73: Temporal evolution of the mutually inhibited module (4.84) for a time dependent subcritical signal $S = 0.5$ with varying duration. After the time Δt we apply the critical stimulus $S = 1.4$. Note, the final steady states depends on the duration of the subcritical stimulus. Only if Δt is long enough, the response signal falls below the separatrix given by the unstable solution $R_{\mathrm{ss}}(1.4)$. After a short signal the system returns to its high level response. Parameters: $k_0 = 0$, $k_1 = 0.05$, $k_2 = 0.06$, $k_2' = 0.5$, $k_3 = 1$, $k_4 = 0.2$, $J_3 = 10$, $J_4 = 0.01$, $R_0 = 1.1$.

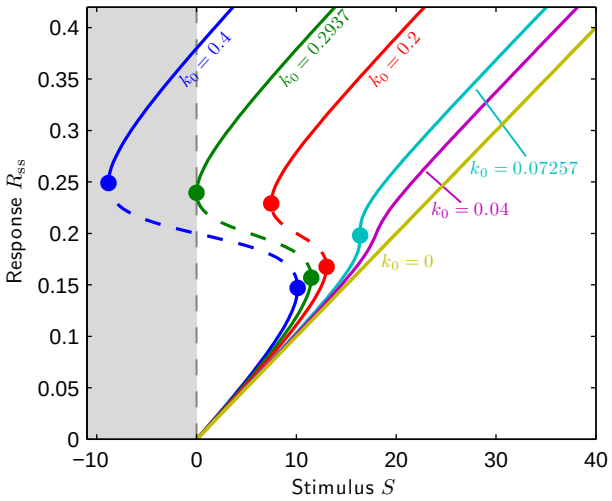


Figure 4.74: Bistability of the mutually activated system, dependent on parameter k_0. The stable branches are drawn as solid lines, unstable as dashed lines. The corresponding critical points are shown as circles. With decreasing k_0 the system changes from an irreversible to a reversible switch. The asymptotic system $k_0 = 0$ is linear. Parameters: $k_1 = 0.01$, $k_2 = 1$, $k_3 = 1$, $k_4 = 0.2$, $J_3 = 0.05$, $J_4 = 0.05$.

Dependency on parameters

The previous two sections discussed two systems with positive feedback/feedforward mechanisms, leading to bistability. In the present section we investigate the dependency of this special systems property on the rate coefficients k_i. As an example for the two systems, let us consider the mutually activated system (4.81) and vary the coefficient k_0. The numerical simulations are shown in Figure 4.74. For a better comparison we extend the graphical representation to negative values of the external signal strength S. Remember, the signal strength is per definition a positive definite quantity. Especially, if we consider biochemical networks the external signal is a concentration of molecules.

For $k_0 = 0$ we have the system discussed in the previous section. The system behaves like an irreversible switch, once activated the system never return to the inactivated state. But if we extend our calculations to negative signals we obtain the same hysteresis-like behavior as for the mutually inhibited system. What happens, if we change the parameter k_0? In Figure 4.74 we show an example investigation.

Beginning with $k_0 = 0.4$, the situation is equivalent to Figure 4.63. For the second curve, k_0 is chosen such that the second critical point is reached at $S = 0$. For $k_0 = 0.2$, the mutually activated system has a hysteresis-like behavior, equivalent to (4.84) (Figure 4.69). By changing one parameter we therefore alter the system from an irreversible switch to a reversible. If we further decrease k_0 the critical points coincide. The system is then continuous. For smaller values of the rate coefficient k_0 no more critical points

exist. For $k_0 = 0$ the system (4.81) is linear:

$$\lim_{k_0 \to 0} R_{\mathrm{ss}} = \frac{k_1}{k_2} S \ . \tag{4.87}$$

Equation (4.87) represents an approximation for small signal strengths S. In the limit $E(R)$ is a small quantity. The product of k_0 and $E(R)$ is negligible in comparison to the remaining terms in (4.81). The then linear system has a signal-response curve corresponding to (4.87).

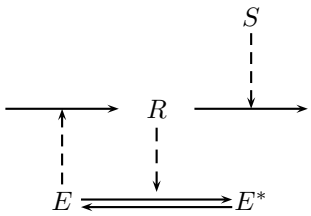
The simple investigation of the properties of a mutually activated system illustrates a major problem in modeling biochemical networks: Often a behavior can be realised by more than one kinetic model. The falsification of these models is usually not possible with kinetic methods alone. The concentration of some hypothetic intermediates is not measurable with direct kinetic methods. As we have seen, by changing one rate coefficient, the behavior of the system change dramatically. In biochemical networks the coefficients depend on properties such as temperature, volume and pressure.

4.9.5.2 Negative feedback - Oscillations

According to the definition of Section 4.9.5, negative feedback means the response counteracts the effect of the stimulus. There are two ways in which a negative influence can be exerted: through an acceleration of degradation and a deceleration of production of the response component.

Homoeostasis

In *homoeostasis*, the response on an external signal (stimulus) is approximately constant over a wide range of signal strength. This behavior may also be described as a kind of imperfect adaption. In contrast to the perfectly-adapted system (4.80) the response component is not changed in response to step-change of the stimulus S (cf. Figure 4.59): homoeostasis



Such a system can be described by the coupled system of differential equations

$$\left.\begin{aligned} \frac{dR}{dt} &= k_0 E - k_2 S R \\ \frac{dE}{dt} &= \frac{k_3\left(1-E\right)}{J_3+1-E} - \frac{k_4\, RE}{J_4+E} \end{aligned}\right\} \tag{4.88}$$

where the response component R inhibits the enzyme catalyzing its synthesis. In (4.88) E is normalised to the total enzyme concentration E_T.

If ones assumes, that the enzyme production reaches its steady state much faster than the whole system, we can simplify (4.88) using (4.79). The enzyme concentration is now

$$\left.\begin{aligned} \frac{dR}{dt} &= k_0 E(R) - k_2 S R \\ \frac{dE(R)}{dt} &= G(k_3, k_4 R, J_3, J_4) \end{aligned}\right\} \tag{4.89}$$

A comparison of production and degradation rate is shown in Figure 4.75.

The rate of production

$$\text{rate of production} = k_0 E(R)$$

implicitly depends from the response component R. The result is a sigmoidal curve. The degradation rate

$$\text{rate of degradation} = k_2 S R$$

is a linear function with slope $k_2 S$. For those chosen range, it is assumed that the steady state response is nearly independent from the external signal. The net rate, shown in

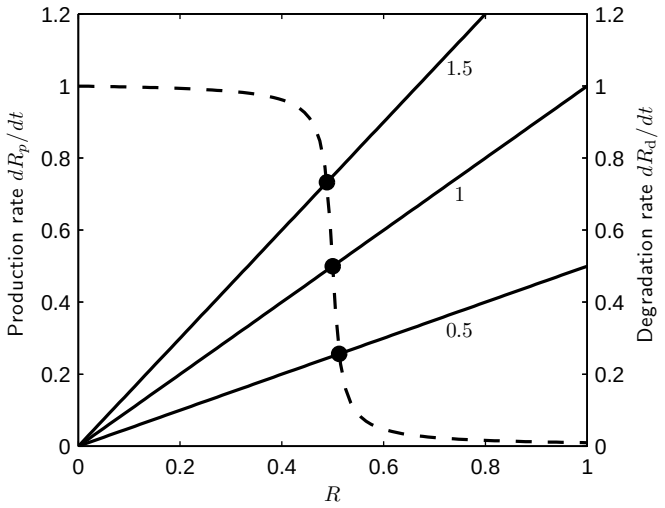


Figure 4.75: Comparison of rate of response and degradation for the homoeostatic system (4.89) for stimulus S. The solid lines are the degradation rates, (4.89), i.e., linear functions with slope $k_2 \cdot S$. The rate of response depends on R and has the typical sigmoidal shape of (4.79). The intersections, denoted by dots, are again steady state solutions. Parameters: $k_0 = k_2 = 1$, $k_3 = 0.5$, $k_4 = 1$, $J_3 = J_4 = 0.01$, $E_T = 1$.

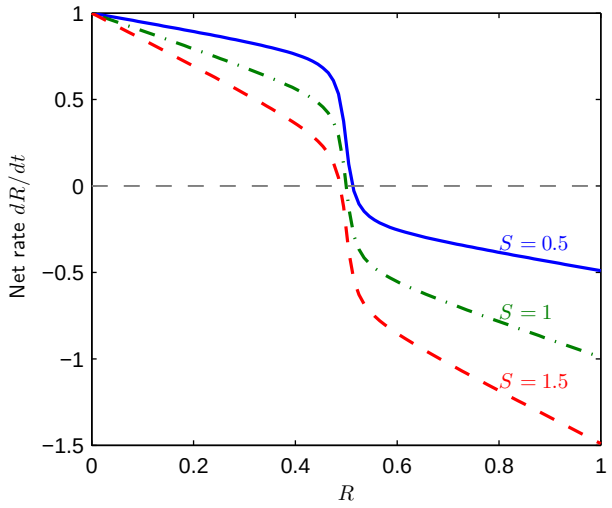


Figure 4.76: Net rate for the homoeostatic system (4.89) for different external stimuli. The intersections with the horizontal dashed line are the corresponding steady states. The steady states are only weakly dependent on the stimulus S. Parameters: $k_0 = k_2 = 1$, $k_3 = 0.5$, $k_4 = 1$, $J_3 = J_4 = 0.01$, $E_T = 1$.

Figure 4.76, displays an analogues behavior. Again, we compare the influence of different stimuli to the overall rate. The intersections with the dashed line, where the rate is zero,

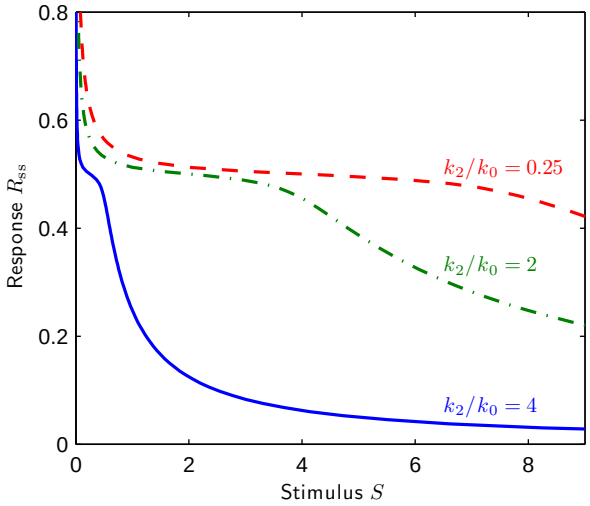


Figure 4.77: Stimulus-response curve for system (4.89) as a function of different ratios of rate coefficients k_0 and k_2. The steady state response shows, for small ratios, a plateau over a range of the signal S. The plateau decreases with increasing ratio and eventually disappears. Parameters: $k_0 = 1$, $k_3 = 0.5$, $k_4 = 1$, $J_3 = J_4 = 0.01$, $E_T = 1$.

represent the steady states. The corresponding balance equation

$$0 = k_0 E(R) - k_2 SR \tag{4.90}$$

can be transformed into

$$\frac{k_2}{k_0} S = \frac{E(R)}{R} \tag{4.91}$$

A solution for the dependence on the response component can only be found numerically. On the other hand, (4.91) gives a rule to calculate the strength of the stimulus S for known R. This implies that the ratio of the rate coefficients k_0 and k_2 plays an important role for the behavior of the considered system. The corresponding signal-response curve is shown in Figure 4.77, where we the ratio k_2/k_0 is varied. The ratio determines the length of the expected plateau. The amplitude is weakly affected. For small signals the steady state response is singular and for strong signals it tends to zero.

For a discussion of steady state properties for a homoeostatic system we neglected the time dependence of the enzyme activation/deactivation reaction in (4.88). For the study of the temporal behavior of such a system we do not make the assumption of a much faster reversible enzymatic reaction. Before we continue this, let us return to the steady state properties, derived from the coupled balance equations

$$\begin{aligned} 0 &= k_0 E - k_2 SR \\ 0 &= \frac{k_3 \left(E_T - E\right)}{J_3 + E_T - E} - \frac{k_4 RE}{J_4 + E} \end{aligned}$$

The solution of the enzymatic equation is again the Goldbeter-Koshland function we used in (4.89). After insertion into the first equation we obtain again (4.91). The

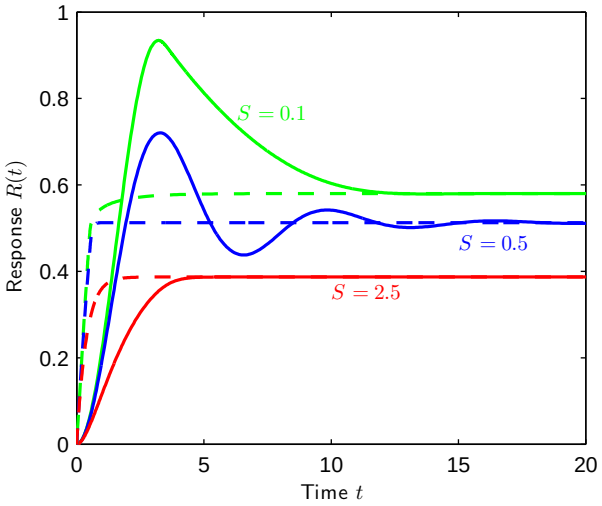
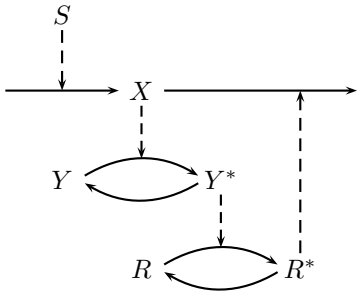


Figure 4.78: Comparison of temporal evolution of homoeostatic system (4.88) and its simplified version (4.89) for different signal strengths S. The solutions of the full system are drawn as solid lines, while dashed lines are used for the simplified case. Parameters: $k_0 = k_2 = 1$, $k_3 = 0.5$, $k_4 = 1$, $J_3 = J_4 = 0.01$, $E_T = 1$.

steady state solution of a coupled system of equations remains therefore unaffected. We thus expect, that the system will reach the same steady state in its temporal evolution. In Figure 4.78 we plot numerical solutions of (4.88) for the response component and compare it with solutions of (4.89). The differences in temporal behavior can be divided into three classes, corresponding to the three ranges in the stimulus-response curve. In the homoeostatic range, the system displays damped oscillations around the steady state. Common for all three cases is a difference in the relaxation time. The system (4.88) takes longer than (4.89) to reach the steady state. The simplification made in (4.89) makes it easier to handle the differential equations and does not affect the steady state. On the other hand, no oscillations occur with this approximation. Nevertheless, the assumption of a much faster process is often used to simplify the treatment.

Negative feedback oscillator

The previous section showed how negative feedback can introduce damped oscillations. We here show how negative feedback can lead to stable oscillations. Therefore we consider a system of three components



The mathematical model for this system is defined by the following equations

$$\left.\begin{aligned}\frac{dX}{dt} &= k_0 + k_1 S - \left(k_2 + k_2' R^*\right) X \\ \frac{dY^*}{dt} &= \frac{k_2 X\,(Y_T - Y^*)}{K_{\mathrm{M3}} + Y_T - Y^*} - \frac{k_4 Y^*}{K_{\mathrm{M4}} + Y^*} \\ \frac{dR^*}{dt} &= \frac{k_5 Y^*\,(R_T - R^*)}{K_{\mathrm{M5}} + R_T - R^*} - \frac{k_6 R^*}{K_{\mathrm{M6}} + R^*}\end{aligned}\right\} \tag{4.92}$$

where X activates the protein Y. The activated protein Y^* activates the next protein R. Its activated form catalyzes the degradation of X. Another possible way to close the negative feedback loop is the inhibition of production from S. We focus on the first case. A numerical simulation of (4.92) is given in Figure 4.79. All three components show oscillations. The third component Y^* introduces a time delay in the feedback loop, causing the control system repeatedly to over- and undershoot its steady state. Within the shown interval the oscillation are damped. The system takes a certain time to establish stable oscillations, depending on the chosen set of parameters. For instance, if one increases the rate coefficient k_2' to 20, the amplitude will initially increase until the limit cycle is reached. A variation of k_0 influences the strength of damping or amplifying and the amplitude of the limit cycle. In the phase-plane representation, Figure 4.80, it is shown, how the three components settle towards a limit cycle.

Let us now focus on steady state properties. The state steady state are derived from the set of coupled balance equations

$$\begin{aligned} 0 &= k_0 + k_1 S - \left(k_2 + k_2'\, R^*\right) X \\ 0 &= \frac{k_2 X\,(Y_T - Y^*)}{K_{\mathrm{M3}} + Y_T - Y^*} - \frac{k_4 Y^*}{K_{\mathrm{M4}} + Y^*} \\ 0 &= \frac{k_5 Y^*\,(R_T - R^*)}{K_{\mathrm{M5}} + R_T - R^*} - \frac{k_6 R^*}{K_{\mathrm{M6}} + R^*} \end{aligned} \tag{4.93}$$

The balance equation for each subsystem itself is solvable but the steady state solution of the coupled system for the response component is lengthy and complicated. Therefore we prefer a numerical solution of (4.93). The corresponding stimulus-response curve is plotted in Figure 4.81. The solution is separated into two stable (solid line) and an unstable range (dashed line). In the unstable range the system performs stable oscillations. The amplitude of the oscillation depends on the stimulus S as shown by the dashed-dotted curve.

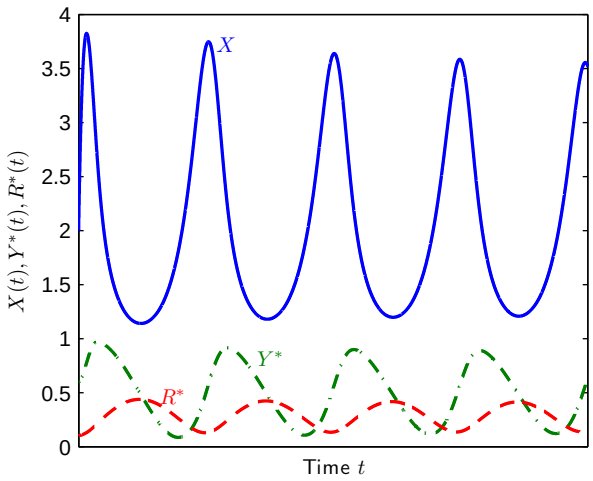


Figure 4.79: The temporal evolution of the negative feedback oscillator (4.92). All three components X, Y^*, and R^* perform oscillations. Within the shown time interval these oscillations are damped. Parameters: $k_0 = 0$, $k_1 = 1$, $k_2 = 0.01$, $k_2' = 10$, $k_3 = 0.1$, $k_4 = 0.2$, $k_5 = 0.1$, $k_6 = 0.05$, $K_{M3} = K_{M4} = K_{M5} = K_{M6} = 0.01$, $Y_T = R_T = 1$, $S = 5$, $X_0 = 2$, $Y^* = 0.6$, $R^* = 0.1$.

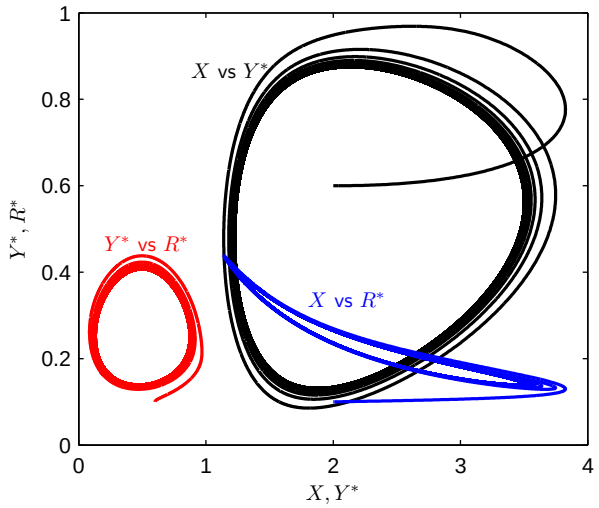


Figure 4.80: Phase-plane representation of numerical solution of (4.92). We combine three possible combinations in this figure. All components perform oscillations and tend to the limit cycle within a certain number of oscillations. Parameters: $k_0 = 0$, $k_1 = 1$, $k_2 = 0.01$, $k_2' = 10$, $k_3 = 0.1$, $k_4 = 0.2$, $k_5 = 0.1$, $k_6 = 0.05$, $K_{M3} = K_{M4} = K_{M5} = K_{M6} = 0.01$, $Y_T = R_T = 1$, $S = 5$, $X_0 = 2$, $Y^* = 0.6$, $R^* = 0.1$.

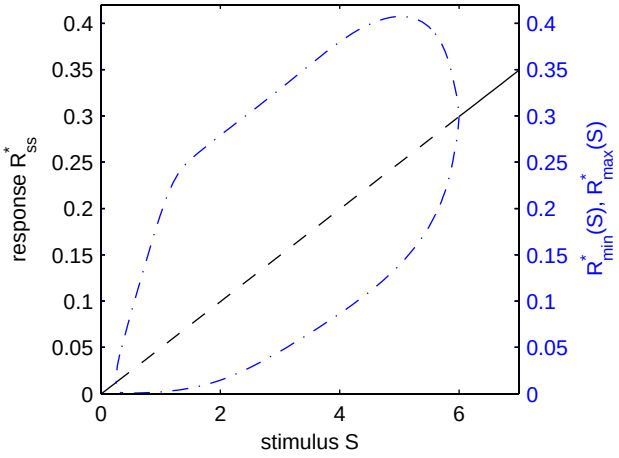


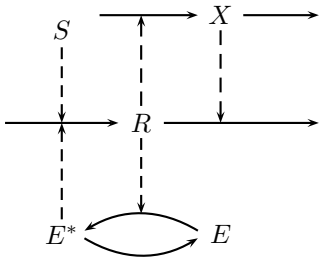
Figure 4.81: Stimulus-response curve as function of the external signal strength S for the negative feedback oscillator (4.92). The straight line is the numerical steady state solution of (4.93). The solid parts denote stable ranges. In the interval $0.26 < S < 6$, the solution is unstable and the system performs oscillations, as in Figure 4.79. The maxima and minima as a function S are plotted as the dash-dotted curve. Parameters: $k_0 = 0$, $k_1 = 1$, $k_2 = 0.01$, $k_2' = 10$, $k_3 = 0.1$, $k_4 = 0.2$, $k_5 = 0.1$, $k_6 = 0.05$, $K_{M3} = K_{M4} = K_{M5} = K_{M5} = 0.01$.

4.9.5.3 Mixed control mechanisms

In the present section, different feedback mechanisms are combined.

Activator-Inhibitor-Oscillator

Our first example combines the mutually activated system (4.81) and an autoinhibition of the response component.



The response component R is produced in an autocatalytic process. It activates the enzyme E^* which accelerate the production of R. On the other hand, the response component promotes the production of the inhibitor X at the same time. The inhibitor speeds up the removal of R. Again, we assume that the enzyme is always in its steady state described by the relation (4.79), assuming the activation/deactivation process is much faster than the other reactions in the system. This assumption simplifies our further discussion and restrict the (mathematical) dimension of the corresponding system

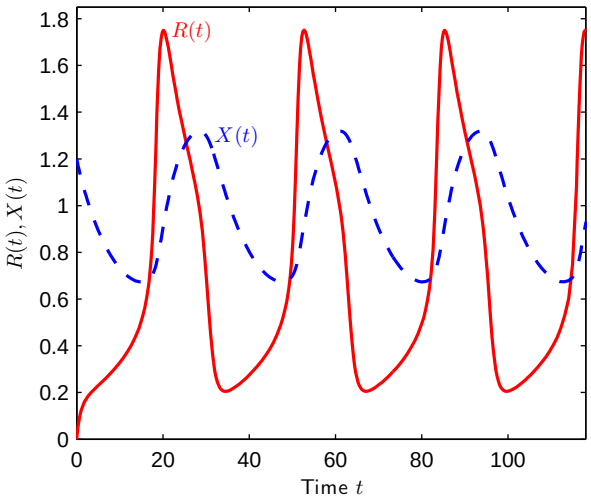


Figure 4.82: The temporal evolution of system of an activator and an inhibitor (4.94) as function of time. The solid line is the response component and the dashed line the inhibitor. After a short starting time the system carries out stable oscillations. Parameters: $k_0 = 4$, $k_1 = k_2 = k_2' = k_3 = k_4 = 1$, $k_5 = 0.1$, $k_6 = 0.075$, $J_3 = J_4 = 0.3$, $S = 0.2$, $R_0 = 0$, $X_0 = 1.2$.

of differential equations

$$\begin{aligned} \frac{dR}{dt} &= k_0 E^*(R) + k_1 S - \left(k_2 + k_2' X\right) R \\ \frac{dX}{dt} &= k_5 R - k_6 X \\ E^*(R) &= G(k_3 R, k_4, J_3, J_4) \end{aligned} \tag{4.94}$$

For this composed system an analysis in terms of rate of production/degradation and the net rate is only possible in three dimensions[12]. The assumption of the steady state for the enzyme avoids an additional fourth dimension. The temporal evolution of the response signal $R(t)$ and the inhibitor $X(t)$ is numerically solvable from the coupled differential equations (4.94). For the chosen set of parameters we obtain stable oscillations in Figure 4.82 for both components. If the amount of R small, the production of the response component is the main process. The degradation of the inhibitor is faster than its production. This results in an increase of the response component. With increasing R also the production of the inhibitor is increasing. The acceleration of the degradation of R leads to a decrease of the response component, returning us to where we started.

The phase plane representation of the oscillations is shown in Figure 4.83. Additionally to the limit cycle of the oscillation we plotted the steady states as a function of the response component R and X. The steady states are obtained from the balance equations

$$\frac{dR}{dt} = k_0 E^*(R) + k_1 S - \left(k_2 + k_2' X\right) R = 0 \tag{4.95}$$

[12] One for the response component R, one for the rate of change of the response component, and one for the rate of change of the inhibitor X.

and

$$\frac{dX}{dt} = k_5 R - k_6 X = 0 \tag{4.96}$$

Equation (4.96) for the inhibitor can be solved analytically leading to a straight line with slope k_6/k_5:

$$R = \frac{k_6}{k_5} X \ . \tag{4.97}$$

In Figure 4.83, the closed curve is the phase-plane representation of the temporal evolution shown in Figure 4.82. After an initial time the system reaches a stable limit cycle. The straight line and the line that is in parts dashed, are the stimulus-response curves of the subsystems. The straight line is the analytic solution (4.97). The solution of (4.95) is numerically found. The stable solutions are shown as solid sections of the line and the unstable solution as a dashed line. The corresponding critical points are shown as filled dots. The solutions of the balance equations (4.95) and (4.96) have an intersection, which is the steady state solution shown as a dot on the straight line. For the chosen parameters the steady-state is unstable leading to an oscillating behavior. As one subsystem tries to reach one of its two stable states, the resulting production or degradation of X forces it back to the unstable state and the cycle starts again. The intersection depends on the signal strength S. By an increase or decrease of S we move the solution of (4.95) in Figure 4.83 until the intersection is stable. For the corresponding signal strengths the activator-inhibitor system has an stable steady state. No more stable oscillations occur. Such a situation is shown in Figure 4.84. At the intersection both solutions are stable, although near the critical point. The system (4.94) shows damped oscillations and reaches a stable steady state. The phase plane shows a typical spiralling curve with decreasing amplitude. For still higher signal strengths the intersection moves further away from the critical point and the damped oscillations will disappear.

As discussed above, the oscillatory behavior of the considered system (4.94) strongly depends on the strength of external signal S. Oscillations occur only if the intersection between the steady states of each subsystem is on the unstable branch of (4.95). The parameter we can change is the external signal S. The internal parameter we assume as inherent, uniquely determined by conditions like temperature, pH-value and so on. The stimulus-response curve shows a new qualitative property. There are intervals in the signal strength S, where the system tends to a steady state. For the chosen values of the rate coefficients, see Figure 4.85, this behavior is established in the intervals $0 < S < 0.066$ and $S > 0.41$. In the intermediate range the response signal R oscillates around the unstable steady state, shown as a dashed line in Figure 4.85. The amplitude and minimal ($R_{\min}$) and maximal ($R_{\max}$) values are functions of the stimulus signal S.

Substrate-Depletion-Oscillator

As second example of an oscillating mixed system we choose a substrate-depletion-oscillator.

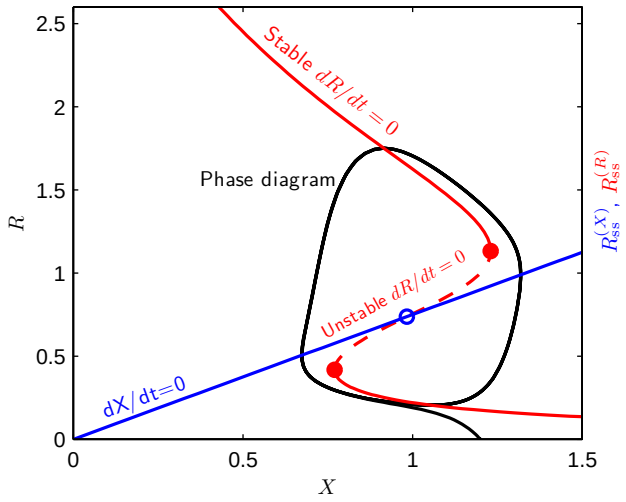


Figure 4.83: Phase-plane representation and stimulus-response curve for the activator-inhibitor system (4.94) with a constant external signal strength. Parameters: $k_0 = 4$, $k_1 = k_2 = k_2' = k_3 = k_4 = 1$, $k_5 = 0.1$, $k_6 = 0.075$, $J_3 = J_4 = 0.3$, $S = 0.2$, $X_0 = 1.2$, $R_0 = 0$.

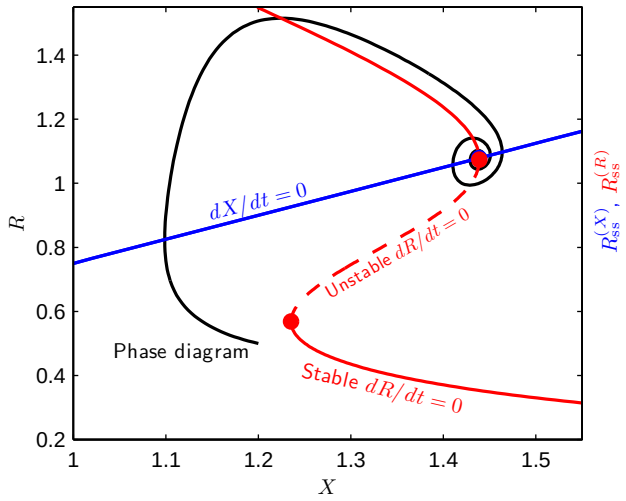


Figure 4.84: Phase-plane representation and stimulus-response curves for the activator-inhibitor subsystems (4.95)-(4.96). In contrast to Figure 4.83 the system reaches a stable steady state. Starting with R_0 the system has damped oscillations around this stable state and ends finally in it. Parameters: $k_0 = 4$, $k_1 = k_2 = k_2' = k_3 = k_4 = 1$, $k_5 = 0.1$, $k_6 = 0.075$, $J_3 = J_4 = 0.3$, $S = 0.43$, $X_0 = 1.2$, $R_0 = 0.5$.

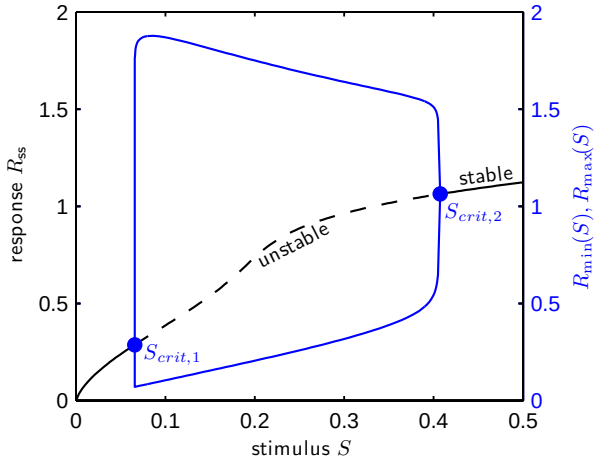
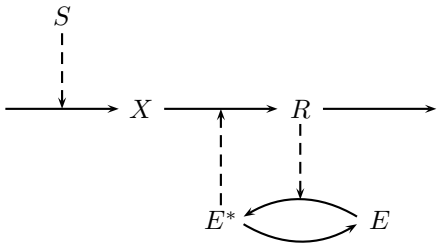


Figure 4.85: The stimulus-response curve for the system (4.94). The steady state solutions of the coupled system of equation (4.95) and (4.96). The stable solutions are drawn as solid and the unstable solution as the dashed section of the line. The closed curve shows the maximal and minimal values of the response signal in the case of stable oscillations. On the critical points $S_1 \approx 0.066$ and $S_2 \approx 0.41$ the system changes is behavior abruptly. Parameters: $k_0 = 4$, $k_1 = k_2 = k_2' = k_3 = k_4 = 1$, $k_5 = 0.1$, $k_6 = 0.075$, $J_3 = J_4 = 0.3$.



The signalling component S is converted into the intermediate X. From the intermediate the response component R is produced. On the other hand the response component activates the enzyme E^*, increasing the conversion rate from X to R. For such a reaction scheme we obtain the following system of coupled differential equation

$$\begin{aligned} \frac{dX}{dt} &= k_1 S - \left(k_0' + k_0\, E^*(R)\right) X \\ \frac{dR}{dt} &= \left(k_0' + k_0 E^*(R)\right) X - k_2 R \\ E^*(R) &= G(k_3 R, k_4, J_3, J_4)\ . \end{aligned} \tag{4.98}$$

We again assume that the activation/deactivation of the enzyme is much faster than the other reaction. Hence, the enzyme E^* is assumed to be always in a steady state. The corresponding steady state solution is then (4.79). In Figure 4.86 the numerical simulation of (4.98) is shown. For the chosen set of parameters the system displays stable oscillations. First, the amount of the intermediate X increases faster than it is converted into the response component R. But R promotes its own production via the enzyme E^*. According to the sigmoidal shape of (4.79), the positive feedback term $k_0 E^*$

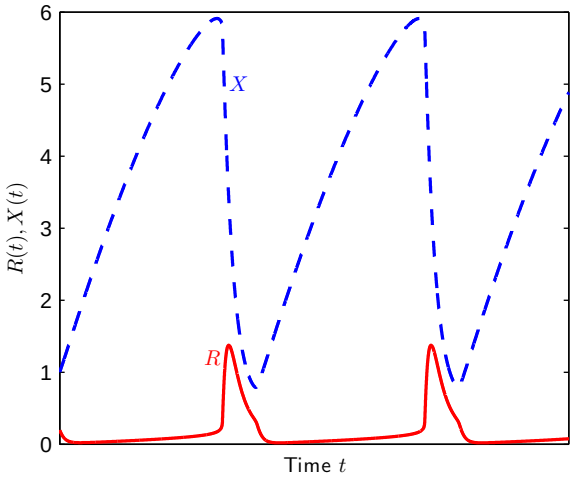


Figure 4.86: The oscillatory behavior of the system (4.98). The response signal $R(t)$ is drawn as solid line and the intermediate $X(t)$ as a dashed line. Parameters: $k_0 = 0.4$, $k_0' = 0.01$, $k_1 = k_2 = k_3 = 1$,$k_4 = 0.3$, $J_3 = J_4 = 0.05$, $S = 0.2$, $X_0 = 1$, $R_0 = 0.2$.

is small in comparison to the direct transformation with k_0'. The response component will increase approximately linear. If R exceeds a critical value the enzyme concentration jumps to a high value. The conversion from X into R is now determined by the reaction of the activated enzyme and the intermediate. This reaction is faster than the production of X. The result is a strong increase of the response component until the intermediate is depleted. Then the fast production of R breaks down. On the other hand, the decay of the response component is now faster as its slow production and its concentration decrease. In the mean time the amount of X increase again.

Next we investigate the system (4.98) in more detail. We start with a computation of the steady state, solving the coupled system of balance equations

$$0 = k_1 S - \left(k_0' + k_0 E^*\right) X \ , \tag{4.99}$$

$$0 = \left(k_0' + k_0 E^*\right) X - k_2 R \ . \tag{4.100}$$

For the respective subsystems we derive

$$X = \frac{k_1\, S}{k_0' + k_0 E^*(R)} \tag{4.101}$$

from (4.99) and

$$R = \frac{k_0' + k_0 E^*(R)}{k_2}\, X \tag{4.102}$$

for the second equation (4.100). Both solutions are shown in Figure 4.87, together with a phase-plane representation of the temporal evolution in Figure 4.86. Whereas, (4.101) is a monostable function, the stimulus-response curve of the R-subsystem is bistable. This bistability triggers the occurrence of oscillations. Remember, the steady state of the

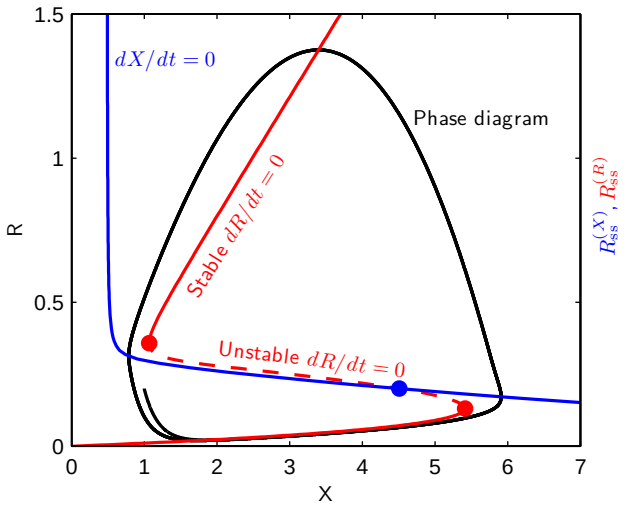


Figure 4.87: Phase-plane representation of the substrate-depletion oscillator (4.98) combined with stimulus-response curves of the subsystems. The solid black line is the limit cycle of the stable oscillations of intermediate X and response component R. The balance equations (4.99) and (4.100) were numerically solved. The steady state solution, the intersection of the stimulus-response curves, is located in the unstable branch of the response component, the system therefore displays oscillations around this point. Parameters: $k_0 = 0.4$, $k_0' = 0.01$, $k_1 = k_2 = k_3 = 1$, $k_4 = 0.3$, $J_3 = J_4 = 0.05$, $S = 0.2$, $X_0 = 1$, $R_0 = 0.2$.

substrate-depletion oscillator is given by the intersection of both curves. Mathematically we have to derive the balance equations (4.99) and (4.100) simultaneously. The result is the linear function

$$R_{\mathrm{ss}} = \frac{k_1}{k_2} S \ . \tag{4.103}$$

But for a stable steady state all subsystems have to be in a stable steady state. This is not the case for the given set of parameters in Figure 4.87. The system performs oscillations around the steady state solution (4.103).

If we choose other rate coefficients the intersection of (4.101) and (4.102) change its position. Again, in analogy to the activator-inhibitor system (4.94) the state space of substrate-depletion oscillator (4.98) is separated into a region of stable oscillation and a non-oscillating part, where the system tends to a stable steady state. In Figure 4.88 this is illustrated with the stimulus-response representation. The straight line are the steady state solutions (4.103) following from (4.99) and (4.100) as a function of the external signal strength S. Solid parts of the line corresponds to a stable steady state. With the dashed part the system performs stable oscillations around the formal solution. The closed line corresponds to maximal and minimal values of the amplitude. This pictures shows a sharp transition between both behaviors. The value of minimal response component is nearly constant, where the maximum depends on the external signal.

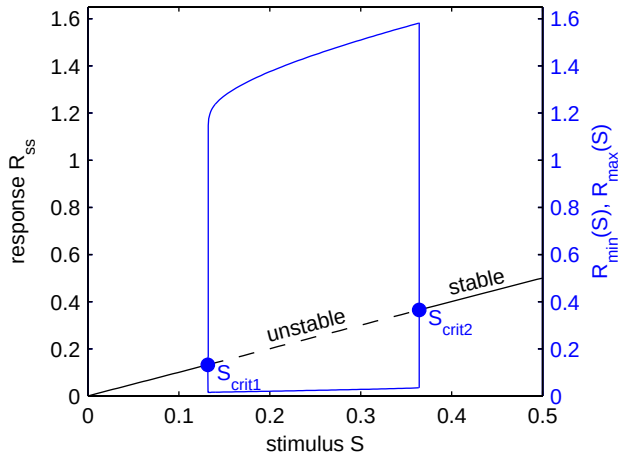


Figure 4.88: Stimulus-response curve for the substrate-depletion oscillator (4.98). The straight line are the steady states. In the interval $0.132 < S < 0.365$ the system has no stable solution of the balance equations, shown as a dashed segment of the line. Under these conditions oscillations occur. The maximal and minimal values of the response component are shown as a closed curve. Again the amplitude of oscillations depends from the signal strength. Outside the unstable region the response component tends to the steady state value. Near the critical points damped oscillations occur. Parameters: $k_0 = 0.4$, $k_0' = 0.01$, $k_1 = k_2 = k_3 = 1$, $k_4 = 0.3$, $J_3 = J_4 = 0.05$.

5 Cell Cycle Modelling

Where a cell arises, there must be a previous cell, just as animals can only arise from animals and plants from plants. This cell doctrine, proposed by the German pathologist Rudolf Virchow in 1858, carried with it a profound message for the continuity of life. Cells are generated from cells, and the only way to make more cells is by division of those already exist. All living organisms, from the unicellular bacterium to the multicellular mammal, are products of repeated rounds of cell growth and division extending back in time to the beginning of life on Earth over three billion years ago.

A cell reproduces by performing an orderly sequence of events in which it duplicates its contents and then divides in two. This cycle of duplication and division, known as cell cycle, is the essential mechanism by which all living things reproduce. In unicellular species, such as bacteria and yeasts, each cell division produces a complete new organism. In multicellular species, long and complex sequences of cell divisions are required to produce a functioning organism. Even in the adult body, cell division is usually needed to replace cells that die. In fact, each of us must manufacture many million of cells every second simply to survive: if all cell division were stopped - by exposure to a very large dose of x-rays, for example - we would die in few days.

The minimum set of processes that a cell has to perform are those that allow it to accomplish its most fundamental task: the passing on of its genetic information to the next generations of cells. To produce two genetically identical daughter cells, the DNA in each chromosome must first be faithfully replicated to produce two complete copies, and the replicated chromosomes must then be accurately distributed (segregated) to the two daughter cells, so that each receives a copy of the entire genome. In addition to duplicating their genetic material, most cells also duplicate their other organelles and macromolecules; otherwise, they would get smaller which each division. To maintain their size, dividing cells must coordinate their growth (i.e., their increase in cell mass) with their division.

5.1 An Overview of the Cell Cycle

For this overview about the cell cycle we follow the representation of Alberts et al. [AJL+02]. The most basic function of the cell cycle is to duplicate accurately the vast amount of DNA in the chromosomes and then segregate the copies precisely into two genetically identically daughter cells. These processes defines the two major phases of the cell cycle. DNA duplication occurs during the S phase (S for synthesis), which requires 10-12 hours and occupies about half of the cell-cycle time in a typical mammalian cell. After S phase, chromosomes segregation and cell division occur in M phase (M for mitosis), which requires much less time (less than an hour in a mammalian cell).

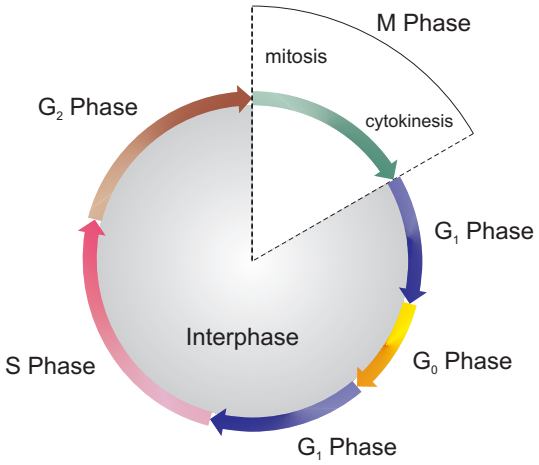


Figure 5.1: The phases of the cell cycle. The cell growths continuously in interphase, which consists of three phases: S phase where the DNA is replicated; G_1 is the gap between M phase and S phase, which can be interrupted by a resting phase G_0. G_2 is the gap between S phase and M phase. In M phase, first the nucleus and then the cytoplasm divide.

Most cells requires much more time to grow and double their mass of proteins and organelles than they require to replicate their DNA and divide. Partly to allow more time for growth, extra gap phases are inserted in most cell-cycles — a G_1 phase between M phase and S phase and a G_2 phase between S phase and mitosis. Thus, the eucaryotic cell cycle is divided into four sequential phases: G_1, S, G_2, and M. G_1, S, and G_2 together are called interphase, see also Fig. 5.1. In a typical human cell proliferating in culture, interphase might occupy 23 hours of a 24 hours cycle, with an 1 hour for M phase.

The two gap phase serve as more than simple time delays to allow cell growth. They also provide time for the cell to monitor the internal and external environment to ensure that conditions are suitable and preparations are complete before the cell commits itself to the major upheavals of S phase and mitosis. The G_1 phase is especially important in this respect. Its length can vary greatly depending on external conditions and extracellular signals from other cells. If extracellular conditions are unfavorable, for example, cells delay progress through G_1 and may even enter a specialised resting state known as G_0, in which they can remain for days, weeks, or even years before resuming proliferation. If extracellular conditions are favorable and signals to grow and divide are present, cells in early G_1 or G_0 progress through a commitment point near the end of G_1. After passing this point, cells are committed to DNA replication, even if the extracellular signals that stimulate cell growth and division are removed. An introduction to underlying modules and control mechanisms is given in Chapter 4.9.

The cell cycle is separated roughly into four phases (excluding an additional resting phase). But Mitosis and S-Phase play a special role in this complex sequence of cellular events. During these phases some remarkable changes in the contents and functionality of the cell are occurring.

5.1.1 Mitosis - Key events

Mitosis is the final step in the process of cell division. But before the cell can divide into two daughter cells some preparing steps are necessary. Mitosis can be separated into four key events as represented [MK89]. In the Prophase, the dispersed duplicated chromosomes are condensed into a state suitable for transport. Additionally, the metabolic activity is reduced. The condensed and inactive chromosomes are position at the equator of the nuclei in the Prometaphase. This phase is followed by the Anaphase, where the duplicated chromosomes are separated into two identical parts. Due to the progress in experimental technique and technology, this can now be observed in experiments. Subsequently, the chromosomes move towards opposite ends of the nuclei. In the Telophase the chromosomes decondensates and become metabolic active. Simultaneously, there is a reformation of the nuclei.

Additionally to the processes inside the nuclei, there are further rearrangements in the cytoplasm providing the cell division. For example, the microtubuli and microfilaments of the cytoskeleton are rearranged, the Golgi apparatus is dispersed into a large number of small vesicles and is distributed throughout the cell. Additionally, the protein synthesis is slow down to about 25% of its normal rate.

5.1.2 S-Phase

As mentioned earlier, the chromosomes are doubled during the S-phase. But the synthesis not only replicates the chromosomes completely and precisely. In addition, the cell also has to duplicate the complex chromosome architecture [LFB89]. Especially, it copies the specific patterns of gene activity and inactivity. For this reason, a specialised cell divides into two daughter cells with the same functionality. A further sub-specialisation of the daughter cell is possible during their life time. Additionally to functional information, also cell cycle specific data are stored and transferred during the S-phase. Eukaryotic cells tag the replicated DNA and distinguish between replicated and unreplicated DNA. Hence, the replication of DNA take place only once between cell divisions. This is a hallmark of eukaryotic cells distinguishing them from prokaryotes. Furthermore the most differentiate cells contains information about the number of replications. It prohibits the infinite division of a cell line. After a certain number of replications the cell cannot duplicate the chromosomes. The cell cycle cannot be finished and internal regulatory mechanisms drive the cell into Apoptosis.

5.2 The Cell-Cycle Control System

Eucaryotic cells have evolved a complex network of regulatory proteins, known as the cell-cycle control system, that governs progression through the cell cycle. The core of this system is an ordered series of biochemical switches that control the main events of the cycle, including DNA replication and the segregation of the replicated chromosomes. In most cells, additionally layers of regulation enhance the fidelity of cell division and allow the control system to respond to various signals from both inside and outside

the cell. Inside the cell, the control system monitors progression through the cell cycle and delays later events until earlier events have been completed. Preparations for the segregation of replicated chromosomes, for example, are not permitted until DNA replication is complete. The control system also monitors the conditions outside the cell. In a multicellular organism, the system is highly responsive to signals from other cells, stimulating cell division when more cells are needed and blocking it when they are not. The cell-cycle control system therefore has a central role in regulation cell numbers in the tissues of the body. When the system malfunctions, excessive cell division can results in cancer.

In principle, one can imagine that the most basic control system should possess the following features:

- A clock, or timer, that turns on each event at a specific time, thus providing a fixed amount of time for the completion of each event.
- A mechanism for initiating events in the correct order; entry into mitosis, for example, must always come after DNA replication.
- A mechanism to ensure that each event is triggered only once per cycle.
- Binary (on/off) switches that trigger events in a complete, irreversible fashion. For example, once initiated the DNA synthesis must be completed.
- Adaptability, so that the system's behavior can be modified to suit specific cell types or environmental conditions.
- Robustness: backup mechanisms to ensure that the cycle can work properly even when parts of the system malfunction.
- Abnormal termination: control mechanisms to ensure that the cycle results in a perfect' copy of the mother cell. For instance, if there are unrecoverable errors in the duplicated genome the cell cycle has to stop definitively and the cell death (apoptosis) must be initiated.

Especially, the required robustness and the ability of an abnormal termination are antagonists in the optimisation process of the cell cycle. A very robust design might be insusceptible to malfunctions but crucial errors, e.g. in the DNA duplication, remains without consequences in the further course. On the other hand, an undersized error tolerance leads to a very fragile and inflexible cell cycle.

An adjustable cell-cycle control system is an very important feature for the adaptation, e.g. on environmental changes, of the cell cycle. The control system of simple embryonic cells is based on a clock. The timer is unaffected by the events it regulates and will progress through the whole sequence of events even if one of those events has not been successfully completed. In contrast, the control system of most cell cycles is responsive to information received back from the processes it is controlling. Sensors, for example, detect the completion of DNA synthesis, and, if some malfunction prevents the successful completion, signals are sent to the control system to delay or stop progression to the

next phase. These delays provide time for the machinery to be repaired and also prevent the disaster that might result if the cycle progressed prematurely to the next stage.

In most cells there are several points in the cell cycle, called checkpoints, at which the cell cycle can be arrested if previous events have not been completed or extracellular conditions are unfavorable. For example, the progression through G_1 and G_2 is delayed by braking mechanisms if the DNA in the chromosomes is damaged by radiation or chemicals. The delays provide time for the damaged DNA to be repaired.

Furthermore, the cell cycle can be regulated by extracellular signals from other cells at the checkpoints. These signals can either promote or inhibit cell proliferation.

Although most checkpoints are not essential for normal cell-cycle progression under ideal conditions, populations with checkpoint defects often accumulate mutations due to occasional malfunctions in DNA replication, DNA repair, or spindle assembly. Some of these mutations can promote the development of cancer.

5.2.1 Cell-Cycle control system and cyclically activated protein kinase

At the heart of the cell-cycle control system is a family of protein kinases known as cyclin-dependent kinases (Cdk). The activity of these kinases rises and falls as the cell progress through the cycle. The oscillations lead directly to cyclical changes in the phosphorylation of intracellular proteins that initiate or regulate the major events of the cell cycle — DNA replication, mitosis, and cytokinesis. An increase in Cdk activity at the beginning of mitosis, for example, leads to increased phosphorylation of proteins that control chromosome condensation, nuclear envelope breakdown, and spindle assembly.

Cyclical changes in Cdk activity are controlled by a complex array of enzymes and other proteins. The most important of these Cdk regulators are proteins known as cyclins. Cdks are dependent on cyclins for their activity: unless they are tightly bound to a cyclin, they have no protein kinase activity. Cyclins were originally named as such because they undergo a cycle of synthesis and degradation in each cell cycle. Cdk levels, by contrast, are constant, at least in the simplest cell cycles. Cyclical changes in cyclin levels result in the cyclic assembly and activation of the cyclin-Cdk complexes. This activation in turn triggers cell-cycle events, as it is outlined in Figure 5.2.

There are four classes of cyclins, each defined by the stage of the cell cycle at which they bind Cdks and function. Three of these classes are required in all eucaryotic cells:

1. G_1/S-cyclins bind Cdks at the end of G_1 and commit the cell to DNA replication.
2. S-cyclins bind Cdks during S-Phase and are required for the initiation of DNA replication.
3. M-cyclins promote the events of mitosis.

The fourth class, the G_1-cyclins, helps promote passage through Start or the restriction point in late G_1. Whereas in yeast a single Cdk protein binds all classes of cyclins and drives all cell-cycle events by changing cyclin partners at different stages of the cycle, in vertebrates cells there are four Cdks. Two interact with G_1-cyclins, one with G_1/S- and S-cyclins, and one with M-cyclins.

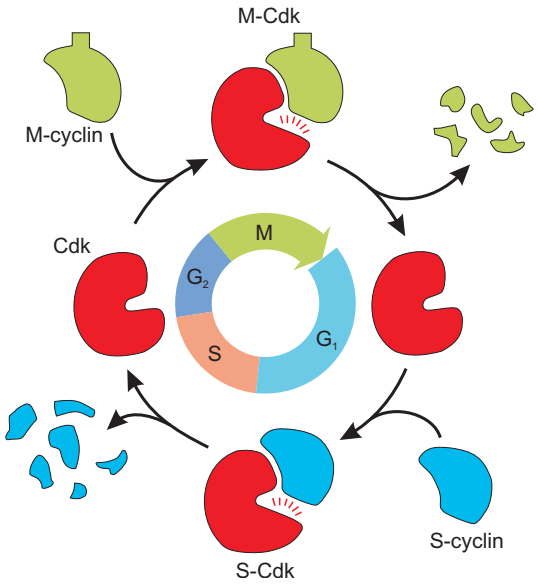


Figure 5.2: Simplified view of the core of the cell-cycle control system [AJL+02]. For demonstration purposes, only the cyclins that act in S phase (S-cyclin (blue)) and M phase (M-cyclin (green)) are shown. The interaction between a single Cdk and a cyclin results in a cyclin-Cdk complex referred to as S-Cdk and M-Cdk, respectively.

The rise and the fall of cyclin concentrations is the primary determinant of Cdk activity during the cell cycle. Several additional mechanisms, however, are important for fine-tuning Cdk activity at specific stages in the cell cycle.

The activity of a cyclin-Cdk complex can be inhibited by phosphorylation at a pair of amino acids in the roof of the active site. Phosphorylation of these sites by a protein kinase known as Wee1 inhibits Cdk activity, while dephosphorylation of these sites by a phosphatase known as Cdc25 increase the activity, as shown in Figure 5.3[a]. This regulatory mechanism is particular important in the control of M-Cdk activity at the onset of mitosis.

A further regulation arises from the binding of Cdk inhibitor proteins (CKI), see Figure 5.3[b]. There are a variety of CKI proteins, and they are primarily employed in the control of G_1 and S phase.

5.2.2 Cell cycle and cell growth

For proliferating cells to maintain a relatively constant size, the length of the cell cycle must match the time it takes the cell to double in size. If the cycle time is shorter than this, the cells will get smaller with each division. The cells will get bigger with each division, if it is longer. Because cell growth depends on nutrients and growth signals in the environment, the length of the cell cycle has to be adjustable to varying environmental conditions. There must be a correlation between cell growth and cell cycle. However, it is not clear how proliferating cells coordinate their growth with the rate of cell-cycle progression to maintain their size.

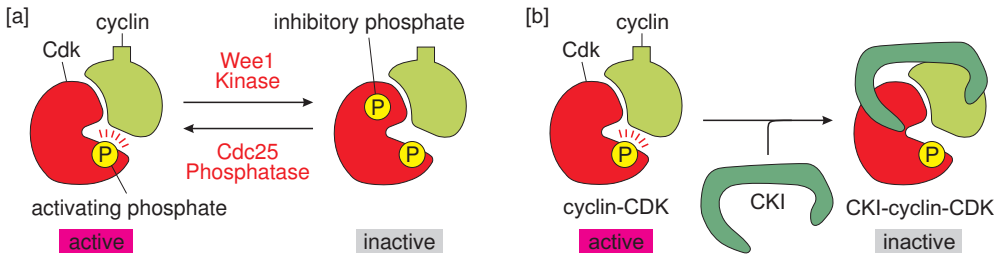


Figure 5.3: Regulation of Cdk activity by inhibitory phosphorylation and a CKI [AJL+02]
[a] Inhibitory phosphorylation: The active cyclin-CDK complex is deactivated when the kinase Wee1 phosphorylates two closely spaced sites above the active site. The phosphatase Cdc25 removes these phosphates and activates the complex.
[b] Inhibition by CKI: The CKI binds to both the cyclin and Cdk in the complex, distorting the active site of Cdk. It also inserts into the ATP-binding site, further inhibiting the enzyme activity.

There is evidence that cells coordinate their growth and cell-cycle progression by monitoring the total amount of a G_1 cyclin, which is synthesised in parallel with cell growth. But the mechanism, how the cell measures the total amount rather than its concentration, is still unclear.

Whereas cells of simple organisms grow and proliferate constitutively if nutrients are plentiful, animal cells generally grow and proliferate only when they are stimulated to do so by signals from other cells. The size at which an animal cell divides depends on, at least in part, on these extracellular signals, which can regulate cell growth and proliferation independently. Animals cells can also completely uncouple cell growth and division so as to grow without dividing or to divide without growing. The eggs of many animals, for example, grow to an extremely large size without dividing. After fertilisation, this relation is reversed, and many rounds of division occur without growth.

Thus, cell division and cell growth are usually coordinated, but they can be regulated independently. Cell growth does not depend on cell-cycle progression. Neurons and muscle cells, for example, grow large after they are withdrawn permanently from the cell cycle.

5.3 Model Systems to Study the Cell-Cycle Control System

The cell-cycle can be investigated experimentally on some 'simple' model systems. In studies of the genetics of the cell cycle a tiny, single-celled fungi, Yeast, is used. Two species are generally used in experiments. The fission yeast *Schizosaccharomyces pombe* is named after the African beer it used to produced. The budding yeast *Saccharomyces cerevisiae* is used by brewers, as well as by bakers. These two species share a number of features that are extremely useful for genetic studies.

They reproduce almost as rapidly as bacteria and have a genome size less than 1% that of a mammal. They are amenable to molecular genetic manipulation, whereby genes can be deleted, replaced, or altered. Most importantly, they have the unusual ability to proliferate in a haploid state, in which only a single copy of each gene is present in the

cell. In such cells one avoids the complication of having a second copy of the gene in the cell and it is easy to isolate and study mutations that inactivate a gene. Many of these mutations cause cells to arrest at a specific point in the cell cycle, suggesting that the normal gene product is required to get the cell past this point.

While yeasts are ideal for studying the genetics of the cell cycle, the biochemistry of the cycle is most easily analyzed in the giant fertilised eggs of many animals, which carry large concentrations of the proteins needed for cell division. The egg of the frog *Xenopus*, for example, is over 1mm in diameter and carries about 10^5 times more cytoplasm than an average cell in the human body. Fertilisation of the egg triggers an astonishingly rapid sequence of cell divisions, in which the single giant cell divides, without growing, to generate an embryo containing thousands of smaller cells. In this process, almost the only macromolecules synthesised are DNA — required to produced the thousands of new nuclei — and a small amount of proteins.

The cells in early embryos of *Xenopus*, as well as those of the clam *Spisula* and the fruit fly *Drosophila* are thus capable of exceedingly rapid division in the absence of either growth or many of the control mechanisms that operate in more complex cell cycles. These early embryonic cells cycles therefore reveal the working of the cell-cycle control system stripped down and simplified to the minimum needed to achieve the most fundamental requirements — the duplication of the genome and its segregation into two daughter cells. Another advantage is their large size. It is relatively easy to inject test substances into an egg to determine their effect on cell-cycle progression. It is also possible to prepare almost pure cytoplasm from *Xenopus* eggs and reconstitute many events of the cell cycle in a test tube.

5.4 Modeling the Cell Cycle of Yeast

A often used model organism for studying the cell cycle, its components and its regulation is the single celled yeast. Because of its known gene sequence we can identify the encoding genes of proteins. With this knowledge knock-out experiments were made to identify the proteins participating on the cell cycle. Furthermore we obtain information of their function during the cycle, e.q. if a specific protein triggers the transition from G_2 to M phase. If we know the components and their function, we are able to build a mathematical model. The numerical simulation of this model provides us with information of the temporal operation of the cell cycle. This we demonstrate now using a model of the yeast cell cycle developed by Novak et al. [NPCT01]. But before we discuss the mathematical representation of the cell cycle engine in yeast we want to shortly introduce the molecular basis of the model.

5.4.1 Molecular components of the yeast cell cycle

Lower eukaryotes, like yeast, use only one essential Cdk subunit (generally called Cdk1). In yeast Cdk1 is often called Cdc2, in recognition of the gene (cdc2) that encodes this protein in fission yeast [Nur90]. In fission yeast complexes between Cdc2 and the B-type cyclin Cdc13 play the major roles in cell cycle regulation [FN95, Nas96b]. Deletion of

the gene encoding this essential cyclin produces mutant cells that cannot enter mitosis [HFWN94].

DNA replication occurs once per cycle because Cdk activity not only triggers DNA replication but also inhibits replication of DNA [SN96]. To start DNA synthesis, Cdk activity causes properly licensed origins of replication to begin copying the chromosomes, and the same time it phosphorylates licensing factor molecules, making them more susceptible to degradation. Disappearance of licensing factors from the nucleus prevent further rounds of replication. Cdk activity increases to higher level in late G_2, thereby initiating M phase. As cells exit M phase, Cdk activity must be destroyed, to permit accumulation of licensing factors at replication origins in the next G_1 phase.

Destruction of Cdk activity as cells exit mitosis is the job of the anaphase promoting complex (APC). The APC is a large protein complex that attaches ubiquitin tags to target proteins, which are then rapidly degraded by proteasomes. The APC has two important functions at anaphase:

1. to initiate degradation of the cohesion proteins that hold two sister chromatids together, thereby initiating mitotic anaphase
2. to initiate degradation of B-type cyclins, thereby permitting cells to re-enter G_1 phase

To recognise the proper substrates for ubiquitination, the APC core requires specific "auxiliary" proteins. Slp1 targets the cohesion complex for disassembly, and both Slp1 and Ste9 present Cdc13 to the APC for ubiquitination [YON00, BSDdPM00]. Proper timing of these events is controlled by phosphorylation and dephosphorylation of Slp1 and Ste9.

5.4.1.1 MPF's enemies and friends

The activity of the Cdc2/Cdc13 complex (also called "M-phase promoting factor" MPF) is controlled by antagonistic interactions with its enemies. These enemies have negative effects on MPF, but MPF can down-regulate all of its enemies. Two of these enemies are active in G_1 phase, while a different group regulates the G_2/M transition.

The first G_1 enemy, Ste9 [YMO97, KMS98] targets Cdc13 to the APC core and promotes its degradation. On the other hand, phosphorylation of Ste9 by MPF inhibits its association with the APC core, rendering it inactive [YON00, BSDdPM00].

The other G_1 enemy of MPF is a stoichiometric inhibitor, called Rum1 [MN94], which can bind to Cdc2/Cdc13 complexes and inhibits their activity [CBN95, MCLM96]. On the other hand, a phosphorylation of Rum1 by MPF promotes its ubiquitination and rapid degradation [BMCM98]. Hence, there is antagonism between MPF and Rum1, as well as between MPF and Ste9.

Because of these antagonistic relationships, MPF and its G_1 enemies cannot coexist. Either the enemies win and the cell is in G_1 phase corresponding to a low MPF activity or MPF wins and the cell is in S/G_2/M phase of the cycle [NCNB+98]. The balance between MPF and its enemies is shifted by helper molecules in one direction or the other.

The Start transition ($G_1 \rightarrow S$) is supported by a "starter" kinase (SK), a group of Cdk/cyclin complexes. It helps the Cdc2/Cdc13 complex (MPF) to get the upper hand by phosphorylating Rum1 and Ste9. The helper molecule for the transition $M \rightarrow G_1$ is the Slp1(APC complex, which promotes the degradation of Cdc13 and activates Ste9. Slp1 can help the enemies because it is not inactivated by MPF phosphorylation, as is Ste9.

The duration of G_2 phase is regulated by different mechanism, namely enzymatic inhibition of MPF activity. The kinases Wee1 and Mik1 [RN87, LWB+91] can inactivate Cdc2. In return, MPF can also phosphorylate and inactivate them. So we have another case of mutual antagonism and alternative steady states: an S/G_2 state and an M state.

The G_2/M transition is accelerated by a direct positive feedback loop. The inhibitory phosphate group of Cdc2 is removed by a specific phosphatase, called Cdc25 [MR92]. This phosphatase is phosphorylated by MPF, but the phosphorylated form is mire active. Here, MPF helps its friend, Cdc25.

5.4.1.2 Feedback loops

For cells to proliferate, to make a repetitive sequence of properly controlled Start, G_2/M and Finish transitions, the helper molecules must be removed after they have done their jobs, because they are inhibitory for the next cell cycle. For instance, the starter kinase (SK) would inhibits the Finish transition, when the MPF enemies must come back. Therefore, MPF inhibits the synthesis of SK by phosphorylating its transcription factor (TF).

In analogy, Slp1 must disappear after Finish transition; otherwise, it would inhibit the next Start transition. The synthesis and activation of Slp1 depends on MPF creating a negative feedback loop [MR92]. It is essential that Slp1/APC complex is not directly activated by MPF, but rather through an intermediary enzyme (IE), which provides a time delay in the loop. This delay gives the chromosomes enough time to align before Slp1/APC breaks down their cohesion.

5.4.2 Surveillance mechanisms, checkpoints

These helper molecules are regulated by surveillance mechanisms (also called checkpoints) [Nas96a]. Start is controlled by cell mass, Finish by the state of the cell's chromosomes, and the G_2/M transition is affected by both. The chromosome cycle, regulated by the cell cycle engine, must run in concert with overall cytoplasmic growth. Else the cells becoming hopeless small or enormously large. Without such a coordination mechanism, cells cannot be kept alive over the long term.

How cytoplasmic mass exerts its control over the cell cycle engine is not clear at present.

5.4.3 Mathematical model

After we identified the key proteins in the cell cycle of yeast and its functional relationships, we want to represent the cycle within a mathematical representation. From

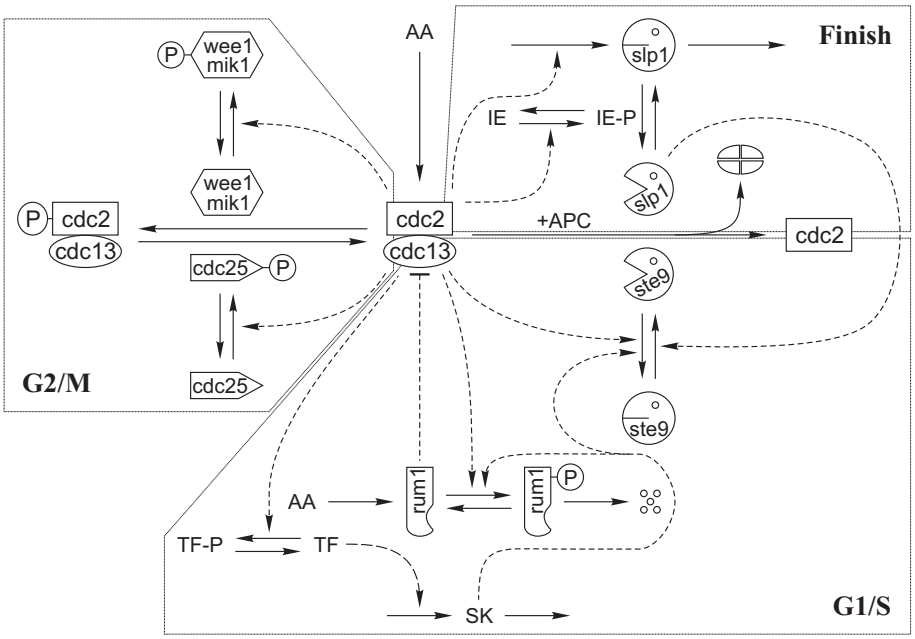


Figure 5.4: The wiring diagram of the fission yeast cell cycle engine according to Novak et al. [NPCT01]. The core of the engine is the Cdc2/Cdc13 (MPF) complex, which is regulated by proteolysis of the Cdc13 component, phosphorylation of Cdc2 subunit, and stoichiometric inhibition of the complex. These processes are arranged according to the cell cycle transitions in which they are involved.

the graphical representation as wire diagram, Figure 5.4, we can establish a model of coupled ordinary differential equations (5.1)-(5.12)[NPCT01].

$$\frac{d\,[\mathrm{Cdc13_T}]}{dt} = k_1\,M - \left(k_2^{'} + k_2^{''}\,[\mathrm{Ste9}] + k_2^{'''}\,[\mathrm{Slp1}]\right)[\mathrm{Cdc13_T}]\,, \tag{5.1}$$

$$\frac{d\,[\mathrm{preMPF}]}{dt} = k_{\mathrm{wee}}\,([\mathrm{Cdc13_T}] - [\mathrm{preMPF}]) - k_{25}\,[\mathrm{preMPF}] - \left(k_2^{'} + k_2^{''}\,[\mathrm{Ste9}] + k_2^{'''}\,[\mathrm{Slp1}]\right)[\mathrm{preMPF}]\,, \tag{5.2}$$

$$\frac{d\,[\mathrm{Ste9}]}{dt} = \left(k_3^{'} + k_3^{''}\,[\mathrm{Slp1}]\right)\frac{1-[\mathrm{Ste9}]}{J_3+1-[\mathrm{Ste9}]} - \left(k_4^{'}\,[\mathrm{SK}] + k_4\,[\mathrm{MPF}]\right)\frac{[\mathrm{Ste9}]}{J_4+[\mathrm{Ste9}]}, \tag{5.3}$$

$$\frac{d\,[\mathrm{Slp1}]}{dt} = k_5^{'} + k_5^{''}\frac{[\mathrm{MPF}]^4}{J_5^4+[\mathrm{MPF}]^4} - k_6\,[\mathrm{Slp1_T}]\,, \tag{5.4}$$

$$\frac{d\,[\mathrm{Slp1^*}]}{dt} = k_7\,[\mathrm{IEP}]\,\frac{[\mathrm{Slp1_T}]-[\mathrm{Slp1}]}{J_7+[\mathrm{Slp1_T}]-[\mathrm{Slp1}]} - k_8\frac{[\mathrm{Slp1}]}{J_8+[\mathrm{Slp1}]} - k_6\,[\mathrm{Slp1}]\,, \tag{5.5}$$

$$\frac{d\,[\mathrm{IEP}]}{dt} = k_9\,[\mathrm{MPF}]\,\frac{1-[\mathrm{IEP}]}{J_9+1-[\mathrm{IEP}]} - k_{10}\frac{[\mathrm{IEP}]}{J_{10}+[\mathrm{IEP}]}, \tag{5.6}$$

$$\frac{d\,[\mathrm{Rum1_T}]}{dt} = k_{11} - \left(k_{12} + k_{12}^{'}\,[\mathrm{SK}] + k_{12}^{''}\,[\mathrm{MPF}]\right)[\mathrm{Rum1_T}]\,, \tag{5.7}$$

$$\frac{d\,[\mathrm{SK}]}{dt} = k_{13}\,[\mathrm{TF}] - k_{14}\,[\mathrm{SK}]\,, \tag{5.8}$$

$$\frac{d\,[\mathrm{M}]}{dt} = \mu\,M, \tag{5.9}$$

$$[\mathrm{Trimer}] = \frac{2\,[\mathrm{Cdc13_T}]\,[\mathrm{Rum1_T}]}{\Sigma + \sqrt{\Sigma^2 - 4\,[\mathrm{Cdc13_T}]\,[\mathrm{Rum1_T}]}}, \tag{5.10}$$

$$[\mathrm{MPF}] = \frac{([\mathrm{Cdc13_T} - [\mathrm{preMPF}]])\,([\mathrm{Cdc13_T}] - [\mathrm{Trimer}])}{[\mathrm{Cdc13_T}]}, \tag{5.11}$$

$$[\mathrm{TF}] = G\left(k_{15}M, k_{16}^{'} + k_{16}^{''}\,[\mathrm{MPF}]\,, J_{15}, J_{16}\right) \tag{5.12}$$

where

$$k_{\mathrm{wee}} = k_{\mathrm{wee}}^{'} + \left(k_{\mathrm{wee}}^{''} - k_{\mathrm{wee}}^{'}\right)G\left(V_{\mathrm{awee}}, V_{\mathrm{iwee}}\,[\mathrm{MPF}]\,, J_{\mathrm{awee}}, J_{\mathrm{iwee}}\right), \tag{5.13}$$

$$k_{25} = k_{25}^{'} + \left(k_{25}^{''} - k_{25}^{'}\right)G\left(V_{\mathrm{a25}}\,[\mathrm{MPF}]\,, V_{\mathrm{i25}}, J_{\mathrm{a25}}, J_{\mathrm{i25}}\right), \tag{5.14}$$

$$\Sigma = [\mathrm{Cdc13_T}] + [\mathrm{Rum1_T}] + K_{\mathrm{diss}}. \tag{5.15}$$

We use the Michaelis-Menten scheme for describing the activation an deactivation reactions of the participating proteins, whereby we assume the formed intermediate complexes in a quasi-steady state. As result we obtain Michaelis-Menten like expressions within our system of coupled differential equations, see for instance Eq. (5.3) for the

protein ste9. A further simplification arises from the assumption of a quasi-stationary state for the following proteins

1. the transcription factor (TF) for the synthesis of the starter kinase (SK), Eq. (5.12),

2. the trimeric complexes of Cdc13/Cdc2 and Rum1 (Trimer), Eq. (5.10),

3. the enzymatic inhibitors Wee1, Eq. (5.13), and cdc25, Eq. (5.14).

Then their concentrations can be expressed in terms of the Goldbeter-Koshland function [GKJ81, TCN03]

$$G(a,b,c,d) = \frac{2ad}{b-a+bc+ad+\sqrt{(b-a+bc+ad)^2-4ad(b-a)}} \tag{5.16}$$

describing the steady state solution of an activation/deactivation cycle (see section 4.9.3).

Furthermore we used conservation laws for the proteins Wee1, cdc25, IE, and ste9 to reduce the model structure. We assume further, that Rum1 binds to unphosphorylated (MPF) and the phosphorylated (preMPF) form of the Cdc13/Cdc2 complex.

For the protein slp1 we have to distinguish between its deactivated form slp1 and its activated form slp1*, Eqs. (5.4) and (5.5).

As shown in the wiring diagram, Figure 5.4, the proteins rum1 and slp1 are also produced outside from the cell cycle engine. We assume a constant external production rate represented by zero-order rate constants k_5' and k_{11}. The cell mass M takes in a special position in the presented model. Because of a lack of knowledge of the feedback regulation and to simplify the model we treat the cell growth as an independent dynamic variable, as shown in the corresponding differential equation (5.9). At the end of mitosis the cell mass is divided by two triggered by a decreasing MPF level through 0.1.In the model we assume an instantaneous cell division into two daughter cells of the same mass and volume, although daughter cells do not physically separate from another until 15-20 min after mitosis [NPCT01]. Due to this boundary condition we simulate a regulatory relationship between the cell cycle and the cell growth. Notice, that during the division the concentrations of the proteins remain unchanged[1]. But the new cell mass influence the production rate of the protein Cdc13, see Eq. (5.1), which affects in succession all other components of the cell cycle engine. A possible resting phase G_0 or differences of growth behavior in G_1, G_2, M, and S phase are not included in this model [NPCT01]. Also, it does not describe the influence of changing external conditions to the cell cycle. We assume constant and ideal external conditions resulting in a exponential growth. The exponential function is characterised by the constant growth rate μ.

Cdc13 synthesis and degradation:
$k_1 = 0.03$, $k_2^{'} = 0.03$, $k_2^{''} = 1$, $k_2^{'''} = 0.1$
Ste9 activation and deactivation:
$k_3^{'} = 1$, $k_3^{''} = 10$, $J_3 = 0.01$, $k_4^{'} = 2$, $k_4^{''} = 35$, $J_4 = 0.01$
Slp1 synthesis, degradation, activation, and deactivation:
$k_5^{'} = 0.005$, $k_5^{''} = 0.3$, $J_5 = 0.3$, $k_6 = 0.1$
$k_7 = 1$, $J_7 = 0.001$, $k_8 = 0.25$, $J_8 = 0.001$
IE activation and deactivation:
$k_9 = 0.1$, $J_9 = 0.01$, $k_{10} = 0.04$, $J_{10} = 0.01$
Rum1 synthesis, degradation, and inhibition:
$k_{11} = 0.1$, $k_{12} = 0.01$, $k_{12}^{'} = 1, k_{12}^{''} = 3$, $K_{\text{diss}} = 0.001$
SK synthesis and degradation:
$k_{13} = k_{14} = 0.1$
TF activation and deactivation:
$k_{15} = 1.5$, $J_{15} = 0.01$, $k_{16}^{'} = 1$, $k_{16}^{''} = 2$, $J_{16} = 0.01$
Wee1 activation and deactivation:
$V_{\text{awee}} = 0.25$, $J_{\text{awee}} = 0.01$, $V_{\text{iwee}} = 1$, $J_{\text{iwee}} = 0.01$
Cdc25 activation and deactivation:
$V_{a25} = 1$, $J_{a25} = 0.01$, $V_{i25} = 0.25$, J_{i25}=0.01
Rate of phosphorylation and dephosphorylation
$k_{\text{wee}}^{'} = 0.15$, $k_{\text{wee}}^{''} = 1.3$, $k_{25}^{'} = 0.05$, $k_{25}^{''} = 5$
Growth rate
$\mu = 0.005$

Table 5.1: Parameter values for the cell cycle model of Novak et al. [NPCT01], Eqs. (5.1)-(5.14) . All constants have units min^{-1}, except the dimensionless Michaelis constants J_i and the dissociation constant of the trimer K_{diss}.

5.5 Numerical Simulations

After we introduced the mathematical model of the cell cycle of yeast and its biological background, we now simulate the cell cycle. We compare the cell cycle of wild-type cells with mutant cells and investigate the differences in the dynamic change of key proteins [NPCT01]. Due to the mathematical complexity of the cell cycle model we can only numerically solve the system of coupled differential equations (5.1)-(5.14). The used parameter values of wild-type fission yeast are given in Table 5.1.

5.5.1 Wild-type cells

The temporal changes of the concentrations of components of the cell cycle of yeast are shown in Figure 5.5. We separate this graphical representation into three parts:

[a] temporal evolution of the cell mass

[b] temporal evolution MPF, preMPF and $Cdc13_T$

[c] temporal evolution of the enemies of MPF (Ste9, Slp1, Rum1) and of the starter kinase (SK)

Additionally we draw the phases of the cell cycle on top of the figure and extend the separations with dashed lines into all three plots. The events determining the different stages are discussed in the further course.

The evolution of the cell mass is shown in Figure 5.5[a]. As mentioned before, it follows an exponential monotone increasing law until the Mitosis is finished. During the time between two divisions a common yeast cell double their mass. According to Figure 5.5[a] it takes about 140 minutes to complete the cell cycle of wild type yeast cells. If we further assume an ideal environment there are no fluctuations in the maximal cell mass and the cycle time. If the Mitosis finishes the cell mass is halved and the arisen daughter cells will growth with the same rate.

The duration of Mitosis is determined by the MPF level plotted in part [b] of Figure 5.5. In order to enter the Mitosis the MPF level has to change from low to high. A requirement for this is a previous production of Cdc13 which can combine with cdc2 to the inactive form of the mitosis promoting factor, preMPF. As shown in the figure all the three components of yeast cell cycle among three levels. After a short time delay, Slp1/APC is activated by the high MPF activity in Mitosis, shown in part [c] of the figure, initiating the rapid degradation of Cdc13. As a consequence, MPF activity drops. Simultaneously, the increase of Slp1/APC activates Ste9. Because MPF inhibits the Ste9 activation, the low level of MPF has a further positive effect on the production of active Ste9. The period of high Ste9 level determines the G_1 phase. The low MPF concentration leads to an inactivation of Slp1 and relieves the inhibition on the transcription factor

[1] Concentrations are intensive quantities, whereas the mass is an extensive one. A doubling of the considered system doubles extensive quantities, but the intensive keep constant. In particular, the ratio of two extensive quantities, for instance number of molecules and volume, results in an intensive quantity, for our example in a concentration

TF responsible for the synthesis of cyclin subunit of the SK. The SK level increases and inhibits the MPF enemies Ste9 and Rum1. Actually, G_1 is so short that Rum1 does not have time to come up [CBN95]. As soon as Ste9 gets inactivated, the Cdc13 level rises and the cell passes the G_1/S transition. However, SK does not inactivate the third enemy, Wee1, which phosphorylates Cdc2/Cdc13. The phosphorylated from has reduced protein-kinase activity, which seems to be enough to initiate S phase but not mitosis. When the cell reaches a critical size, the positive feedbacks for G_2/M transition turn on. Abrupt activation of MPF by Cdc25 drives the cell into mitosis. During this process virtually all Cdc13 is bound in MPF, the precursor form preMPF is completely converted into MPF. Hence, we come back to our starting point and a new cell cycle begins.

The SG_2 transition was arbitrarily chosen. A theoretical model of this transition is not included in the presented cell cycle model.

5.6 Mutations

Mutations play an important role in the development of organisms. Due to the increasing knowledge of proteins and their encoding genes and the arising new technologies to alter the genome of an organism in a well defined manner, it is possible to investigate the role of individual proteins and their influence on the cell cycle in experiments.

5.6.1 *wee*1$^-$ - mutants

The inhibition of MPF by Wee1 is an crucial step in the cell cycle of wild type cells. In order to inactivate Wee1, cells must grow to a critical size, which necessitates an extended $S+G_2$ phase (about 100 minutes). Hence, a change in the Wee1 activity strongly affects the duration of the cell cycle and the cell mass. The cell cycle of these *wee*1$^-$ cells is investigated in Figure 5.6. Again, we use the same representation as described in the previous section. The activity of Wee1 is expressed in the rate constants for its inactivation k'_{wee} and for its activation k''_{wee}. A smaller activation constant of $k''_{\text{wee}} = 0.3$ (for wild-type cells see Table 5.1) shifts the concentration of activated Wee1 to a decreased concentration [NPCT01]. As consequence a smaller critical mass is needed to pass the S transition. As in Figure 5.6[a] is shown, the cell mass is smaller than the mass of a wild-type cell. This is the defining characteristic of 'wee mutants [Nur75].

The G_2/M transition is not size controlled. Consequently, the $S+G_2$ phase is much shorter ($\approx$ 45 min) than in wild type. To adjust their cycle time to the mass doubling time (140 min), *wee*1$^-$ cells have an extended G_1 phase, stabilised by up-regulation of Rum1 and Ste9 (Figure 5.6[c]).

Because of the extend M phase, $\approx$ 25 min, which is about the twice of the M phase of wild-type cells, and the stronger increase of MPF during the $S+G_2$ phase, the starter kinase shows a pronounced oscillatory behavior. The suppression of TF activation during the M phase leads to an almost complete degradation of SK. If the MPF concentration decreases at the end of mitosis the SK level rises until the comparable small cell size prefers the degradation again. First with increasing cell mass the SK production gets

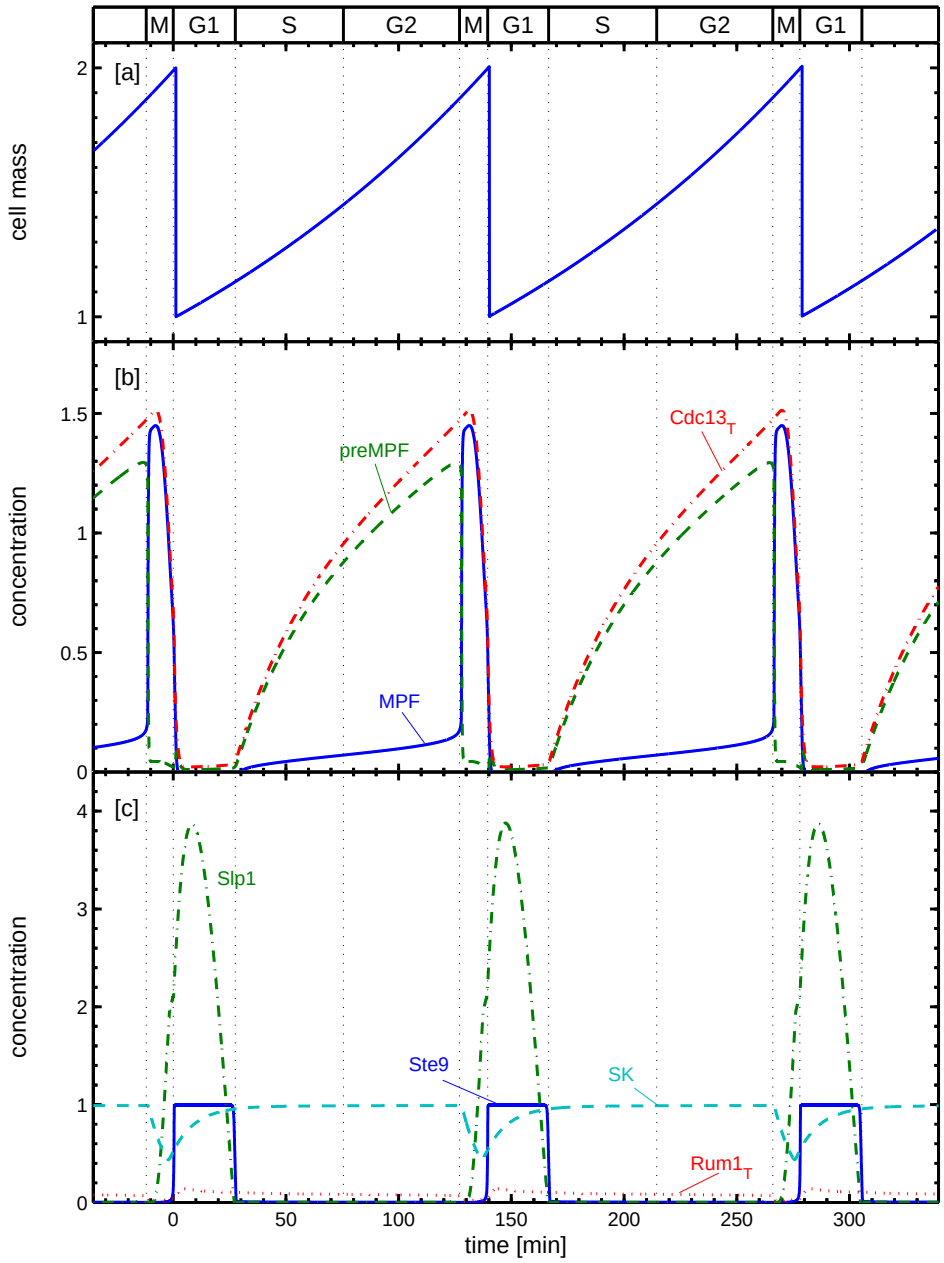


Figure 5.5: Numerical solution of the cell cycle model of yeast of Novak et al. [NPCT01], Eqs. (5.1)-(5.14), for wild-type cells. The corresponding parameters values are given in Table 5.1. Two complete cell cycles are shown. The figure is divided into three parts:
[a] The temporal evolution of the cell mass.
[b] The temporal evolution of the Mitosis-promoting factor (MPF), its inactive form (preMPF) and the cyclin Cdc13, one of its main proteins.
[c] The temporal evolution of enemies and friends of MPF.

the upper hand until the negative feedback of MPF gets the dominant part in Eq. (5.12). The $\mathrm{Rum1_T}$, on the other hand, benefits from the low level of starter kinase during the $\mathrm{G_1}$ phase. Its antagonist SK is down-regulated and its inhibitor MPF can be neglected within this period. Hence, it can reach a pronounced maximum in $\mathrm{G_1}$. If the MPF concentration and the SK concentration increase, $\mathrm{Rum1_T}$ is degraded to a similar low level as in wild-type cells, see Figure 5.5.

5.6.2 *wee*1$^-$ *cdc*25Δ mutants

Wee1 and Cdc25 are the major tyrosine-modifying enzymes in fission yeast. If Wee1 (the inhibitor) is missing, then cells should not need Cdc25 (the activator). Indeed, *wee*1$^-$ *cdc*25Δ double-mutant cells are viable, but they exhibit abnormal progression through the cell cycle.

For the simulation of these yeast mutants we use the rate constants $k'_{\mathrm{wee}} = 0.15$ and $k''_{\mathrm{wee}} = 0.3$ (for *wee*1$^-$ mutation) and $k'_{25} = k''_{25} = 0.02$ (for *cdc*25Δ mutation) [NPCT01]. In comparison to wild-type cells the mutated rate constants are reduced (see Table 5.1).

The numerical results are shown in Figure 5.7. The cells alternate between short cycles (about 110min) and long cycles (about 170min). As observed experimentally [SNM99], the mutant cells always divide at size larger than wild type. Cells born at larger size are committed to the shorter cell cycle, and smaller newborns have longer cycles. Due to the larger cell size the mutants have a much shorter $\mathrm{G_1}$ phase than wild-type cells.

The reason for the quantised cell cycle is the weak positive feedback in mitotic control due to lacking of activated Wee1 and Cdc25 [NPCT01]. The mitosis promoting factor is not activated abruptly, Figure 5.7[c], when cells are supposed to enter the M phase. As in Figures 5.5 and 5.6 we use the abrupt change in MPF behavior as the transition $\mathrm{G_2}/M$. Then the *wee*1$^-$ *cdc*25Δ double-mutant cells have an extended M phase compared to the previously considered fission yeast cells. Furthermore, only a small part of preMPF is activated during this phase. During the M phase of wild-type cells and *wee*1$^-$ mutants the conversion to MPF was almost complete.

Because MPF rise is sluggish, it may not turn on the fully the exit-from-mitosis pathway. The negative feedback loop generates a rebound in MPF activity leading to a degradation and inactivation of the involved proteins. After a delay a 'second try' is started when the cell size is larger which leads to a successful mitosis.

Because of the comparable small MPF concentration in the cell cycle of *wee*1$^-$ *cdc*25Δ double-mutants the inhibitory effect on the transcription factor TF is weak and the concentration of the starter kinase SK is almost constant and on a high level over the whole cycle. Consequently, the MPF enemy Rum1 is suppressed.

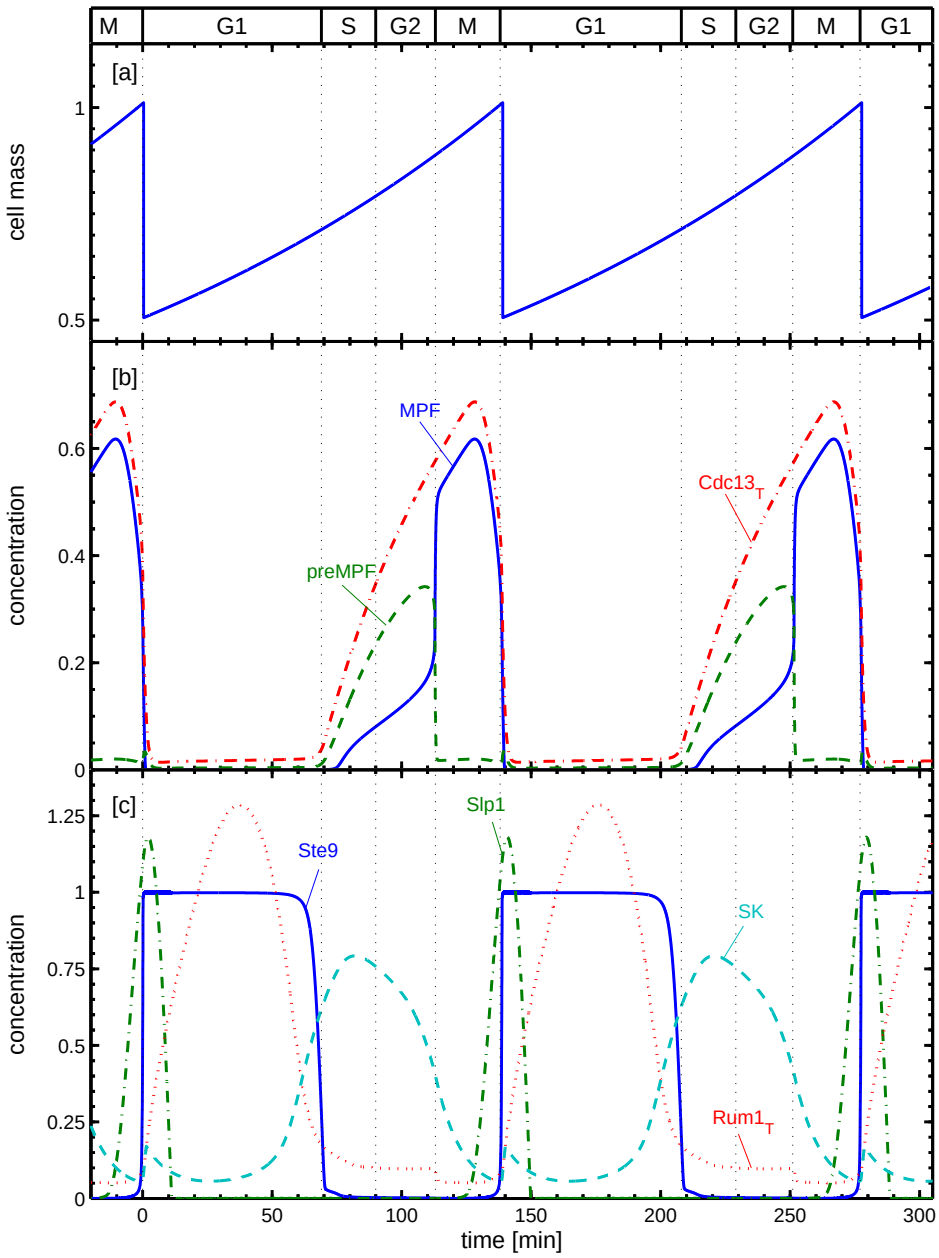


Figure 5.6: Numerical solution of the cell cycle model of yeast of Novak et al. [NPCT01], Eqs. (5.1)-(5.14), for wee1$^-$ mutants. The corresponding parameters values are given in Table 5.1, except $k''_{\rm wee} = 0.3$. Two complete cell cycles are shown. The figure is divided into three parts:

[a] The temporal evolution of the cell mass.

[b] The temporal evolution of the Mitosis-promoting factor (MPF), its inactive form (preMPF) and the cyclin Cdc13, one of its main proteins.

[c] The temporal evolution of enemies and friends of MPF.

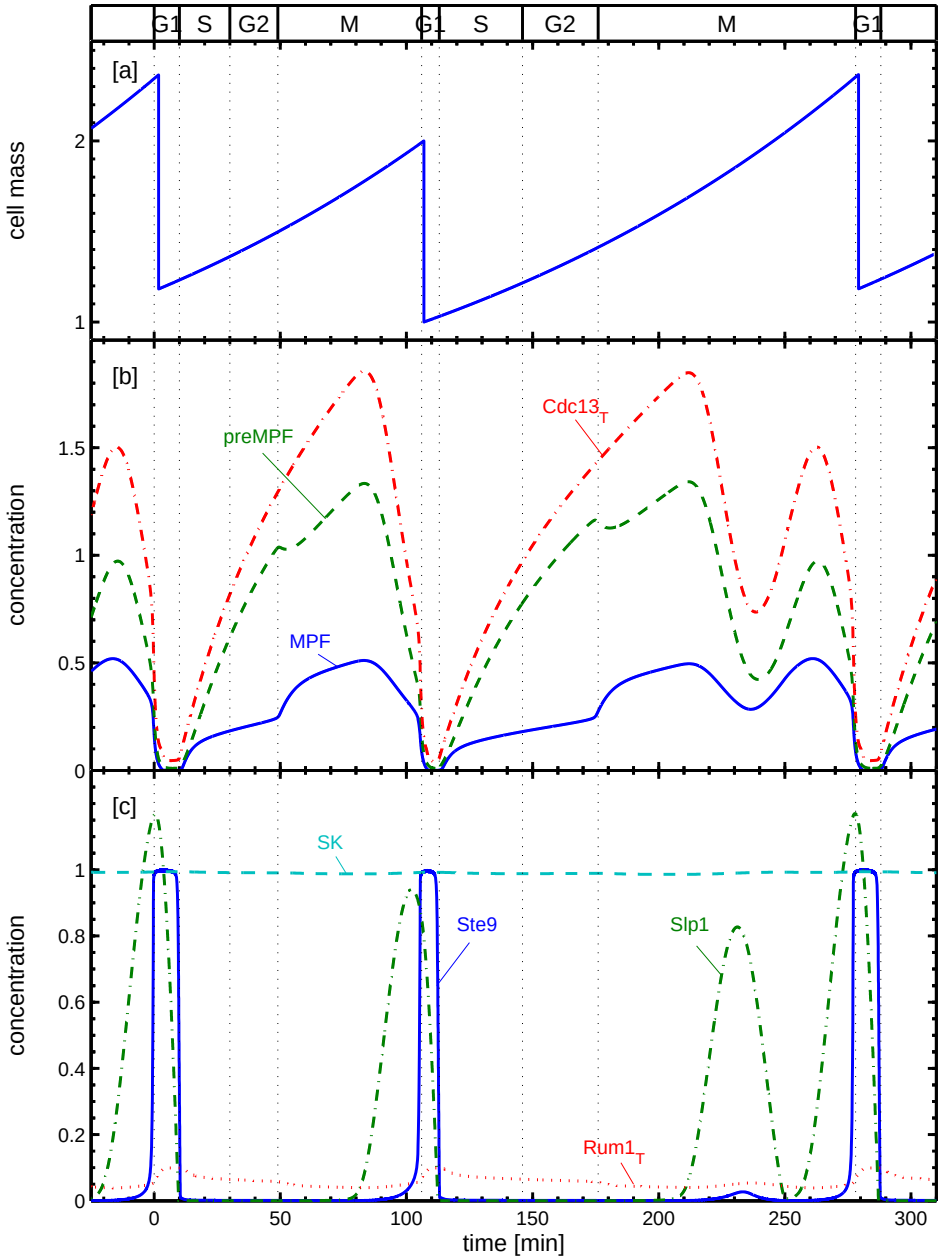


Figure 5.7: Numerical solution of the cell cycle model of yeast of Novak et al. [NPCT01], Eqs. (5.1)-(5.14), for wee1$^-$ cdc25Δ mutants. The corresponding parameters values are given in Table 5.1, except $k''_{\text{wee}} = 0.3$ and $k'_{25} = k''_{25} = 0.02$. Two complete cell cycles are shown. The figure is divided into three parts:
[a] The temporal evolution of the cell mass.
[b] The temporal evolution of the Mitosis-promoting factor (MPF), its inactive form (preMPF) and the cyclin Cdc13, one of its main proteins.
[c] The temporal evolution of enemies and friends of MPF.

h

6 Metabolic Control Analysis

In the previous chapters we dealt with autonomous nonlinear systems and investigated, how the structure of a set of ordinary differential equations determines the properties and the structure of the corresponding solutions. There we focused on

- Structural properties of biochemical networks;
- Dependencies of the dynamic and transient behavior on initial states;
- Dependencies of the dynamic and transient behavior on rate coefficients.

Several methods for the analysis of nonlinear systems were introduced in Section 4.5 and, from a more system theoretic point of view, in subsequent sections. Simple examples were given that demonstrate how changes in parameters, initial conditions, and structure can change systems properties.

In this chapter, we introduce a general approach to investigate and quantify such changes in a systematic way. We are interested in the control exerted by the various reactions over the fluxes or rates and over the concentrations, especially at steady state. The main questions, which to be addressed during the course are:

- How much does the steady state change if one varies the concentration of participating species?
- How much does the steady state change if one varies the rate coefficients?
- How is the reaction network controlled and regulated?

This approach was developed for investigations of metabolic networks thus, it is called Metabolic Control Analysis (MCA). Nevertheless, it should be mentioned here that the mathematical framework is not restricted to metabolism. It can be applied to any problem that considers the transformations of elements or, more general, which considers the fluxes of some elements, e.g. economics, traffic and crowd control. The mathematical formalism, introduced in the next sections, provides us with a method to describe the control of metabolic systems. For this reason, we will use the usual terminology of MCA. The first species of a reaction network are referred to as substrates, intermediary components as metabolites and the final components of the metabolic pathway are products. Thereby, we follow the convention that the direction of the reaction goes from the substrates to the products [HS96]. This implies that there is a continuous conversion of substrate into products. The amount of converted material per unit of time is described as flux, for example as number of particles per second. In analogy to the definition of fluxes in Irreversible Thermodynamics, the fluxes in the Metabolic Control Analysis

have a unique direction. However, a generalisation for arbitrary direction is possible [Fel97]. Nevertheless, before we introduce the mathematical framework of MCA we will introduce some fundamental quantities in the next section.

Our notation follows the convention proposed by several leading authors of the MCA community [WGW84, BCBG+85]. Even if this notation differs from the one used in the original papers of Kacser and Burns [KB73] and Heinrich and Rapoport [HR73, HR74, HR75], it becomes more and more accepted in recent publications.

6.1 Fundamentals

6.1.1 Control and regulation

control
regulation

Metabolic control analysis is a means to study the control of metabolic processes. It ascribes a specific meaning to the term control. The terms control and regulation are often used interchangeable, but here they have a distinct meaning. *Control* is the ability to make changes in the behavior of a system, whereas *regulation* is the maintenance (or the lack of it) of constant internal conditions when a system is exposed to external disturbances.

In this course, we consider control as the ability to influence something. For example, if an enzyme controls the flux of a pathway, a change in the concentration of that enzyme changes the flux of the pathway.

Regulation describes whether any control (by a controller), is actually affected or not. In metabolism, regulatory mechanisms often exist to maintain a degree of homoeostasis. As well as maintaining constant conditions, regulatory mechanisms can also improve the performance of control.

When describing control we must be careful to define what is being 'controlled' and what is doing the 'controlling'. A controller is a component that controls some other component(s) of the system. In contrast to this, a component is controlled, if it is determined by another component. As we see later, these definitions are closely related to parameters and variables. However, it is important to note that the definition of what is 'controlled' (a variable) and what is the controller (a parameter) is arbitrary and depends on the system under investigation.

6.1.2 Parameters and variables

Parameters are quantities that can be changed independently, and they typically (but not always) remain constant during the evolution of the system toward its steady state. Examples include kinetic coefficients, enzyme concentrations, and external inhibitors, but also physical quantities like temperature and pH-value. Parameters can be classified as internal and external [HCB91]. Internal parameters have values that, although not invariant in the absolute sense, are fixed characteristics of a particular system, e.g. the thermodynamic and kinetic coefficients of the enzyme reactions. External parameters are those quantities through which the environment can affect the system. Most realistically external parameters are concentrations, e.g. of enzymes, terminal metabolites,

and external effectors, such as inhibitors, activators, hormones or growth factors; their levels are determined by the environment of the system and do not depend in any way on what happens in the system. Consequently, changes in the environment can only be communicated to the system through changes of external parameters.

Variables are quantities determined by the system, and they are time-dependent before reaching their steady state. The most common variables are metabolite concentrations. Other important variables are the rates of reactions. These are functions of concentration variables and kinetic parameters. In most cases, the reaction rates will be described by enzyme kinetics. However, they may also represent other types of reactions or transport processes. A change in systems parameters will lead to a change of systems variables.

6.2 Control Coefficients

Control coefficients are fundamental to Metabolic Control Analysis. In general, a *control coefficient* is a relative measure of how much a perturbation of a systems parameter affects a systems variable. In other words, a control coefficient describes how sensitive a system variable is towards changes of a system parameter. Hence, an approach that determines control coefficients is called a sensitivity analysis.

control coefficient

The above definition of control coefficients contains an implicit assumption – the measurement of control coefficients requires a reference point. In metabolic systems, a natural choice for the reference point is the steady state. Hence, unless noted otherwise, we measure the relative changes in the considered system against the steady state determined by the system parameters. Usually, this means for metabolism that we investigate the modification of the steady state through parameter variation.

However, before we can investigate metabolic pathways in more detail, we have to identify a set of variables and parameters that determines the considered metabolic system uniquely. As an example, we use the linear pathway represented in Figure 6.1. Its steady state can be characterised by different quantities. First, we can consider the concentrations of substrates, intermediates, and products as in the investigation of transient phenomena in the previous sections. Second, the network can be characterised by the overall rate at steady state. From the definition of an open system and the convention that the net conversion of the investigated pathway is directed from the substrates to the products [HS96][1], it follows that the flux is always positive. The rate describing the amount of substrates that are converted into products per unit of time is usually called flux of the pathway in MCA.

$$S \underset{E_1}{\overset{v_1}{\rightleftharpoons}} X_1 \underset{E_2}{\overset{v_2}{\rightleftharpoons}} X_2 \underset{E_3}{\overset{v_3}{\rightleftharpoons}} X_3 \underset{E_4}{\overset{v_4}{\rightleftharpoons}} P$$

Figure 6.1: Linear four step pathway. The substrate S is converted into the product P via four subsequent steps with the intermediates X_1, X_2, and X_3. Each reaction is facilitated by a specific enzyme E_i and is characterised by the reaction rate v_i.

[1] Nevertheless, a generalisation to arbitrary directions is discussed in [Fel97]

At steady state, each enzymatic reaction of the metabolic pathway is operating at the same rate. No reaction is running slower than any other. Hence, we cannot deduce a controlling step from the fluxes. However, this does not exclude that only few enzymes or even one single enzyme may control the steady state of the considered pathway(Principle of rate-determining step introduced in Section **??**). The relation of MCA and this principle shall be discussed later in Section 6.3.

Flux and concentrations are determined by kinetic parameters, such as rate coefficients, Michaelis constants or Hill coefficients, depending on the chosen level of mathematical description. In our example, at steady state, all fluxes v_i are equal. Nevertheless, we identify a flux for each intermediary step, because these fluxes are determined by the parameters of the corresponding step and can therefore be of different importance to the control of the considered pathway. Furthermore, the concentration of enzyme facilitating a step of the metabolic pathway is treated as a parameter. Finally, we assume that the substrate and product concentrations remain constant and can be treated as system parameters.

The fluxes v_i of the pathway and the concentrations of the intermediates are variables that are determined by the system parameters. The changes in the system variables are described by various control coefficients, which we introduce now.

6.2.1 Flux control coefficient

First, we want to investigate, how a change of parameters changes the overall flux J of a pathway. Originally, Kacser and Burns [KB73] investigated how the concentration $[E_i]$ of the enzymes acting as catalysts of the individual reaction i affects the metabolic flux. Correspondingly, they defined a *flux control coefficient* as follows

flux control coefficient

$$C = \frac{\Delta J}{\Delta [E_i]}\,, \tag{6.1}$$

which relates the fractional change in the steady-state flux ΔJ to the fractional change in the total enzyme concentration $\Delta[E_i]$. In the present representation, the control coefficient depends on the units and the actual experimental conditions. Consequently, this absolute control coefficient will be different for other experimental setups and a comparison of different experiments and situations is, thus, not possible. This problem is resolved by normalisation with the former steady-state flux J and the corresponding enzyme concentration $[E_i]$

$$C_{E_i} = \frac{[E_i]}{J}\frac{\Delta J}{\Delta [E_i]}\,. \tag{6.2}$$

We consider only infinitesimal changes in the enzyme concentration, and obtain the flux control coefficient [KB73]

$$C^J_{E_i} = \frac{[E_i]}{J}\frac{\partial J}{\partial [E_i]}\,, \tag{6.3}$$

where the subscript E_i denotes the considered enzyme of the metabolic pathway. The superscript J indicates the control coefficient as a flux control coefficient. For further

simplification of (6.3), we use the mathematical properties of the derivative of the logarithmic function

$$\frac{\partial \ln f(x)}{\partial x} = \frac{1}{x}\frac{\partial f(x)}{\partial x}$$

which leads to the logarithmic flux control coefficient

$$C^J_{E_i} = \frac{\partial \ln J}{\partial \ln [E_i]} \,. \tag{6.4}$$

As represented in Figure 6.2, the control coefficient is the tangent of the curve flux J versus the enzyme concentration $[E_i]$ at the considered steady state.

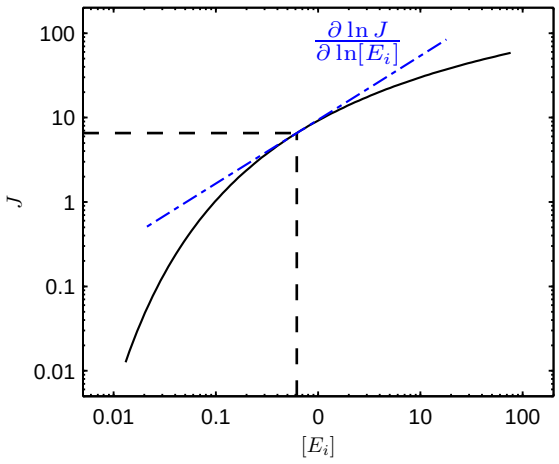


Figure 6.2: Logarithmic representation of flux J of a metabolic pathway as function of the enzyme concentration $[E_i]$ (solid line). The flux control coefficient (6.4) defines the tangent (blue dashed-dotted line) of the curve at a considered steady state (dashed lines).

Taking into account that other parameters than enzyme concentrations may affect reaction rates v_i and, thus, steady-state fluxes, Heinrich and Rapoport [HR73, HR74] defined the flux control coefficient in a more general way:

$$C^J_{v_i} = \frac{v_i}{J}\frac{\partial J/\partial p_k}{\partial v_i/\partial p_k} = \frac{v_i}{J}\frac{\partial J}{\partial v_i} \,, \tag{6.5a}$$

$$= \frac{\partial \ln J/\partial \ln p_k}{\partial \ln v_i/\partial \ln p_k} = \frac{\partial \ln J}{\partial \ln v_i} \,, \tag{6.5b}$$

where p_k is the changing parameter. Mathematically, the fluxes J cannot be directly expressed as function of the rates v_i, thus, the simpler forms at the right are not strictly correct and have to be regarded as an abbreviated notation. However, it is acceptable keeping in mind that there is always an implied external parameter p_k even if it is not shown explicitly.

From Eqs. (6.4) and (6.5) it follows that the flux control coefficient is state dependent. It varies according to the steady state (i.e. the value of J) at which it is measured. Note,

that the analysis assumes that only one parameter, e.g. the concentration of enzyme E_i, is changed and all other parameters remain constant. Hence, each parameter in the considered pathway is characterised by one flux control coefficient. The overall flux of a metabolic pathway is determined by a set of control coefficients, where the number of coefficients is equal to the number of parameters. Usually, the metabolic pathway is investigated in terms of its enzymes. Then, we have one control coefficient for each enzyme.

6.2.2 Concentration control coefficient

Apart from the flux, we can also use the concentrations of intermediary metabolites X_i to characterise the considered pathway. In analogy to the previous section substituting the system variable flux J for the concentration of the metabolites X_i, we obtain the *concentration control coefficient* [HR73, HR74] as

concentration control coefficient

$$C^C_{v_i} = \frac{v_i}{[X_i]} \frac{\partial [X_i]}{\partial v_i}, \tag{6.6a}$$

$$= \frac{\partial \ln [X_i]}{\partial \ln v_i}, \tag{6.6b}$$

where we have to keep in mind that the rates depend on the external parameters p_k[2]. The concentration control coefficient measures the degree of control of an external parameter, as the total enzyme concentration, to the concentration of metabolites. Again, the concentration control coefficient is calculated at steady state.

6.2.3 The summation theorem

In the previous two sections we introduced control coefficients measuring the effect that a change of an enzyme concentration has on the flux or the metabolite concentration. From the definitions of flux control coefficient (Eqs. (6.3) and (6.4)) and concentration control coefficient (Eqs. (6.6a) and (6.6b)) it appears that the control coefficients for different enzymes are independent from each other. However, one can show within the framework of MCA that this is not the case. This dependency conclusion is one of most important results of this approach to biochemical networks. It concerns the *summation of all the control coefficients* of a pathway. By various procedures [KB73, HR75, Gie88, Red88] it can be demonstrated that for a given flux the sum of its flux control coefficients of all steps in the system is constant.

summation theorem

The simplest way in which this property can be derived is by considering a simultaneous small relative increase Δv in all reaction rates of a metabolic system. Since for each metabolite the relative rates of its production increase exactly by the same amount Δv as the relative rates of its consumption, the metabolite concentrations remain unchanged. The flux of the system increases exactly by Δv. This means that the flux is a homogeneous function of first order and the metabolite concentrations homogeneous

[2] See the comment to flux control coefficients below Eq. (6.5).

functions of zeroth order[3]. The summation theorems follow by applying a corollary of the Euler's homogeneous function theorem[4] for homogeneous functions [Gie88]. Due to the normalisation, Euler's theorem simplifies to the *summation theorem for fluxes* [KB73]

flux summation theorem

$$\sum_i C^J_{v_i} = 1 \tag{6.7}$$

and to the *summation theorem for concentrations* [HR75]

concentration summation theorem

$$\sum_i C^{X_j}_{v_i} = 0\,. \tag{6.8}$$

Note, that the sum over the control coefficients is equal to the order of homogeneity of the corresponding system's variable.

It clearly follows from both summation theorems that the control coefficients are not independent of each other. If, for example, one coefficient increases, one or more of the other coefficients have to decrease.

If and only if a flux control coefficient $C^J_j \approx 1$ and all other $C_{i\neq j} \approx 0$, then the j-th enzymatic reaction is rate-determining. In general, the control of the pathway may be distributed over more than one steps. This is summarised in the famous dictum – "*Control can be shared!*"

However, the flux summation theorem does not restrict the flux control coefficients to the interval $[0;1]$. Some coefficients may well be negative and some coefficients of the considered pathway can exceed unity.

For a concentration control coefficient $C^{X_i}_{E_i} \approx 0$, the enzymatic reaction is limited by other factors than the enzyme concentration, and a change of the enzyme concentration does not change the intermediate concentration.

6.2.3.1 Number of summation theorems

The number of summation theorems is determined by the structure of the investigated metabolic pathway. A summation theorem exists for each systems variable, i.e. flux. A short list is given in the following table

[3] A function $f(\alpha\, x_1, \alpha\, x_2, \ldots)$ is homogeneous of order n if the equation

$$f(\alpha\, x_1, \alpha\, x_2, \ldots, \alpha\, x_k) = \alpha^n\, f(x_1, x_2, \ldots, x_k)$$

holds for all α, $x_i \in \mathbb{R}$.

[4] Euler's homogeneous function theorem

$$n \cdot f(x_1, x_2, \ldots, x_k) = \sum_{i=1}^{k} \frac{\partial f}{\partial x_i} x_i$$

relates a homogeneous function $f(x_i)$ to its partial derivatives, its coordinates x_i (i.e. enzyme concentrations), and its order of homogeneity n.

structure	# of fluxes	# of summation theorems
linear pathway	1	1
two branches	2	2
⋮	...	...

The number of summation theorems will increase with the complexity of metabolic pathways.

6.3 Metabolic Control Analysis and Rate-Determining Step

In previous sections we already mentioned the principle of a rate-determining step. In metabolic control analysis this principle leads to a controversial discussion, see for instance [Fel97]. The origin is a traditional view of metabolic control which assumes that there has to be a rate-limiting step in a metabolic pathway. On the other hand, from the summation theorem (6.7) follow that control can be distributed over several enzymes. This is also supported by some experiment, i.e. see [GWW+82]. Before we continue our discussion, we will cite two different definitions of the principle of rate-determining step. The first one is from Blackman (1905) [Bla05]:

> *When a process is conditioned as to its rapidity by a number of separate factors, the rate of the process is limited by the pace of the slowest factor.*

and the second definition is a combination from the famous book of Atkins, "Atkins' Physical Chemistry" [AdP02] and from the book of Schwetlik [SDPS73] about Chemical Kinetics :

> *If there is in a consecutive reaction a step much slower than all subsequent steps,* see Figure 6.3[a] and [b], *this step is rate-determining. The rate-determining step controls the rate of all subsequent steps. However, the rate-determining step is not just the slowest step: it must be slow and be a crucial gateway for the formation of products. If a faster reaction can also lead to products, then the slowest step is irrelevant because the slow reaction can then be side-stepped,* as in the example in Figure 6.3[c].

In comparison to the modern definition of the rate-determining step in Chemical Kinetics, the definition of Blackman is vague. On the one hand, the kind of process or biochemical reaction is not specified. On the other hand, the slowest factor is a very imprecise specification. Considering, one factor may have a value of $f_1 = 10$ and another the value $f_2 = 10.5$. Then, both factors are similar. According to Blackman, the factor f_1 is rate-determining. However, as investigations of consecutive reactions show, see Section 2.6.3, similar reaction rates do not fulfill the principle of a rate-determining step.

In general, the principle of rate-determining step is defined only for consecutive reactions or processes[5]. These kind of processes have at least one irreversible step [SDPS73].

[5] For a more detailed discussion of complex reactions see Section 2.6 and for consecutive reactions Section 2.6.3.

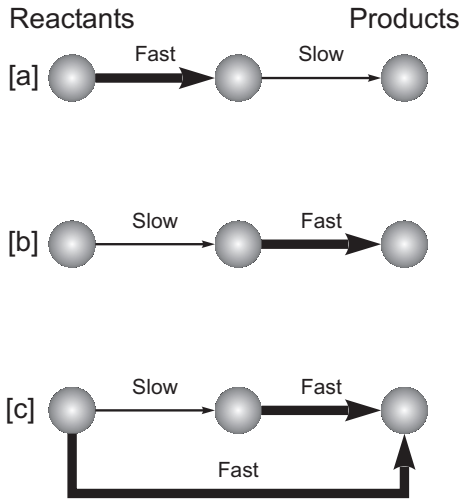


Figure 6.3: Rate-determining steps in consecutive reactions. Heavy arrows represent fast steps and light arrows represent slow steps. [a] The second step is rate-determining; [b] the first step is rate determining; [c] although one step is slow, it is not rate-determining because there is a fast route that circumvents it.

Then the principle is a direct consequence from the principle of cause and effect. The cause triggers the effect.

Whereas this condition is clearly fulfilled in industrial production processes, where it leads to the so-called bottleneck problem, it is rare in chemical processes. A further well-known example for a rate-determining step can be often observed at construction areas of highways. At the construction area the lanes are reduced, for instance from three lanes to a single lane. If the flux of cars exceed the passage limit of the bottleneck it determines the rate of all subsequent events. consequently, this will result in an increasing traffic jam before the construction area. In metabolic pathways such bottlenecks are usually the result of some limitations, for example in food, light, or necessary additional reactants. Only, in these limiting cases a metabolic pathway may have a rate-determining step. However, this is not the usual case. Interestingly, Blackman [Bla05] mentioned this already in 1905.

However, the modern definition characterises only the necessary conditions for a rate-determining step, it does not predict the existence of such a step. An *a priori* assumption of a rate-determining step is a misinterpretation of the principle.

In the investigation of the dynamics of a biochemical system, the principle of a rate-determining step plays a major role, although it is often not obvious. Whereas in consecutive reactions the rate-determining step controls the overall reaction rate, the situation in more complex biochemical reaction networks is different. Here, there is usually not a global rate-determining step, similar to metabolic networks. If we model the network, we decompose it into a set of smaller subnetworks. Some of them may consist of consecutive reactions. For these reactions, local rate-determining steps can exist that enable a simplification of the mathematical representation. As example, the Michaelis-Menten

equation 2.150, discussed in Section 2.7.1, is often used as a template to describe the conversion, the (de)activation or degradation of a protein. Here, the dissociation of the enzyme-substrate complex into product and enzyme has to be rate-determining, otherwise the quasi-steady-state assumption (see Section 2.6.3.2) is not applicable and the Michaelis-Menten equation is not an appropriate description of the dynamics. Nevertheless, the identification of a rate-determining step is crucial because it is dependent on the current situation. During the time course or through a change in the environmental conditions, the local rate-determining step may change or disappear.

6.4 Elasticity Coefficients

In Section 6.2 we investigated how a change in enzyme concentrations changes the flux and the metabolite concentrations. These are properties of the whole intact metabolic pathway, referred to as global properties. In contrast, biochemists have traditionally studied enzymes in isolation, as represented in Fig. 6.4. For this reason, we now investigate, how MCA can be used to describe the effect of changes in substrate, product or enzyme concentration on the individual reaction rates v_i. Furthermore, the rate v_i

$$X_1 \underset{E_1}{\overset{v_1}{\rightleftharpoons}} X_2$$

Figure 6.4: Simple example of an isolated intermediary step of a metabolic pathway. The metabolite X_1 is converted to X_2 with the rate v_1. The reaction is facilitated by the enzyme E_1.

elasticity coefficient

depends on external parameters, such as kinetic coefficients. The *elasticity coefficients* measure the changes to the individual reaction as a consequence of a change of these parameters. According to Heinrich and Schuster [HS96], we distinguish between to kinds of parameters: a) ε-elasticities describing the effects of changes in the concentrations of participating species, and b) π-elasticities describing the effects of varying kinetic coefficients to the rate of change. The elasticities are local or systemic coefficients because they consider only an isolated reaction.

6.4.1 Elasticity coefficients for metabolites

The rate of change of the conversion of a metabolite X_i to a metabolite X_{i+1} depends on the concentrations of the metabolites, of the enzymes facilitating the reaction, and of inhibitors or further influencing proteins. The elasticity coefficients are defined as [KB73, BCBG+85]:

$$\varepsilon_{v_i}^{X_j} = \frac{[X_i]}{v_j} \frac{\partial v_j}{\partial [X_i]}\,, \tag{6.9}$$

where the subscript i denotes the i-th metabolite X_i and the superscript j the rate v_j of the j-th reaction. The derivation of this expression is equivalent to the control coefficients represented in Section 6.2.1. The ε-elasticity quantifies how the reaction rate is modulated by the metabolite X_i. A positive coefficient corresponds to an increase of

the rate and a negative coefficient to a decrease. A zero-valued elasticity coefficient may have two different causes. First, the metabolite X_i is no part of the considered reaction and, therefore, the rate remains unchanged. Second, the metabolite X_i participating on the reaction but does not change the rate. This is common for reactions that follow a Michaelis-Menten regime. If the enzyme is saturated, a further increase of the metabolite does no longer increase the reaction rate. For a more detailed discussion of different reaction mechanisms and their related elasticities, see Section 6.9.

6.4.2 Elasticity coefficients for parameters

As mentioned earlier, the rates v_i may also depend on external parameters. Kinetic parameters, such as rate coefficients and Michaelis-Menten constants, are functions of environmental conditions, e.g. temperature, pH-value or water balance. They vary over time and experiments. Hence, the rates v_i change if kinetic parameters change. Additionally, the reaction rates also depend on some model parameters. Depending on the used framework, these could be for example Hill parameters or kinetic orders. Due to the uncertainty of experimental data, these model parameters can be determined only with a certain limited accuracy. Within the confidence interval the values may change. Therefore, it is important to investigate how these changes affect the rate of the considered reaction.

A relative measure of how a variation in external parameters modulate the rate is the $\pi - elasticity$. It is defined as [KSA90, SK90]

$$\pi_{p_k}^{v_i} = \frac{p_k}{v_i} \frac{\partial v_i}{\partial p_k}, \tag{6.10}$$

where p_k is the changing external parameter and v_i the considered rate. If the π-elasticity is positive, the rate v_i increases if the parameter p_k increases. For negative values it decreases if the parameter decreases. If the π-elasticity is zero, the parameter p_k does not affect the rate v_i. Either no step of the considered reaction is determined by this parameter or the rate is fully determined by other factors, e.g. a rate-determining step which is not affected by p_k.

6.4.3 The connectivity theorem

One of the major objectives of MCA is to reveal how the properties of a whole system, i.e. its control coefficients, can be explained in terms of the properties of its components, i.e. in terms of the elasticities. This is especially important, because the control coefficients are usually not easy to measure. However, many data are available for isolated (in vivo) enzymatic reactions. Thus, one is trying to assess the systemic behavior that is of fundamental interest but difficult to measure in terms of quantities that can be measured very easily. It turns out that for any pathway there are exact relationships between the control coefficients and the ϵ-elasticities. The first *connectivity theorem* relates the flux control coefficients and the ϵ-elasticities in a sum over products between both coefficients.

connectivity theorem

It was found by Kacser and Burns as [KB73]:

$$\sum_i C^J_{v_i} \varepsilon^{X_j}_{v_i} = 0\,. \tag{6.11}$$

As opposed to the summation theorem (6.7), the flux connectivity theorem depends on the kinetic expressions through the ϵ-elasticities.

An analogous theorem was derived for metabolite concentrations [WC84]:

$$\sum_i C^{X_j}_i \varepsilon^{X_j}_{v_i} = -1\,. \tag{6.12}$$

It connects the concentration control coefficients $C^{X_j}_i$ to the elasticities of metabolites that participate in the i-th step of the pathway.

With summation theorems (6.7) for fluxes and (6.8) for concentrations and the above connectivity theorems, we have now a tool to investigate metabolic pathways from their global and their local properties. Moreover, we are able to determine control coefficients from elasticity coefficients using these relations. In the next section, we demonstrate how one can use summation and connectivity theorems on the example of a linear pathway. Later on, we generalise the treatment for arbitrary structures introducing the matrix representation of metabolic control analysis.

6.5 Using the Summation and the Connectivity Theorem

As mentioned earlier, the connectivity theorems (6.11) and (6.12) relate the control coefficients and the elasticities. For linear pathways, this fact can be used to express the control coefficients in terms of elasticity coefficients. As an example, we investigate a three step pathway which converts the substrate S into the product P via the intermediary metabolites X_1 and X_2, see Fig. 6.5. We restrict the analysis to the flux control coefficients, but an analogues treatment is possible for concentrations control coefficients.

$$S \underset{E_1}{\overset{v_1}{\rightleftharpoons}} X_1 \underset{E_2}{\overset{v_2}{\rightleftharpoons}} X_2 \underset{E_3}{\overset{v_3}{\rightleftharpoons}} P$$

Figure 6.5: Linear three step pathway converting substrate S into product P. v_i are the reaction rates of the conversion reactions, E_i the catalyzing enzymes and X_i intermediary metabolites.

Since, we consider a linear pathway, there is only one flux and, thus, we have only one summation theorem. Nevertheless, due to the two intermediates, we have to take into account two connectivity theorems, one for the metabolite X_1 and a second for the metabolite X_2. The corresponding theorems are:

Summation theorem (6.7):	$C^J_1 + C^J_2 + C^J_3 = 1$	(6.13a)
Connectivity theorem (6.11) for X_1:	$C^J_1 \varepsilon^1_{X_1} + C^J_2 \varepsilon^2_{X_1} = 0$	(6.13b)
Connectivity theorem (6.11) for X_2:	$C^J_2 \varepsilon^2_{X_2} + C^J_3 \varepsilon^3_{X_2} = 0$	(6.13c)

where we can rearrange both connectivity relations with respect to the flux control coefficients:

$$C_1^J = -C_2^J \frac{\varepsilon_{X_1}^2}{\varepsilon_{X_1}^1}$$

and

$$C_3^J = -C_2^J \frac{\varepsilon_{X_2}^2}{\varepsilon_{X_2}^3} .$$

Substitution into the summation theorem (6.13a) leads to the equation

$$-C_2^J \frac{\varepsilon_{X_1}^2}{\varepsilon_{X_1}^1} + C_2^J - C_2^J \frac{\varepsilon_{X_2}^2}{\varepsilon_{X_2}^3} = 1 ,$$

holding only on flux control coefficient C_2^J and all elasticities. A transformation with respect to the flux control coefficient leads to the final result

$$C_2^J = \frac{\varepsilon_{X_1}^1 \varepsilon_{X_2}^3}{\varepsilon_{X_1}^1 \varepsilon_{X_2}^3 - \varepsilon_{X_1}^2 \varepsilon_{X_2}^3 - \varepsilon_{X_1}^1 \varepsilon_{X_2}^2} \tag{6.14}$$

for the second control coefficient. It combines global properties (C_2^J) with local properties described by the elasticities ε_i^j. These local properties can be measured by in vitro enzyme kinetic experiments and can be applied around the considered steady state. Nevertheless, a support with in vivo data is strongly recommended to exclude differences for in vitro and in vivo experiments.

Similar expressions can be derived for the remaining flux control coefficients C_1^J and C_3^J. For real metabolic pathways, with many intermediary metabolites and branching points, this method becomes intractable very fast. Moreover, it is not accessible for systematic numerical calculation. Hence, we want to generalise the method in the next section within a matrix approach that resolves the mentioned disadvantages.

6.6 Matrix Methods in Metabolic Control Analysis

As shown in the previous Section 6.5, the combination of summation and connectivity theorem enables the determination of global flux control coefficients through the local elasticity coefficients. Furthermore, if we consider all the theorems for an investigated metabolic pathway, we obtain a system of n^2 coupled equations. In this system, n is the number of all participating metabolites, where the substrate is the zeroth component and the product the n-th component, as in the generalised linear pathway

$$\underbrace{X_0}_{\text{substrate}} \rightleftharpoons X_1 \rightleftharpoons X_2 \rightleftharpoons \cdots \rightleftharpoons X_{n-1} \rightleftharpoons \underbrace{X_n}_{\text{product}}$$

It is clear that with increasing complexity of metabolic pathways the treatment introduced in the last section becomes intractable. Hence, we use a more convenient and

elegant method to represent the metabolic pathway in terms of the Metabolic Control Analysis. The matrix approach was originally introduced by [Red88] and developed further e.g. by [SSF87, SF89, HCBR93, HS96, HW01].

Prior we introduce the matrix approach in more detail, we summarise the theorems of MCA, that play a central role in this approach. First, there are the summation theorems

$$\sum_i C_i^J = 1 \qquad \text{Flux Control Coefficients} \tag{6.15a}$$

$$\sum_i C_i^{X_j} = 0 \qquad \text{Concentration Control Coefficients} \tag{6.15b}$$

for the flux and the concentration control coefficients. The connectivity theorems relate control coefficients to elasticity coefficients. For the combination of the flux related coefficients, they hold:

$$\sum_i C_i^J \epsilon^i_{X_j} = 0\,. \tag{6.16}$$

The connectivity theorem for concentrations (6.12) has to be generalised in order to distinguish between perturbed and unperturbed metabolites [WC84]. The theorem for the perturbed metabolite X_j was introduced in Section 6.4.3 as

$$\sum_i C_i^{X_j} \varepsilon^i_{X_j} = -1\,.$$

However, we have to consider also the case where the metabolite X_j is not perturbed. Then, the connectivity theorem becomes:

$$\sum_i C_i^{X_j} \varepsilon^i_{X_k} = 0\,.$$

Summarising both equations, we obtain the generalised version of the connectivity theorem [WC84]

$$\sum_i C_i^{X_j} \varepsilon^i_{X_k} = -\delta_{jk}\,, \tag{6.17}$$

where δ_{ij} is the Kronecker-Delta [AS72]

$$\delta_{jk} = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases}\,.$$

These information we now combine into in some matrices. This allows us to use matrix algebra to derive control coefficients in terms of elasticity coefficients. Since the introductional nature of this chapter, we here follow a more practical guide [SSF87, CB95b] instead of the theoretical approach based on structural properties of the pathway in [Red88, HCBR93, HS96, HW01].

The information about the metabolic pathway are represented in three quadratic matrices of order n. In the so-called E-matrix we collect elasticities, where the rows

contains information about the fluxes J_i and the metabolites X_i. The rates are arranged over the columns of the matrix. The C-Matrix contains the control coefficients of the considered system. The arrangement with respect to the fluxes, metabolites and rates is now transposed in comparison to the E-matrix. Matrix multiplication of both matrices results in the M-matrix. In fact, this resulting matrix consists of a combination of the theorems which we summarised above. Finally, the metabolic pathway under investigation is described by the matrix equation

$$\begin{bmatrix} \varepsilon_1^1 & \varepsilon_1^2 & \cdots & \varepsilon_1^n \\ \varepsilon_2^1 & \varepsilon_2^2 & \cdots & \varepsilon_2^n \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_n^1 & \varepsilon_n^2 & \cdots & \varepsilon_n^n \end{bmatrix} \begin{bmatrix} C_1^1 & C_1^2 & \cdots & C_1^n \\ C_2^1 & C_2^2 & \cdots & C_2^n \\ \vdots & \vdots & \ddots & \vdots \\ C_n^1 & C_n^2 & \cdots & C_n^n \end{bmatrix} = \begin{bmatrix} M_1^1 & M_1^2 & \cdots & M_1^n \\ M_2^1 & M_2^2 & \cdots & M_2^n \\ \vdots & \vdots & \ddots & \vdots \\ M_n^1 & M_n^2 & \cdots & M_n^n \end{bmatrix} \tag{6.18}$$

or in its short representation

$$\mathbf{E} \cdot \mathbf{C} = \mathbf{M} \,. \tag{6.19}$$

This compact representation of the metabolic control analysis has some advantages over the classical treatment of the previous section: I.) The matrix representation makes it better amenable to numerical and symbolic approaches with suitable software packages or programming languages and to analytical investigation. II.) The matrices can be constructed directly from stoichiometric data as is shown in [Red88, HS96]. Hence, this approach can be connected to graphical tools.

Whereas the M-matrix is determined by the summation (6.15a) and connectivity theorems (6.17), and the E-matrix by local properties of the pathway, the C-matrix and its elements are unknown. Therefore, we are now seeking an expression for $\mathbf{C}$ in terms of the other matrices. To this end, we calculate the inverse matrix of $\mathbf{E}$ which is defined by the relation

$$\mathbf{E} \cdot \mathbf{E}^{-\mathbf{1}} = \mathbf{1} \,.$$

The multiplication of a matrix with its inverse results in the identity matrix. The inverse matrix we multiply from left to matrix equation (6.19)

$$\mathbf{C} = \mathbf{E}^{-\mathbf{1}} \cdot \mathbf{M} \tag{6.20}$$

and obtain the matrix of control coefficients as the product of the inverse of the E-matrix and the M-matrix[6].

In the above definition (6.18) we only introduced the elements of this matrix. Now, we give some rules for the construction of the matrices. Towards this end, we relate the matrix elements to the properties of the metabolic pathway. Here, we distinguish between two kinds of elements, flux related and elasticity related. If an element is part of a row describing a flux through the metabolic pathway, then its value is simply 1. In order

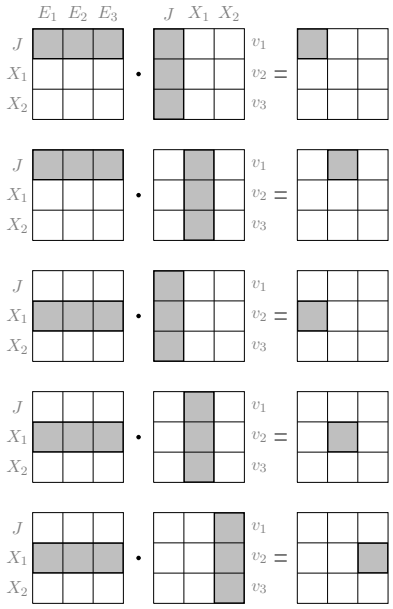
[6] Note the order of multiplication of matrices! Generally, the matrix multiplication is not commutative. This means

$$\mathbf{A} \cdot \mathbf{B} \neq \mathbf{B} \cdot \mathbf{A} \,.$$

Hence, we distinguish between multiplication from left and multiplication from right.

to determine the elasticity related elements we have to consider the current metabolite X_i and the current rate v_j. If the metabolite X_i participate on the j-th reaction, then the value of element ε_i^j follows from the definition of the elasticity coefficient (6.9), else it is zero.

After defining the structure and the values of the E-matrix, we can now calculate the M-matrix. The element-wise calculation of the elements M_i^j reproduces exactly the theorems (6.15a) and (6.17). Thus, we can represent the derivation of the matrix in a simple schematic way [CB95b].



$$\sum_{i=1}^{n} C_i^J = 1$$ Flux control coefficient summation theorem

$$\sum_{i=1}^{n} C_i^{X_j} = 0$$ Concentration control coefficient summation theorem

$$\sum_{i=1}^{n} C_i^J \varepsilon_{X_j}^i = 0$$ Flux connectivity theorem

$$\sum_{i=1}^{n} C_i^{X_j} \varepsilon_j^i = -1$$ Concentration connectivity theorem (homogeneous)

$$\sum_{i=1}^{n} C_i^{X_j} \varepsilon_k^i = 0$$ Concentration connectivity theorem (heterogeneous)

Figure 6.6: Schematic representation of the element-wise calculation of the M-matrix [CB95a]. It is obvious that the matrix equation (6.19) contains all information about the metabolic pathway which we derived in the previous sections. In general, for a system of n enzymes there are n^2 such relations which fall into five classes expressed by Eqs. (6.15 - 6.17).

Note, that in principle the arrangement of the elements is arbitrary, but most authors prefer the convenient rule that one starts with fluxes in the first rows followed by the metabolites. Also, some authors use the matrix equation

$$\mathbf{C} \cdot \mathbf{E} = \mathbf{M}$$

instead of (6.19). Due to the permuted order of the matrix multiplication one has to use the transposed versions of the E- and C-matrix which was introduced in (6.18). Furthermore, the multiplication with the inverse E-matrix have to carried from right.

In the next subsections, we apply the matrix method to the example of an unbranched and a branched pathway.

6.6.1 Control coefficients in an unbranched pathway

The introduced matrix approach is used to investigate an unbranched pathway. We consider again a linear metabolic pathway with two intermediary metabolites (Figure 6.7). The corresponding matrix equation is

$$S \underset{E_1}{\overset{v_1}{\rightleftharpoons}} X_1 \underset{E_2}{\overset{v_2}{\rightleftharpoons}} X_2 \underset{E_3}{\overset{v_3}{\rightleftharpoons}} P$$

Figure 6.7: Three step unbranched pathway. The v_i are the reaction rates of the conversion reactions, E_i are the catalyzing enzymes and X_i are intermediary metabolites.

$$\begin{array}{c} \\ J_1 \\ X_1 \\ X_2 \end{array} \begin{array}{c} \begin{array}{ccc} E_1 & E_2 & E_3 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 \\ \varepsilon_1^1 & \varepsilon_1^2 & 0 \\ 0 & \varepsilon_2^2 & \varepsilon_2^3 \end{bmatrix} \end{array} \begin{array}{c} \begin{array}{ccc} J_1 & X_1 & X_2 \end{array} \\ \begin{bmatrix} C_1^J & C_1^{X_1} & C_1^{X_2} \\ C_2^J & C_2^{X_1} & C_2^{X_2} \\ C_3^J & C_3^{X_1} & C_3^{X_2} \end{bmatrix} \end{array} \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \end{array} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} . \quad (6.21)$$

Due to the linear character of the pathway, there is only one flux which is determined by the three rates v_i. Furthermore three enzymes E_i affect these rates. Hence, the matrix equation consists of matrices of the order 3×3. As noted in Eq. (6.21), the first row of the E-matrix is related to flux J and is, therefore, a unity matrix. The other two rows describe the elasticities of the reactions involved in the linear pathway. The elements ε_1^3 and ε_2^3 are zero because the enzyme E_3 does not affect the metabolite X_1 and the enzyme E_1 does not affect the metabolite X_3. We conclude that in general a metabolite of a linear pathway has only two non-vanishing elasticities, namely one elasticity for the producing reaction and the second one for the degrading reaction. The control coefficients for the metabolic pathway are collected in the C-matrix. The elements of the first column are the flux control coefficients. The remaining columns consist of the concentration control coefficients with respect to the metabolites X_1 and X_2. The M-matrix can be constructed according to the scheme represented in Figure 6.6. The first column of the M-matrix is given by the flux summation theorem (6.15a) and the second and third element by the connectivity theorem between flux control coefficients and elasticities. All non-diagonal elements of the next columns vanish as a consequence of the concentration summation theorem (6.15b) and the connectivity theorem (6.17) for the heterogenous case. The diagonal elements are determined by the control coefficients and the elasticities for the same metabolite (homogenous). Hence, these matrix elements have the value -1. Note also the diagonal structure of the M-matrix. This is typical for linear pathways. Due to their special stoichiometric form, all non-diagonal elements have to equal zero.

After we constructed all three matrices, we now calculate the inverse E-matrix

$$E^{-1} = \frac{1}{\varepsilon_1^1\varepsilon_2^2 - \varepsilon_1^1\varepsilon_2^3 + \varepsilon_1^2\varepsilon_2^3} \begin{bmatrix} \varepsilon_1^2\varepsilon_2^3 & \varepsilon_2^2 - \varepsilon_2^3 & -\varepsilon_1^2 \\ -\varepsilon_1^1\varepsilon_2^3 & \varepsilon_2^3 & \varepsilon_1^1 \\ \varepsilon_1^1\varepsilon_2^2 & \varepsilon_2^2 & \varepsilon_1^2 - \varepsilon_1^1 \end{bmatrix} \quad (6.22)$$

which we multiply from left with matrix equation (6.21). As result we obtain the matrix of control coefficients in terms of a new matrix

$$\begin{bmatrix} C_1^J & C_1^{X_1} & C_1^{X_2} \\ C_2^J & C_2^{X_1} & C_2^{X_2} \\ C_3^J & C_3^{X_3} & C_3^{X_3} \end{bmatrix} = \frac{1}{\varepsilon_1^1\varepsilon_2^2 - \varepsilon_1^1\varepsilon_2^3 + \varepsilon_1^2\varepsilon_2^3} \begin{bmatrix} \varepsilon_1^2\varepsilon_2^3 & \varepsilon_2^2 - \varepsilon_2^3 & -\varepsilon_1^2 \\ -\varepsilon_1^1\varepsilon_2^3 & \varepsilon_2^3 & \varepsilon_1^1 \\ \varepsilon_1^1\varepsilon_2^2 & -\varepsilon_2^2 & \varepsilon_1^2 - \varepsilon_1^1 \end{bmatrix} \tag{6.23}$$

consisting of combinations of the elasticity coefficients. If analytical rate laws are known, the control coefficients can be obtained simply by calculating the partial derivatives of the definitions of the elasticity coefficient (6.9). Otherwise, they have to be measured experimentally.

6.6.2 Control coefficients in a branched pathway

As a further example, we consider a branched pathway. The substrate S is converted via the intermediate X_1 into the products P_1 and P_2 (Figure 6.8). Because of the branching

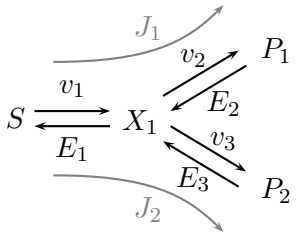


Figure 6.8: An example of a branched pathway. The net rate of each step is denoted as v_i, wheras E_i is the facilitating enzyme. Note, due to the branching point, the flux is splitting into two parts, J_1 and J_2.

point we now have to consider two fluxes J_1 and J_2. An important property of this branching point follows from the mass conservation in chemical reactions. The flux through the reaction v_1 has to be equal to the sum of the fluxes through the reactions v_2 and v_3. This fact is a chemical equivalent to Kirchhoff's law for electrical circuits and is expressed mathematically as:

$$\sum_{i=1}^{n} J_i = 0\,. \tag{6.24}$$

Furthermore, we follow the usual convention that ingoing fluxes are positive and outgoing fluxes are negative. This law (6.24) can be used to search for missing branches or to estimate the crosstalk between different pathways. If one knows, e.g. all ingoing fluxes, and their sum is not equal to the sum of all considered outgoing fluxes, it is probable that there are further outgoing fluxes which one has to consider. On the other hand, one can use the normalised outgoing fluxes to estimate the crosstalk or the contribution of branches to the whole investigated system, whereby the normalisation factor is the sum over all ingoing fluxes.

However, the matrix equation of the branched pathway of Figure 6.8 is:

$$\begin{array}{c} \\ J_1 \\ J_2 \\ X_1 \end{array} \begin{array}{c} \begin{array}{ccc} E_1 & E_2 & E_3 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ \varepsilon_1^{X_1} & \varepsilon_2^{X_1} & \varepsilon_3^{X_1} \end{bmatrix} \end{array} \begin{array}{c} \begin{array}{ccc} J_1 & J_2 & X_1 \end{array} \\ \begin{bmatrix} C_1^{J_1} & C_1^{J_2} & C_1^{X_1} \\ C_2^{J_1} & C_2^{J_2} & C_2^{X_1} \\ C_3^{J_1} & C_3^{J_2} & C_3^{X_1} \end{bmatrix} \end{array} \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \end{array} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \tag{6.25}$$

with consideration of two fluxes and the metabolite. In contrast to the E-matrix for an unbranched pathway, the flux related rows are not unity matrices. The element in the matrix corresponding to J_1 and E_3 has to be zero, because the reaction catalyzed by the enzyme E_3 have no contribution to the flux J_1. In analogy, the element ε_2^2 has to be zero, too. The metabolite X_1 participates on all reactions, and thus it has three elasticities. The construction of the M-matrix follows again the scheme 6.6. The inverse of E-Matrix is then

$$E^{-1} = \frac{1}{\varepsilon_1^{X_1} - \varepsilon_2^{X_1} - \varepsilon_3^{X_1}} \begin{bmatrix} -\varepsilon_2^{X_1} & -\varepsilon_3^{X_1} & 1 \\ \varepsilon_1^{X_1} - \varepsilon_3^{X_1} & \varepsilon_3^{X_1} & -1 \\ \varepsilon_2^{X_1} & \varepsilon_1^{X_1} - \varepsilon_2^{X_1} & -1 \end{bmatrix} \tag{6.26}$$

which leads finally to a C-Matrix of

$$\begin{bmatrix} C_1^{J_1} & C_1^{J_2} & C_1^{X_1} \\ C_2^{J_1} & C_2^{J_2} & C_2^{X_1} \\ C_3^{J_1} & C_3^{J_2} & C_3^{X_1} \end{bmatrix} = -\frac{1}{\varepsilon_1^{X_1} - \varepsilon_2^{X_1} - \varepsilon_3^{X_1}} \begin{bmatrix} \varepsilon_2^{X_1} + \varepsilon_3^{X_1} & \varepsilon_2^{X_1} + \varepsilon_3^{X_1} & 1 \\ -\varepsilon_1^{X_1} & -\varepsilon_1^{X_1} & -1 \\ -\varepsilon_1^{X_1} & -\varepsilon_1^{X_1} & -1 \end{bmatrix}. \tag{6.27}$$

The control coefficients of the branched pathway are expressed in terms of elasticities which can be measured experimentally.

6.7 Response Coefficients

Studies of metabolism have identified different mechanisms that regulate and coordinate the activity of metabolic pathways. Generally, these mechanisms involve 'effectors'. These effectors may originate 'internally' from the pathway (e.g. a pathway intermediate) or 'externally' (e.g. a hormone, drug or change in the physical conditions).

Analogous to a control coefficient expressing the dependence of a system variable, such as flux, on an internal parameter, such as enzyme concentration, one can derive a *response coefficient* R_z^J to express the dependence of a system variable on an external parameter, such as the concentration $[Z]$ of an external effector Z. Towards this end, we define the flux as a multiparametric function

response coefficient

$$J(k_1, \ldots, k_i; X_1, \ldots, X_j; E_1, \ldots, E_k; Z, T, pH, ...) = J(\xi_i)$$

and collect all the different parameters in the parameter set ξ. The effect of a change in the parameter P on the flux can be evaluated applying the chain rule

$$\frac{\partial J}{\partial P} = \sum_i \frac{\partial J}{\partial \xi_i} \frac{\partial \xi_i}{\partial P}. \tag{6.28}$$

Hence the parameter P may affect more parameters we have to sum over the whole parameter set ξ in the above equation.

To normalize the expression (6.28) we now transform the left hand side

$$\begin{aligned}\frac{\partial J}{\partial P} &= \sum_i \frac{\partial J}{\partial \xi_i}\frac{\xi_i}{J}\frac{\partial \xi_i}{\partial P}\frac{J}{\xi_i} \\ &= \sum_i \frac{\partial \ln J}{\partial \ln \xi_i}\frac{\partial \ln \xi_i}{\partial P} J\end{aligned}$$

and multiplay the equation with P and divide it with J which leads to the expression

$$\begin{aligned}\frac{P}{J}\frac{\partial J}{\partial P} &= \sum_i \frac{\partial \ln J}{\partial \ln \xi_i}\frac{\partial \ln \xi_i}{\partial P} P \\ \underbrace{\frac{\partial \ln J}{\partial \ln P}}_{R_P^J} &= \sum_i \underbrace{\frac{\partial \ln J}{\partial \ln \xi_i}}_{C_i^J}\underbrace{\frac{\partial \ln \xi_i}{\partial \ln P}}_{\varepsilon_i^P} .\end{aligned}$$

The response coefficient is then defined as the logarithmic derivative

$$R_P^J = \frac{\partial \ln J}{\partial} \ln P \,. \tag{6.29}$$

which can be calculated as the product of corresponding control coefficients and elasticities

$$R_P^J = \sum_i C_i^P \varepsilon_i^P \tag{6.30}$$

$$= \sum_i \frac{\partial \ln J}{\partial \ln \xi_i}\frac{\partial \ln \xi_i}{\partial \ln P} \tag{6.31}$$

In the case of an external effector that acts on a single enzyme, the degree to which it influences the pathway flux depends on the degree to which it affects the activity of the target enzyme, expressed by its elasticity, and the degree of control this enzyme has on the pathway flux, expressed by the flux control coefficient of the enzyme. This relationship was proven by Kacser and Burns [KB73] and formulated in the so-called *partitioned response*

partitioned response

$$R_{P,i}^J = C_i^J \varepsilon_i^P \,. \tag{6.32}$$

The partitioned response generalizes our previous definitions of control and elasticity coefficients. It follows from Eq. (6.32) that all the introduced coefficients can be treated as response coefficients. Let's consider the change of a single enzyme concentration E_j. The corresponding response coefficient is given by Eq. (6.31). Because of the enzyme E_i affects no other parameters of the parameter set ξ the second derivative in Eq. (6.31) follows by the relation

$$\frac{\partial \ln \xi_i}{\partial \ln E_i} = \begin{cases} 0 & \text{for} \quad \xi_i \neq E_i \,, \\ 1 & \text{for} \quad \xi_i = E_i \,. \end{cases} \tag{6.33}$$

As consequence the some in the definition of the response coefficient (6.30) has only one non-zero element which is the partioned response with respect to the considered enzyme E_i. Additionally, relation (6.33) simplifies Eq. (6.32) to

$$R^J_{E_i} = \frac{\partial \ln J}{\partial \ln E_i} = C^J_i \,. \tag{6.34}$$

The partioned response of the enzyme E_i is determined by the control coefficient only. In general, if the parameter ξ_i affects no further parameters, the response coefficient reduces to a control or elasticity coefficient.

6.8 Limits of Metabolic Control Analysis

At present, metabolic control analysis is largely concerned with steady states. Nevertheless, there are first approaches to extend it to transient phenomena [KC82, ASK89, MHTSK90, HR91, IS03, HBB+05].

Moreover, it assumes that a steady state exists and that it is stable and unique. As it is relatively easy to construct models, even for systems of only two enzymes, that have no steady state, one should not assume the existence of a steady state automatically.

Compartmentation might also seem to pose a problem but, in fact, MCA can readily accommodate multiple compartments as long as there are well defined connections between them and the components are distributed homogeneously within each compartment.

Some further restrictions follow from the mathematical framework of MCA. It deals with infinitesimal changes only, as it was shown in the schematic derivation in Section 6.2. Hence, it cannot predict the consequences of large changes in an experiment. Analytical considerations are more flexible but once a certain steady state was chosen as reference point, one is restricted to its neighborhood. This leads directly to the next point. As mentioned earlier, the coefficients are state depended and cannot be assigned to other situations.

6.9 Elasticity Coefficients for Specific Rate Laws

As we demonstrated in the previous sections, the control coefficients are closely related to the elasticities of an isolated reaction in the metabolic pathway. Therefore, we want to discuss the elasticities of some well-known rate laws in the this section [HS96].

6.9.1 Michaelis-Menten equation

A commonly used reaction motif of a conversion of a substrate into a product was proposed by Michaelis and Menten [MM13] as follows:

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_2} P + E \,, \tag{6.35}$$

where the enzyme E catalyzes the conversion of the substrate S into the product P. One can derive the famous Michaelis-Menten equation

$$v = \frac{V_{\max} S}{K_{\mathrm{M}} + S} \tag{6.36}$$

for the reaction rate if one assumes the intermediary complex C in a quasi-steady state (see Sec. 2.7.1 for a more detailed discussion). $V_{\max} = k_2 E^{\mathrm{T}}$ is the limiting rate, E^{T} the total enzyme concentration, and $K_{\mathrm{M}} = (k_{-1} + k_2)/k_1$ the Michaelis constant. According to the definition (6.9), the elasticity with respect to the substrate is

$$\begin{aligned} \varepsilon_{\mathrm{S}} &= \frac{S}{v}\frac{\partial v}{\partial S}\,, \\ &= \frac{S}{v}\frac{\partial}{\partial S}\frac{V_{\max} S}{K_{\mathrm{M}} + S}\,, \\ &= \frac{S}{v}\left[\frac{V_{\max}}{K_{\mathrm{M}} + S} - \frac{V_{\max} S}{(K_{\mathrm{M}} + S)^2}\right]\,, \end{aligned} \tag{6.37}$$

where we use the quotient rule to get the derivative. Then we use the Michaelis-Menten equation (6.36) to substitute the terms inside the brackets by the reaction rate v

$$\varepsilon_{\mathrm{S}} = \frac{S}{v}\left[\frac{v}{S} - \frac{v}{K_{\mathrm{M}} + S}\right]$$

which enables us to cancel the reaction rate v from the left hand side. After a reduction to the common denominator we obtain the final result

$$\varepsilon_{\mathrm{S}} = \frac{K_{\mathrm{M}}}{K_{\mathrm{M}} + S} \tag{6.38}$$

for the substrate elasticity of the Michaelis-Menten equation. Since the Michaelis constant K_{M} is positive, thus is a monotonically decreasing function with respect to the substrate concentration S.

In the limiting case of a small substrate concentration $S \ll K_{\mathrm{M}}$, we can neglect the substrate compared to the Michaelis constant, and the elasticity becomes $\varepsilon_{\mathrm{S}} \approx 1$. A small change in the substrate concentration has a huge effect on the reaction rate. On the other hand, in the limit of a exceedingly substrate concentration $S \gg K_{\mathrm{M}}$, the elasticity is vanishing. A change of the concentration does not affect the reaction rate. The enzyme is completely saturated. It follows from both limits that the substrate elasticity is in the range

$$0 \leq \varepsilon_{\mathrm{S}} \leq 1\,.$$

Besides the substrate concentration, the reaction rate also depends on the Michaelis constant and the limiting rate. Both quantities are assumed to be parameters of the enzyme kinetic reaction. The effect of a change in their values is measured by the π-elasticities which we calculate now. From Eq. (6.10) follows for the elasticity with respect

to the Michaelis constant K_M

$$\begin{aligned}\pi_{K_\mathrm{M}} &= \frac{K_\mathrm{M}}{v}\frac{\partial v}{\partial K_\mathrm{M}} \\ &= \frac{K_\mathrm{M}}{v}\left[-\frac{V_\mathrm{max}\,S}{(K_\mathrm{M}+S)^2}\right].\end{aligned} \tag{6.39}$$

Again we can substitute the rational expression in the brackets with the reaction rate. It follows for the elasticity

$$\pi_{K_\mathrm{M}} = -\frac{K_\mathrm{M}}{K_\mathrm{M}+S} = -\varepsilon_\mathrm{S} \tag{6.40}$$

which is a monotone decreasing function in the range $[-1;0]$. The lower limit is valid in the case $K_\mathrm{M} \gg S$ where a further increase of the Michaelis constant will decelerate the reaction. Additionally, the reaction rate is very sensitive to a change in the constant in that regime. In the limit of $K_\mathrm{M} \ll S$ the elasticity is approximately zero. Here, the saturation of the enzyme is the dominating effect. A change of K_M has no or only a small effect on the reaction rate.

The limiting rate V_max is the second parameter in the Michaelis-Menten equation (6.36). Its elasticity coefficient is

$$\pi_{V_\mathrm{max}} = 1\,. \tag{6.41}$$

The reaction rate is direct proportional to the limiting rate.

6.9.2 Reversible Michaelis-Menten equation

The original Michaelis-Menten scheme (6.35) does not include a reverse reaction in the enzymatic reaction. For a well-founded investigation of biochemical pathways one has to consider such a reaction [HCB97]. Otherwise, the chosen description might be valid only in a restricted region of the parameter and variable space. Often these are very special limiting cases. Therefore we now include a reverse reaction into the enzyme kinetic reaction

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E + P \tag{6.42}$$

which results in the rate equation (see Sec. 2.7.6)

$$v = \frac{V_\mathrm{max1}\,S/K_\mathrm{M1} - V_\mathrm{max2}\,P/K_\mathrm{M2}}{1 + S/K_\mathrm{M1} + P/K_\mathrm{M2}}\,. \tag{6.43}$$

The parameters are now

$$\begin{aligned} K_\mathrm{M1} &= \frac{k_{-1}+k_2}{k_1} \qquad & V_\mathrm{max1} &= k_2\,E^\mathrm{T}\,, \\ K_\mathrm{M2} &= \frac{k_{-1}+k_2}{k_{-2}} \qquad & V_\mathrm{max2} &= k_{-1}\,E^\mathrm{T}\,. \end{aligned}$$

From Eq. (6.9) we obtain for the substrate elasticity

$$\begin{aligned}\varepsilon_{\mathrm{S}} &= \frac{S}{v}\frac{\partial v}{\partial S}\,,\\ &= \frac{S}{v}\left[\frac{V_{\mathrm{max1}}/K_{\mathrm{M1}}}{1+S/K_{\mathrm{M1}}+P/K_{\mathrm{M2}}} - \frac{1}{K_{\mathrm{M1}}}\frac{V_{\mathrm{max1}}\,S/K_{\mathrm{M1}} - V_{\mathrm{max2}}\,P/K_{\mathrm{M2}}}{(1+S/K_{\mathrm{M1}}+P/K_{\mathrm{M2}})^2}\right]\,. \end{aligned} \tag{6.44}$$

We can substitute Eq. (6.43) in the second term and cancel the reaction rate from the prefactor. The first term we multiply with the inverse reaction rate. Then we reduce the resulting expression to the common denominator

$$\varepsilon_{\mathrm{S}} = \frac{S}{K_{\mathrm{M1}}}\left[\frac{V_{\mathrm{max1}}}{V_{\mathrm{max1}}S/K_{\mathrm{M1}} - V_{\mathrm{max2}}P/K_{\mathrm{M2}}} - \frac{1}{1+S/K_{\mathrm{M1}}+P/K_{\mathrm{M2}}}\right]$$

and obtain the substrate elasticity

$$\varepsilon_{\mathrm{S}} = \frac{\left[V_{\mathrm{max1}} + (V_{\mathrm{max1}}+V_{\mathrm{max2}})\,\frac{P}{K_{\mathrm{M2}}}\right]\frac{S}{K_{\mathrm{M1}}}}{\left(\frac{V_{\mathrm{max1}}S}{K_{\mathrm{M1}}} - \frac{V_{\mathrm{max2}}P}{K_{\mathrm{M2}}}\right)\left(1+\frac{S}{K_{\mathrm{M1}}}+\frac{P}{KM2}\right)} \tag{6.45}$$

which is positive if $v > 0$ corresponding to a flux from substrate to product. An increase of substrate concentration will increase the flux and the production rate of the product, respectively.

In contrast to the irreversible Michaelis-Menten equation (6.36) the reaction rate depends now also on the product. The corresponding product elasticity is

$$\begin{aligned}\varepsilon_{\mathrm{P}} &= \frac{P}{v}\frac{\partial v}{\partial P}\,,\\ &= \frac{P}{v}\left[\frac{V_{\mathrm{max2}}/K_{\mathrm{M2}}}{1+S/K_{\mathrm{M1}}+P/K_{\mathrm{M2}}} - \frac{1}{K_{\mathrm{M2}}}\frac{V_{\mathrm{max1}}\,S/K_{\mathrm{M1}} - V_{\mathrm{max2}}\,P/K_{\mathrm{M2}}}{(1+S/K_{\mathrm{M1}}+P/K_{\mathrm{M2}})^2}\right]\,,\\ &= \frac{P}{K_{\mathrm{M2}}}\left[\frac{V_{\mathrm{max2}}}{V_{\mathrm{max1}}S/K_{\mathrm{M1}} - V_{\mathrm{max2}}P/K_{\mathrm{M2}}} - \frac{1}{1+S/K_{\mathrm{M1}}+P/K_{\mathrm{M2}}}\right]\,,\\ &= \frac{\left[V_{\mathrm{max2}} + (V_{\mathrm{max1}}+V_{\mathrm{max2}})\,\frac{S}{K_{\mathrm{M1}}}\right]\frac{P}{K_{\mathrm{M2}}}}{\left(\frac{V_{\mathrm{max1}}S}{K_{\mathrm{M1}}} - \frac{V_{\mathrm{max2}}P}{K_{\mathrm{M2}}}\right)\left(1+\frac{S}{K_{\mathrm{M1}}}+\frac{P}{KM2}\right)}\,, \end{aligned} \tag{6.46}$$

which one obtains in the same way as described for the substrate elasticity above. The product elasticity is negative for a reaction rate $v > 0$. Hence, an increase of the product will decrease the reaction rate.

6.9.3 Hill equation

Many enzymes show a much more complicated behavior than considered in the derivation of the Michaelis-Menten equation. The corresponding reaction rates can be described by the phenomenological Hill equation

$$v = \frac{V_{\mathrm{max}}(S/K)^n}{1+(S/K)^n}\,, \tag{6.47}$$

where $V_{\max}$ is a limiting rate, K an enzyme specific constant, n the Hill coefficient and S the substrate concentration. The substrate elasticity for the Hill equation is

$$\begin{aligned} \varepsilon &= \frac{S}{v}\frac{\partial v}{\partial S} \\ &= \left[\frac{V_{\max}(S/K)^{n-1}}{1+(S/K)^n} - \frac{V_{\max}(S/K)^n(S/K)^{n-1}}{[1+(S/K)^n]^2}\right]\frac{S}{K}\frac{n}{v}\,, \end{aligned} \tag{6.48}$$

which we can transform with Eq. (6.47) into the expression

$$\varepsilon_{\mathrm{S}} = \left[v - v\frac{(S/K)^n}{1+(S/K)^n}\right]\frac{n}{v}$$

leading to the final result

$$\varepsilon_{\mathrm{S}} = \frac{n}{1+(S/K)^n}\,. \tag{6.49}$$

The substrate elasticity is a decreasing function with respect to the substrate concentration.

In the limit of small concentrations $S \ll K$ we neglect the substrate dependent term in the denominator and obtain the asymptotic elasticity

$$\varepsilon_{\mathrm{S}} \approx n\,.$$

Again, the reaction rate is very sensitive to changes in the substrate concentration. The sensitivity increases if the Hill coefficient increases. In the limit of high substrate concentrations $S \gg K$ the elasticity becomes

$$\varepsilon_S \approx \frac{n}{(S/K)^n}$$

which simplifies further if we assume $(S/K)^n \gg n$ to

$$\varepsilon_{\mathrm{S}} \approx 0\,.$$

The Hill equation is insensitive to changes in the substrate concentration. Here, the enzyme saturation is dominant and determines the reaction rate.

Similar to the substrate elasticity we obtain the π-elasticity with respect to the enzyme-specific constant K

$$\begin{aligned} \pi_K &= \frac{K}{v}\frac{\partial v}{\partial K} \\ &= -\frac{n}{1+(S/K)^n} = -\varepsilon_{\mathrm{S}}\,. \end{aligned} \tag{6.50}$$

In analogy to the treatment of the Michaelis-Menten equation the reaction rate decreases with increasing constant K. Since the limiting rate $V_{\max}$ acts only as constant factor in the Hill equation (6.47), its π-elasticity is again

$$\pi_{V_{\max}} = 1\,.$$

An increase in the limiting rate will increase the reaction rate in the same manner.

6.9.4 S-Systems

S-Systems [Sav69b] can be used to approximate the reaction rate near the steady state. Within this approach we aggregate all contributions to the formation of X_i in one single term and all contributions to the degradation of X_i in a second one. The rate law for the j-th reaction is now

$$v_j = \underbrace{k_{+j} \prod_i X_i^{n_{ij}^-}}_{\text{production}} - \underbrace{k_{-j} \prod_i X_i^{n_{ij}^+}}_{\text{degradation}} \tag{6.51}$$

with the forward rate constant k_{+j} and the reverse rate constant k_{-j} and the stoichiometric coefficients n_{ij}^- for reactants and n_{ij}^+ for products. In contrast to the kinetic orders of elementary reactions introduced in Section 2.5, the kinetic orders $n_{ij}^{\pm}$ can have real values. Note that a stoichiometric coefficient of zero has two possible interpretations: i.) the metabolite does not take part at the considered reaction and ii.) the rate with respect to the considered species is limited by physical processes (see also Section 2.5.5.1). The steady state is then determined by a generalised law of mass action:

$$K_{\text{eq}} = \frac{k_{+j}}{k_{-j}} = \frac{\prod_i X_i^{n_{ij}^+}}{\prod_i X_i^{n_{ij}^-}} \,. \tag{6.52}$$

The elasticity for the j-th reaction and with respect to the metabolite X_i is in the S-Systems approach

$$\begin{aligned} \varepsilon_i^j &= \frac{X_i}{v_j} \frac{\partial v_j}{\partial X_i} \\ &= \frac{X_i}{v_j} \left[k_{+j} \frac{\partial}{\partial X_i} \prod_l X_l^{n_{lj}^-} - k_{-j} \frac{\partial}{\partial X_i} \prod_l X_l^{n_{lj}^+} \right] . \end{aligned}$$

The partial derivatives of products in the above equation is

$$\frac{\partial}{\partial X_i} \prod_l X_l^{n_{lj}^{\pm}} = n_{ij}^{\pm} X_i^{n_{ij}^{\pm}-1} \prod_{l \neq i} X_l^{n_{lj}^{\pm}} \,,$$

where the metabolite X_i is excluded from the product. Furthermore, its kinetic order is reduced by one. But a multiplication with the normalisation factor restores the original order, and we can group again all participating metabolites in a common product. If we apply the definition of the reaction in the S-System approach (6.51), the elasticity becomes

$$\varepsilon_i^j = \frac{k_{+j}\, n_{ij}^- \prod_l X_l^{n_{lj}^-} - k_{-j}\, n_{ij}^+ \prod_l X_l^{n_{lj}^+}}{k_{+j} \prod_l X_l^{n_{lj}^-} - k_{-j} \prod_l X_l^{n_{lj}^+}} \,. \tag{6.53}$$

The π-elasticities of the rate constants are easy to derive. We obtain an elasticity of

$$\pi_{k_{\pm i}}^j = \pm \delta_{ij} \frac{k_{\pm i} \prod_l X_l^{n_{lj}^{\mp}}}{k_{+j} \prod_l X_l^{n_{lj}^-} - k_{-j} \prod_l X_l^{n_{lj}^+}} \,, \tag{6.54}$$

where

$$\delta_{ij} = \begin{cases} 0 & \text{for } i \neq j \\ 1 & \text{for } i = j \end{cases}$$

is the Kronecker-Delta [AS72]. The π-elasticity is non-vanishing only if the rate constant determines the considered reaction. Otherwise, the rate does not depend on the investigated rate constant and the elasticity is zero. If $v_j > 0$, then the π-elasticity with respect to the forward rate constant is always positive. Its increase further increases the reaction rate. In contrast, the π-elasticity of the reverse rate constant is negative, corresponding to a decreasing reaction rate with increasing rate constant.

6.9.5 Elasticity calculus

As demonstrated in the previous sections the derivation of elasticity coefficients becomes more and more complicated if the complexity of the considered reactions increases. Due to the resulting partial derivatives the expressions are very complex. Formally, due to the normalisation often a simplification arises because of cancelations between numerator and denominator terms with the normalisation factor. Hence, one can use an elasticity calculus [Fel97] to avoid the generation of terms that will subsequently cancel.

Often the reaction rate is a rational form of

$$v = \frac{N}{D} \tag{6.55}$$

where both function N and D depend on the metabolite concentration X_i. The corresponding elasticity is

$$\begin{aligned} \varepsilon_i &= \frac{X_i}{v}\frac{\partial v}{\partial X_i} \\ &= X_i \frac{D}{N}\left[\frac{\partial N/\partial X_i}{D} - N\frac{\partial D/\partial X_i}{D^2}\right] \end{aligned} \tag{6.56}$$

which simplifies if both terms in the brackets are multiplied with the inverse reaction rate of the normalisation factor. Finally we obtain with

$$\varepsilon_i = X_i\left[\frac{1}{N}\frac{\partial N}{\partial X_i} - \frac{1}{D}\frac{\partial D}{\partial X_i}\right] \tag{6.57}$$

a simple rule for the calculation of the elasticity coefficients. Nevertheless, we must emphasise, that the introduced calculus is applicable only if the reaction rate has a rational form as proposed in Eq. (6.55). Otherwise, one has to use the general definition (6.9). In this case a step by step derivation, as shown on the examples in the previous section, cannot be avoided.

7 Inferring Interactions in networks

It is important to note at the beginning of this chapter that in other chapters we used capital letters for matrices and small letters for elements. This would not be possible in this chapter and we will therefore introduce bold font to denote matrices.

Motivation

Consider the example of a four-gene network model depicted in Figure 7.1(a). Note that genes do not actually interact directly with each other (neither do the corresponding mRNAs); instead, gene induction or expression occurs through the action of specific proteins, metabolites, or complexes. By abstracting these actions, we represent the interaction of genes in a *gene network* and refer to the interactions as *functional interactions*. The functional interactions in a gene network are also called *wiring*. Any metabolic pathway, signal transduction pathway, or gene interaction circuit can be modelled as a network with nodes corresponding to metabolites, proteins, or genes, and links between nodes corresponding to the interactions of nodes. These interactions proceed through a number of protein products (for example, transcription factors) and metabolic intermediates.

gene network
functional interactions
wiring

The functional interactions between the nodes of an interaction network can be expressed at different levels. Figure 7.1(a) shows only the topology of connections, where a link represents either some association between two genes, or correlation between their expression profiles. In Figure 7.1(b), the concept of direction is added to each interaction. Direction implies some causality in a relation. Causality can be direct or through a chain of relations. For example, gene 1 directly affects the expression of gene 3, but not vice versa. On the other hand, gene 3 affects gene 1 indirectly through a chain of connections which can be shown as $3 \rightarrow 4 \rightarrow 2 \rightarrow 1$. Another possible chain of connections through which gene 3 indirectly affects gene 1 would be $3 \rightarrow 4 \rightarrow 1$. It is however impossible to answer the question, how much gene 3 affects gene 1 with the information inferred from Figure 7.1(b). To answer such a question, more information should be added to the interaction network.

The effect of one gene on another one can be stimulatory or inhibitory. A stimulatory connection directly increases the transcription rate of the target gene. Such a connection is shown with a positive sign on the corresponding connection in Figure 7.1(c). An inhibitory connection in contrast, directly decreases the transcription rate of the target gene. Such a connection is shown with a negative sign on the corresponding connection in Figure 7.1(c). Figure 7.1(d) shows another way to illustrate Figure 7.1(c). In Figure 7.1(e) the strength of each interaction which is the strength of the direct stimulatory or inhibitory affect of one node on another one is added to the interaction network. In the

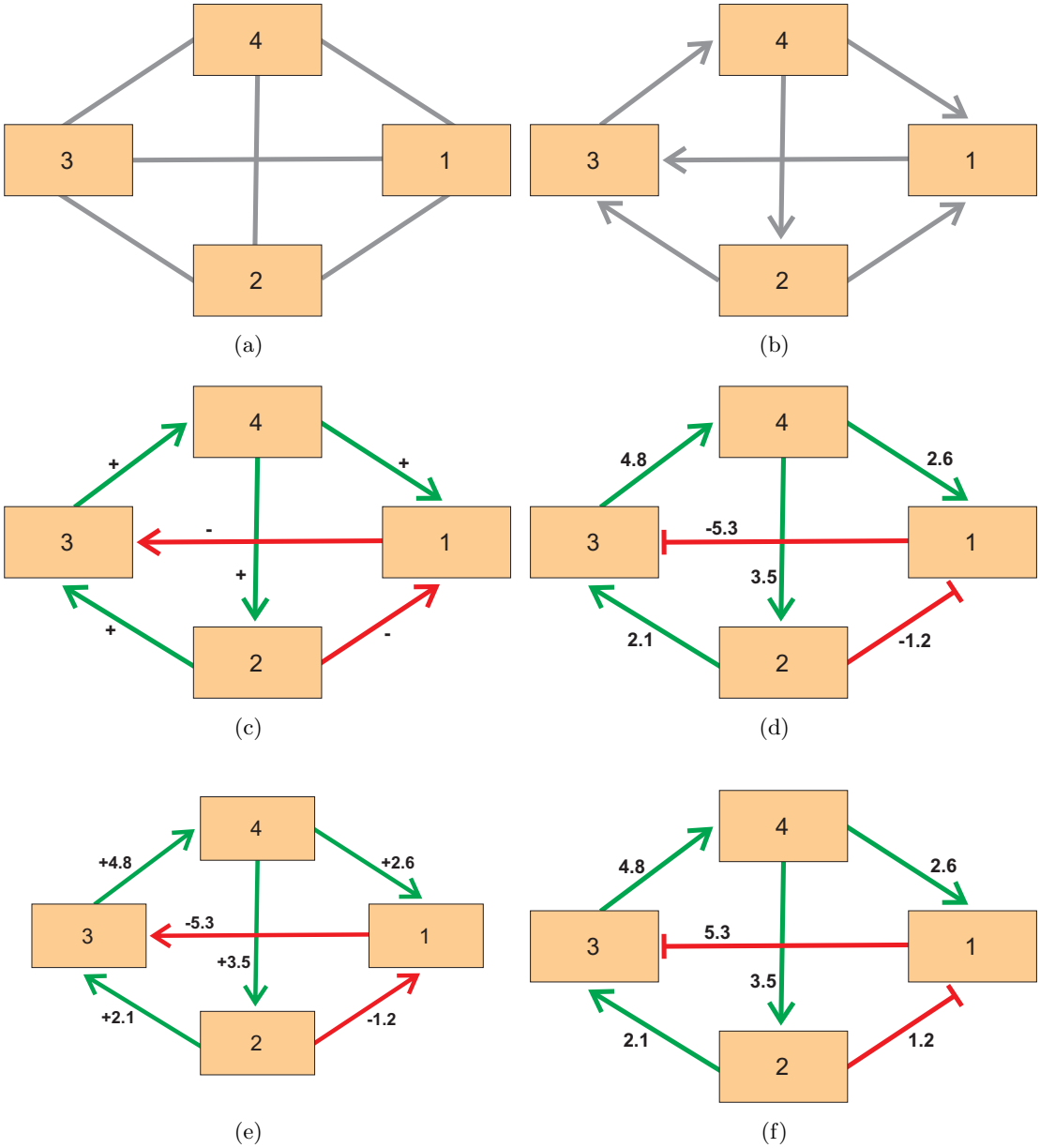


Figure 7.1: Different descriptions of a four-gene transaction network.

former case, the strength can be shown with a positive number; whereas, in the latter case with a negative number. In Figure 7.1(e), the absolute values of the connection strengths can be regarded as a measure of comparison between connection strengths. Figure 7.1(f) shows another way to illustrate Figure 7.1(e).

It is important to find out in which level of description the considered gene or signalling network can be accessible. Different description levels of a gene or signalling network require different types of mathematical model, different algorithms for constructing the network from stationary or time series data, different methods of analysis of their behaviour, and yield different types of biological information [K+02].

If the variables $x_1, ..., x_4$ represent the expression levels of genes 1 to 4 respectively, one can describe the dynamical behaviour of the network shown in Figure 7.1(e) by a set of differential equations. In each equation, the rate of change of the transcription level of one gene is described as a function of the expression levels of other genes as follows.

$$\begin{aligned} \frac{d}{dt}x_1 &= f_1(x_1, x_2, x_3, x_4) \\ \frac{d}{dt}x_2 &= f_2(x_1, x_2, x_3, x_4) \\ \frac{d}{dt}x_3 &= f_3(x_1, x_2, x_3, x_4) \\ \frac{d}{dt}x_4 &= f_4(x_1, x_2, x_3, x_4) \end{aligned} \tag{7.1}$$

In most cases, these functions are known but likely to be non-linear. The fact that the rate of transcription of a gene would saturate after reaching a maximum, justifies the likelihood of f-functions being non linear. It is also typical that the transcription rate of a gene would only depend on the expression level of some of other genes in the network. Thus, for instance, in Figures 7.1(e), gene 1 depends directly on gene 2 and 4; gene 2 depends directly only on gene 4; gene 3 depends directly on gene 1 and gene 2; gene 4 depends directly only on 3. Hence, the above equations simplify to the following equations.

$$\begin{aligned} \frac{d}{dt}x_1 &= f_1(x_2, x_4) \\ \frac{d}{dt}x_2 &= f_2(x_4) \\ \frac{d}{dt}x_3 &= f_3(x_1, x_2) \\ \frac{d}{dt}x_4 &= f_4(x_3) \end{aligned}$$

Generally, a node of an interaction network can be a single protein, or a group of proteins, a gene (as in the interaction network of the example shown in Figure 7.1), a group of genes, enzymes or other cellular components of arbitrary complexity. These nodes are connected by physical and/or chemical interactions which together perform one or more identifiable tasks. A node of the interaction network is also called a *module*. module

According to the type of the network nodes, an interaction network can be classified as signalling, gene or metabolic network. For example, in gene networks, modules can involve mRNAs of a particular gene or gene cluster with regulatory interaction loops running through metabolic and signalling pathways.

The basic concept of quantifying the strengths of interactions between nodes of an interaction network is to analyse the direct effect of a small change in one node on another node, while keeping the remaining nodes unchanged to prevent the spread of the perturbation. For example, continuing with the example of the gene network shown in Figure 7.1, we ask how much the transcription level of gene 1 changes in response to a change in the transcription level of gene 4, while the transcription level of all other gene are kept constant. This can be shown by the partial derivatives of f_1 with respect to x_4, written as:

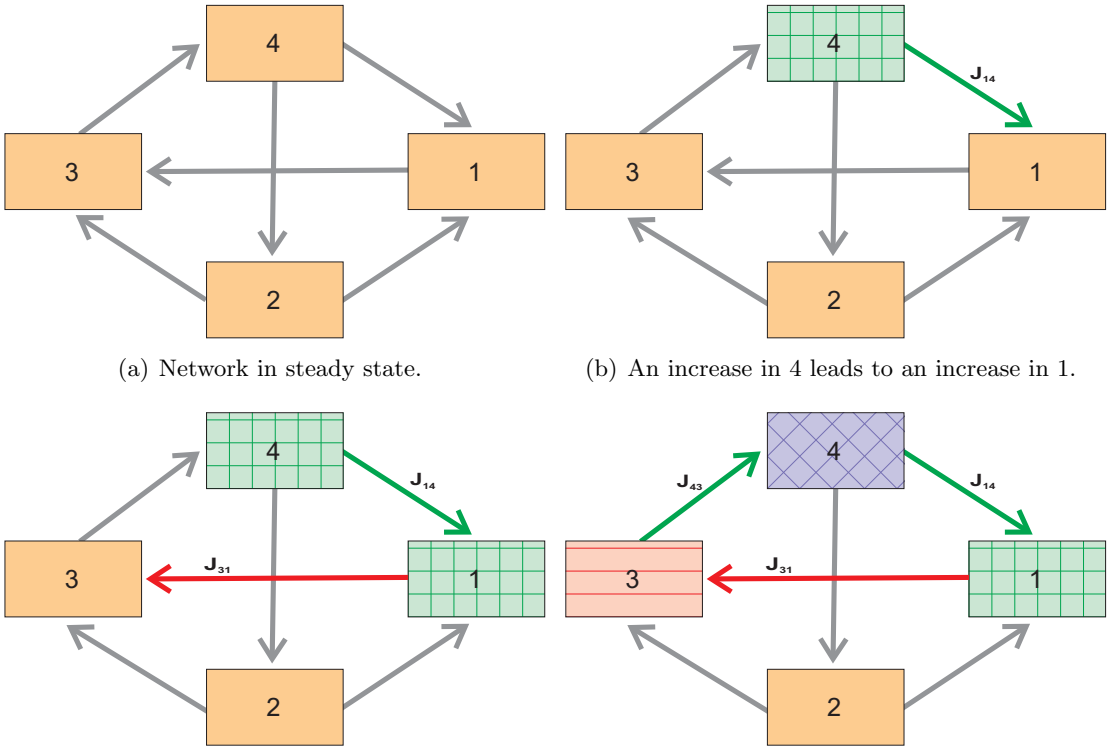
$$J_{14} = \frac{\partial f_1}{\partial x_4}$$

These quantities can conveniently be arranged in a matrix $\mathbf{J}$, known in as the Jacobian matrix. The Jacobian matrix is here of particular interest. Each off-diagonal element of the Jacobian matrix describes the change of one node, relative to another. Positive and negative entries correspond to activation, respectively inhibition of one node by another. The definition of J_{14} is illustrated in Figure 7.2(b).

In Figure 7.2, the propagation effect of a change in the level of one gene through the whole interaction network is partly shown. In Figure 7.2(a), the network is assumed to be in a reference steady state. Therefore, there are no observable changes in the transcription levels of the genes depicted as the nodes of the network. In this case, right hand sides of the differential equations shown in equation (7.1) equal zero. As in Figure 7.2(b) is shown, a perturbation increasing the level of gene 4 leads to an increase in the transcription level of gene 1. This in turn results in a decrease in the transcription level of gene 3 shown in Figure 7.2(c). The change in the level of gene 3 affects the transcription level of gene 4 as highlighted in Figure 7.2(d). It is important to note that the propagation does not stop at this point. The propagation of the changes spread through the whole interaction network until all the genes have found a new steady state level.

Another point to consider is the possibility of having different paths for the propagation of changes in an interaction network. This fact is shown in Figure 7.3. The network is assumed to be in a reference steady state in Figure 7.3(a). A perturbation then increases the level of gene 4, which has the effect of increasing the rate of transcription of 2 as illustrated in Figure 7.3(b). The sequence is continued in Figure 7.3(c) to highlight the stimulating effect of an increase in the level of 2 on the transcription level of 3. This in turn leads to an increase in 4 as shown in Figure 7.3(d). Again, the changes propagate throughout the whole interaction network until a new steady state is established.

As once mentioned, the amount of change in the transcription level of gene 1 as a result of a change in the level of gene 4 is quantified by the partial derivative J_{14} shown in Figure 7.2(b). This quantifying is done in the context of a process called linearisation. Linearisation of a non-linear system around a critical point requires infinitesimally small perturbations around the reference state, so that in practise we replace J_{14} by a finite



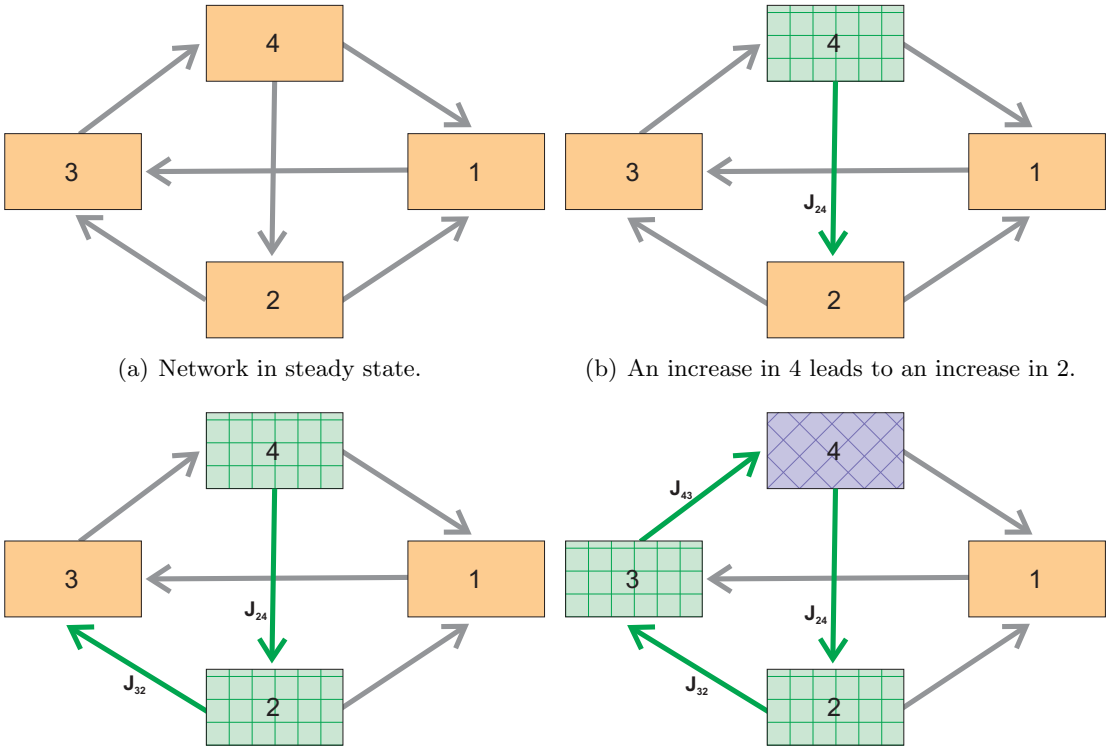
(a) Network in steady state.

(b) An increase in 4 leads to an increase in 1.

(c) An increase in 1 then leads to a decrease in 3.

(d) The change in 3 leads to a change in 4.

Figure 7.2: Propagating effect of an increase in the level of gene 4 through the whole interaction network is shown partly.



(a) Network in steady state.

(b) An increase in 4 leads to an increase in 2.

(c) An increase in 2 then leads to a decrease in 3.

(d) The change in 3 leads to a change in 4.

Figure 7.3: Another path for the partial propagation of an increase in the level of gene 4 through the whole interaction network.

difference approximation. If this requirement is encountered, the system of differential equations of 7.1 can be replaced by a system defined by the Js using a first order Taylor series approximation of the functions $f_1, ..., f_4$.

Functions $f_1, ..., f_4$ in equation 7.1 determine the dynamics of the interaction network as well as its architecture. Therefore, to discover the architecture of an interaction network, the f functions should be estimated from experimental data. Using the notion of linearisation, this problem deforms to the problem of estimating the direct influence (or local response) of the variation of each node on all other nodes in a network. Having this information, one can identify the local network structure and from this refer the global network structure. In other words, the identification of the functional interaction structure of a biomolecular network can be considered as unraveling all of the corresponding local interaction networks.

Global interactions in a biomolecular network can be considered as a sum of local interactions of individual nodes. This is illustrated in Figure 7.4. Note that an arrow indicates a positive direct influence. On the other hand, a line with the end of a vertical segment shows a negative direct influence and a line with the end of a black circle represents no direct influence. A positive influence is referred also as an *up-regulation* influence whereas a negative influence can be referred as a *down-regulation* influence.

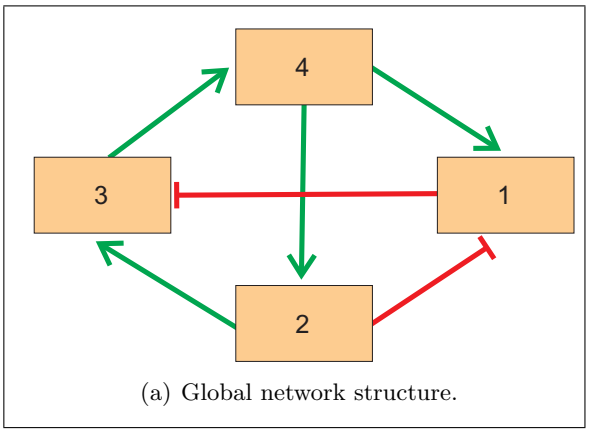
up-regulation

down-regulation

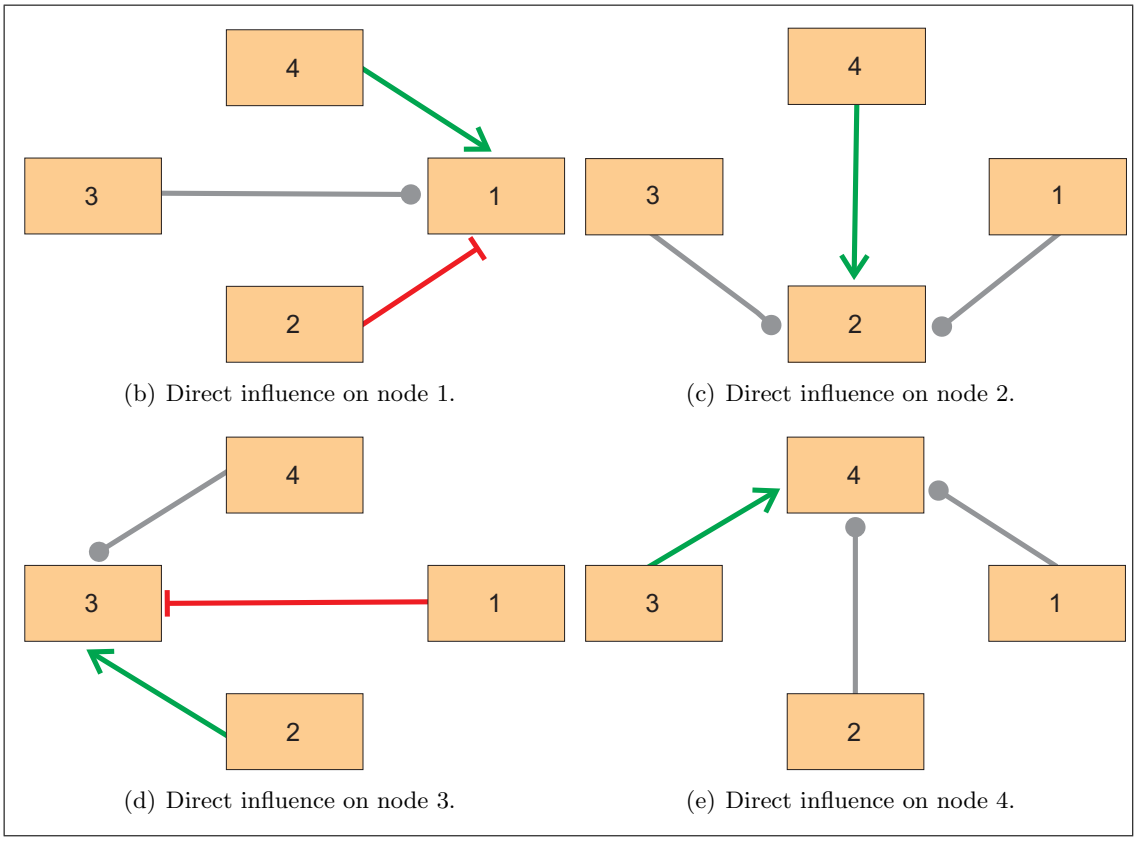
Fortunately, advances in genomics and proteomics analysis facilitate the monitoring of the expression levels of large gene sets and the activity states of signalling proteins in living cells. However, inferring and quantifying interconnections within complex signalling and gene networks is still considered a fundamental problem in cell biology. It is difficult to directly measure the local responses (for example, estimating J_{14}) from experiments. As it is highlighted in Figure 7.2 and Figure 7.3 , the variation of one node promptly spreads over the other nodes of the network till the whole network reaches a new steady state. Carrying out practical experiments in order to quantify the change in the transcription level of one gene, owing to a change in another gene while keeping everything else constant is practically impossible. Instead, the global change in the transcription level of a gene owing to some perturbation to the system can be measured. Thus, a method which can be used to indirectly infer the local responses based on information from global changes measured in a practical experiment is required.

Various qualitative and mechanistic modelling methods have been used to infer the structure of bio-molecular networks. These include Boolean networks, genetic algorithms, dynamic simulations, Bayesian models (to infer the gene circuitry) [LFS98], and metabolic control analysis (to determine interactions between metabolites and enzymes) [LD93, Bra96, VH02]. These mechanistic *bottom-up approaches* have the advantage of being readily testable against experiments as a computer replica of cellular networks and providing a wealth of detail. However, a major disadvantage of the mechanistic modelling is the necessity of modelling a large number of molecular processes , complicated by the fact that the values of multiple kinetic parameters may be unknown. Moreover, a button-up approach inevitably misses the interactions and regulatory feedbacks which are still not discovered. Therefore, a button-up approach would be less applicable to poorly characterised networks [dlFBM02, K$^+$02, SKK04]. In the reverse engineering of a gene network, detailed information about the network is in general deficient when

bottom-up approaches



(a) Global network structure.



(b) Direct influence on node 1.

(c) Direct influence on node 2.

(d) Direct influence on node 3.

(e) Direct influence on node 4.

Figure 7.4: Global interactions of a biomolecular network (the upper box) can be seen as a sum of local interactions of individual nodes (the lower box).

compared to the case of a metabolic network. In this context, the metabolic control analysis (MCA) techniques have been extended by de la Fuente in [dlFBM02] to infer a gene network structure with the assumption that each node of the network is a single gene and the parameters corresponding to gene transcription rates can be perturbed. Recently, however two important approaches to this problem have been proposed by Kholodenko *et al.* [K+02] and Sontag *et al.* [SKK04]. These two approaches can be classified as *top-down approaches* of studying interaction networks.

top-down approaches

The bottom-up approaches rely on the numerical simulation of a mechanistic model of an interaction network and quantitatively compare the results of such simulations to experimental observations [HMIC01, KDMH99, MMD+02, SHY+02, SBB98, Tys01, vDMMO00]. In contrast, top-down approaches infer the architecture of the interaction network based on observed global responses of the system to a series of signals or experimental perturbations. Hormones, growth factors, neurotransmitters, or experimental interventions such as chemical inhibition are some examples of perturbation to a system. The global responses that could be measured include changes in the phosphorylation states or activities of proteins, mRNA levels, or transcription rates. From these global responses, top-down analysis methods attempt to recover the local interactions between interaction network nodes, which in turn form a map of the network. Inferring the architecture of an interaction network is also called reverse engineering of an interaction network [AKLS05]. Contrary to the bottom-up approaches which are less applicable to poorly characterised networks, top-down approaches are more appropriate for network discovery in the cases where the bottom-up approaches by nature miss interactions and regulatory feedbacks that still await discovery [Bra96, dlFBM02, K+02]. Top-down approaches have also the great advantage of being applicable to regulatory networks of arbitrary complexity.

Top-Down Methods

Definitions

The concept of a modular framework [HHLM99, Lau00] was considered in top-down approaches as a way to facilitate the process of understanding the coordinated behaviour of numerous molecular interactions in an interaction network and inferring its architecture. The extensive use of the concept of modules gives the top-down approaches the advantage of being applicable to networks of arbitrary complexity. This way, instead of considering numerous molecular reactions happening inside an interaction network, we divide it into m reaction groups referred to as functional units or modules. The concept of a module was introduced in the last section. As a remark, a module can be a group of genes, enzymes, or other cellular components of arbitrary complexity, connected by physical and/or chemical interactions which together perform one or more identifiable tasks. Generally, a module involves many cellular components (intermediates) connected by chemical reactions. However, the task(s) which a module performs usually depends on one or a small number of components of the module. Each of these components is called

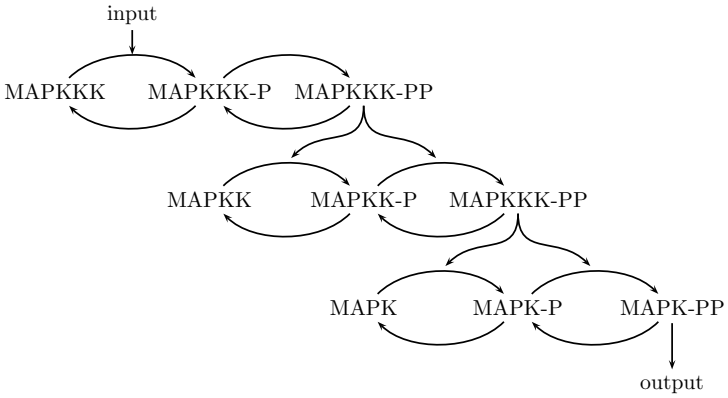


Figure 7.5: General outline of the MAPK pathway.

communicating intermediate

intermodular interactions

a *communicating intermediate* . A communicating molecule may be the active form of a kinase, mRNA, or transcription factor influencing other modules. Thus, communicating intermediates form molecular connections between modules, referred to as *intermodular interactions* . The top-down approaches black-box the molecular organisation of network modules. In order to untangle and quantify the web of intermodular interactions, only the concentration or chemical states of the communicating intermediates (the module outputs) is monitored. The probably complex internal structure of the module plays no role in the analysis of the interaction network. This is considered the big advantage of using top-down models to find network interaction maps.

In [K+02, AKLS05], the mitogen-activated protein kinase (MAPK) signalling network shown in Figure 7.5 is considered as an example to show the usage of the module concept in simplifying an interaction network. Each of the three tires of the mitogen-activated protein kinase (MAPK) cascade can be considered as a functional module that involves unphosphorylated, mono-phosphorylated, and biphosphorylated forms of a protein kinase and the reactions converting these forms. Note that modules need not be rigid, and entire MAPK cascades can serve as functional modules in a signalling network that involves growth factor and stress-activated pathways. However, if we consider a tire as a module, we see that only the doubly phosphorylated forms of the kinases have any interactions with components on other tires/levels of the network. The unphosphorylated and singly phosphorylated forms only interact with components on the same tire/level. If we black-box the tires and consider them as modules, the dimensionality of the problem reduces by a factor of 3. This is due to the fact that although there are a total of nine components in this prototypical network, we can treat it from a modular point of view as consisting of three modules. The modules correspond to the MKKK, MKK, and MAPK levels and are represented by the communicating intermediates MKKK-PP, MKK-PP, and MAPK-PP, respectively. The order of simplification is considerably greater in case of the networks having more complex internal structure in each module [AKLS05]. Another point to consider in Figure 7.5 is that not all the connections between modules are shown. Especially, no repression is shown in this figure. Actually, modules of an interaction network can be interconnected in multiple way,many

of which may be unknown. The aim in the present chapter is to represent some of the methodologies capable of unraveling and quantifying unknown modular connections in signalling and gene networks.

Consider that the interaction network is divided into m reaction groups referred to as functional units or modules and let $x_i, i = 1, ..., m$ be the activities or concentrations of communicating intermediates. Then, the intermodular interactins can be quantified in terms of the fractional changes $(\frac{\Delta x_i}{x_i})$ in the activity of communicating intermediates (x_i) of a particular module (i) brought about by a change in the (output) activity (x_j) of another module (j). Output activities of all other modules $(x_k, k \neq i, k \neq j)$ are assumed to remain fixed, whereas the affected module (i) is allowed to relax to its steady state. A mathematical definition requires the changes $(\frac{\Delta x}{x})$ to be infinitesimally small, resulting in log-to-log derivatives,

$$r_{ij} = \lim_{\Delta x_j \to 0} \left(\frac{\frac{\Delta x_i}{x_i}}{\frac{\Delta x_j}{x_j}} \right) = \left(\frac{\partial \ln x_i}{\partial \ln x_j} \right)_{\text{module } i \text{ in steady state}}, i \neq j \tag{7.2}$$

Where r_{ij} is referred to as *local response* or *local coefficient* which quantifies the sensitivity of module i to module j. The term "local" indicates that the response results from immediate interactions between two modules when all other network interactions are held constant. A response coefficient r_{ij} less than 1 means that (small) fractional changes in module j output are attenuated in module i, whereas a response greater than 1 means that these fractional changes are amplified by the factor r_{ij}. A response coefficient of 0 means that module j has no direct effect on module i, whereas a negative response coefficient means inhibition.

local response
local coefficient

If each module is assumed to have a single communicating intermediate, all interactions between network modules are quantified by $m \cdot (m-1)$ intermodular response coefficients r_{ij}. These "connection" coefficients indicate how the network is "wired" and compose the $m \times m$ matrix, $\mathbf{r}$, hereafter referred to as the *network interaction map*. The ith row of the matrix $\mathbf{r}$ quantifies how module i is affected by each network module through intermediate interaction, whereas the jth column of $\mathbf{r}$ measures how module j directly influences each network module. The diagonal elements (r_{ii}) of the matrix $\mathbf{r}$ are assigned to be -1, i.e, $r_{ii} = -1, i = 1, ..., m$.

network interaction map

If we consider the module i "in isolation" from the network, the local response coefficient r_{ip_i} of x_i to a perturbation of parameter p_i, intrinsic to module i can be determined as follows,

$$r_{ip_i} = \left(\frac{\partial \ln x_i}{\partial p_i} \right)_{\text{module } i \text{ in steady state}}$$

It is clear that when module i is isolated from the network, changes in parameters p_j, influencing other module j, have no effect on module i, and therefore the local response of x_i to a perturbation in p_j equals zero. Local responses to perturbations, affecting single modules only, form the diagonal $m \times m$ matrix, $diag\mathbf{r}_P$, with diagonal elements r_{ip_i} and all off-diagonal elements equal to zero.

If following a parameter p_i perturbation intrinsic to module i an entire network is allowed to relax to a steady state, this perturbation not only causes changes in those

modules directly affected by module i, but also propagates further into the network through interactions between other modules. The resulting time-dependent or stationary responses are called "global" responses of the network. We designate by R_{jp_i} the global response coefficient of module j to a perturbation in p_i and by $\mathbf{R}_P$ the $m \times m$ matrix composed of these coefficients,

$$R_{jp_i} = \left(\frac{d \ln x_j}{dp_i} \right)_{\text{system in steady state}} \qquad j, i = 1, ..., m. \tag{7.3}$$

Understanding the difference between the local and global interaction network responses to perturbations is of great importance. The difference between the local diag$\mathbf{r}_P$ and global $\mathbf{R}_P$ response matrices is that only module i is allowed to reach the steady state to determine r_{ip_j}, whereas an entire network is permitted to relax to its steady state to measure R_{ip_j}.

Global responses to perturbations can be measured in experiments with intact cellular systems. However, local responses governed by the network interaction map can not be captured using intact cells. To measure the kinetics of local interactions between two modules (proteins) directly, they should be isolated from the network. Sometimes the interaction of interest can be reconstituted *in vitro*, but often only an entire system is accessible experimentally. The question of how to determine quantitatively the network interaction map if only the global responses can be assessed is addressed in the rest of this chapter. It is showed that by making parameter perturbations to all modules and measuring the global network responses, the unknown interaction map can be retrieved.

7.1 Method of Kholodenko et al.

7.1.1 Mathematical Derivation

In Appendix 2 published as supporting information to [K+02] on the PNAS web site, an abstract mathematical derivation of the method presented in [K+02] is provided.

Let us assume that the modules have just one communicating intermediate as the output, i.e, the considered perturbations affect only single modules. Consider a dynamic system represented as,

$$\frac{dx}{dt} = f(x, p) \tag{7.4}$$

where the vector of variables $x = (x_1, ..., x_n)$ and the vector of parameters $p = (p_1, ..., p_n)$ belong to open subsets of the Euclidean space. It is assumed that the system has a stable steady state (x_0, p_0) and $f(x_0, p_0) = 0$. Then, where the Jacobian matrix $\mathbf{J} = \frac{\partial f}{\partial x}$ is non-singular, there exists a unique vector $x(p)$ solving $f(x, p) = 0$ in some neighbourhood of a particular p_0 . The objective is then to determine the Jacobian matrix $\mathbf{J}$, assuming that one can determine the global response matrix $\mathbf{R}_P = \frac{\partial x}{\partial p}$.

Unfortunately, such an objective is impossible to achieve. The reason is that the equation $f(x, p) = 0$ is equivalent to the equation $2f(x, p) = 0$, and thus there will be no way to distinguish between $\left(\frac{\partial f}{\partial x}\right)$ and $2\left(\frac{\partial f}{\partial x}\right)$. Thus, the objective should be

restated as finding the matrix $\mathbf{r}$ of r_{ij}. The coefficients r_{ij} correspond to the elements J, "normalised" by the diagonal elements $\left(\frac{\partial f_i}{\partial x_i}\right)$, i.e.,

$$r_{ij} = \frac{\partial x_i}{\partial x_j} = -\frac{\left(\frac{\partial f_i}{\partial x_j}\right)}{\left(\frac{\partial f_i}{\partial x_i}\right)}$$

In matrix notation,

$$\mathbf{r} = -\left(diag\mathbf{J}\right)^{-1} \cdot \mathbf{J} \tag{7.5}$$

Solving $f(x,p) = 0$ in some neighbourhood of a particular p_0 allows us to relate the global response matrix $\mathbf{R}_P$, the Jacobian matrix $\mathbf{J}$, and the matrix of the partial derivatives of functions f with respect to the vector of parameters p,

$$\mathbf{R}_P = \frac{\partial x}{\partial p} = -\left(\frac{\partial f}{\partial x}\right)^{-1} \cdot \left(\frac{\partial f}{\partial p}\right) = -\left(\mathbf{J}\right)^{-1} \cdot \left(\frac{\partial f}{\partial p}\right) \tag{7.6}$$

It is assumed that the matrix $\left(\frac{\partial f}{\partial p}\right)$ is nonsingular in the vicinity of the state (x_0, p_0). Because the perturbation of each of the parameters, namely (p_i) affects only a single module i, the matrix $\left(\frac{\partial f}{\partial p}\right)$ turns out to be diagonal,

$$\frac{\partial f}{\partial p} = diag\mathbf{f}_P$$

The matrix $diag\mathbf{f}_P$ is related to the local response matrix, $diag\mathbf{r}_P$ according to the following relation,

$$diag\mathbf{r}_P = -\left(diag\mathbf{J}\right)^{-1} \cdot \left(diag\mathbf{f}_P\right) \tag{7.7}$$

Equation (7.7) is obtained by differentiation of the equation $f_i(x,p) = 0$ with respect to p_i assuming that all other variables except x_i remain fixed. Using (7.5), (7.6), (7.7) , we find,

$$\mathbf{R}_P = -\left(\mathbf{J}\right)^{-1} \cdot \left(diag\mathbf{f}_P\right) = -\left(\mathbf{J}\right)^{-1} \cdot \left(diag\mathbf{J}\right) \cdot \left(diag\mathbf{J}\right)^{-1} \cdot diag\mathbf{f}_P = -\mathbf{r}^{-1} \cdot diag\mathbf{r}_P \tag{7.8}$$

From equation (7.8) , the matrix $\mathbf{r}$ is expressed as follows,

$$\mathbf{r} = -diag\mathbf{r}_P \cdot {\mathbf{R}_P}^{-1} \tag{7.9}$$

Because all the diagonal elements of the matrix $\mathbf{r}$ are equal to -1, it can be written,

$$\mathbf{I} = diag\mathbf{r}_P \cdot diag\left({\mathbf{R}_P}^{-1}\right) \tag{7.10}$$

Where $\mathbf{I}$ is the identity matrix and $diag\left({\mathbf{R}_P}^{-1}\right)$ is a diagonal matrix with diagonal elements $\left({\mathbf{R}_P}^{-1}\right)_{ii}$ and all off-diagonal elements equal to zero. By expressing $diag\left(\mathbf{r}_P\right)$ from this equation, we obtain,

$$\mathbf{r} = -\left(diag\left({\mathbf{R}_P}^{-1}\right)\right)^{-1} \cdot \left({\mathbf{R}_P}^{-1}\right) \tag{7.11}$$

This final expression gives us the answer: If the global responses of a cellular network to perturbations to all modules have been measured, the network interaction map **r** can be retrieved by the inversion of the global response matrix $\mathbf{R}_P$.

Now we generalise the method for the case where modules have more than one communicating intermediate as output. In Appendix 1, which is published as supporting information to [K+02] on the PNAS web site, an abstract mathematical derivation of this generalised case is provided. In this case, the resulting expression for the network interaction map becomes slightly more complicated. Instead of the diagonal matrix $\left(diag\left({\mathbf{R}_P}^{-1}\right)\right)^{-1}$, the block diagonal matrix **B** is determined using the elements of the inverse matrix ${\mathbf{R}_P}^{-1}$ as follows,

$$\mathbf{B} = \left(blockdiag\left({\mathbf{R}_P}^{-1}\right)\right)^{-1} \tag{7.12}$$

Then, instead of using (7.11), the following equation would be used in order to find the network interaction map.

$$\mathbf{r} = -\mathbf{B} \cdot \left({\mathbf{R}_P}^{-1}\right) \tag{7.13}$$

As an example, remember that we assumed that the interaction network has m modules. For such a network, the matrix **B** has m central minors/square blocks with nonzero elements. All other elements in **B** are zero. The dimension of each block correspond to the number of the communication intermediates or in other words, outputs of the related module.

Suppose that the mth module has c communicating intermediates/outputs. Then, the mth principal minor of the block diagonal matrix **B** looks as follows,

$$\begin{bmatrix} B_{m,m} & B_{m,m+1} & \dots & B_{m,m+c} \\ B_{m+1,m} & B_{m+1,m+1} & \dots & B_{m+1,m+c} \\ \vdots & \vdots & \ddots & \vdots \\ B_{m+c,m} & B_{m+c,m+1} & \dots & B_{m+c,m+c} \end{bmatrix} = \begin{bmatrix} R^{-1}_{m,m} & R^{-1}_{m,m+1} & \dots & R^{-1}_{m,m+c} \\ R^{-1}_{m+1,m} & R^{-1}_{m+1,m+1} & \dots & R^{-1}_{m+1,m+c} \\ \vdots & \vdots & \ddots & \vdots \\ R^{-1}_{m+c,m} & R^{-1}_{m+c,m+1} & \dots & R^{-1}_{m+c,m+c} \end{bmatrix}^{-1} \tag{7.14}$$

7.1.2 Enhancement

There are cases where it is difficult or impossible to measure or estimate the parameter changes (Δp_i). In order to enhance the applicability of the proposed approach in these cases one can simply consider the global ($\Delta_i \ln x_j$) fractional changes in communicating intermediates (x_j) caused by a parameter change (Δp_i). This representation in terms of the relative values may help where quantitation of the absolute activities is difficult. For example, when Western blotting is used to quantify the relative amount of a protein or determining the ratio of the fluorescence intensities from gene arrays [dlFBM02]. To reach the goal, the global response matrix $\mathbf{R}_P$ with the coefficients R_{jp_i} is redefined to be determined by the global fractional changes brought about by a perturbation (Δp_i).

$$R_{jp_i} = \left(\Delta_i \ln x_j\right)_{\text{system in steady state}}, i, j = 1, ..., m. \tag{7.15}$$

Here the derivatives, which were considered in

$$R_{jp_i} = \left(\frac{d \ln x_j}{dp_i}\right)_{systeminsteadystate} \quad j, i = 1, ..., m.$$

are substituted by the finite changes (divided by the initial or mean value). However, the crucial distinction is that according to the equation (7.3), the parameter changes (Δp_i), should be known, whereas equation (7.15) merely considers the differences in intermediates (x_j) before and after perturbations in order to determine the global response matrix $\mathbf{R}_P$. Using equation (7.15) , one obtains exactly the same relationship as (7.11), which expresses the network interaction map in terms of the measured changes in the levels of the communicating intermediates without requiring any knowledge about the values of parameter changes.

7.1.3 Practical Application

According to the method of Kholodenko *et al.*, the following four steps should be applied in an experiment in order to discover the network interaction map by making systematic perturbations and measuring global responses.

- Conceptually divide the network under consideration into interacting modules and identify communicating intermediates.
- Identify the modules which have just one communicating intermediate as output. For these modules, use an inhibitor or other perturbation which affects that single network module only, e.g., module j, and measure the difference in the steady state levels of communicating intermediates before ${x_j}^0$ and after ${x_j}^1$ the perturbation. Then, calculate the jth column of the matrix $\mathbf{R}_P$ by using, e.g., the central fractional differences defined as the finite difference in the activities divided by the mean value,

  $$\Delta_j \ln x_j \approx 2\frac{{x_j}^1 - {x_j}^0}{{x_j}^1 + {x_j}^0} = 2\frac{\left(\frac{{x_j}^1}{{x_j}^0} - 1\right)}{\left(\frac{{x_j}^1}{{x_j}^0} + 1\right)} \tag{7.16}$$

  Repeat for the remaining network modules which also have just communicating intermediate/output, using a perturbation directly affecting that module only, and calculate the corresponding columns of the matrix $\mathbf{R}_P$ $(\Delta_i \ln x_1, ..., \Delta_i \ln x_m)^T$.
- This step is applied to the remaining modules (which have more than one communicating intermediate as output). For these modules, the number of independent perturbations applied to each module should be more than one and equal to the number of communicating intermediates in that module. Determine the block diagonal matrix $\mathbf{B}$ using equation (7.12) and according to (7.14).

  An important point to notice during the last two steps is that perturbations of different kinds can be applied to the modules. This shows the fact that network interactions to be detected by the method would not depend on which particular

modular processes within each module are affected. This emphasises the fact that the method black-boxes the modules.

- Apply the equation (7.13) to reveal and quantify the network interaction map in terms of the matrix **r** of intermodular (local) response coefficients. For the case a single communicating intermediate in each module, the matrix **B** is identical to the diagonal matrix $\left(diag\left(\mathbf{R}_P{}^{-1}\right)\right)^{-1}$.

In [K+02], two examples of applying the proposed methodology are represented. We show some important aspects of these applications in the following.

Unraveling the MAPK cascade interaction map

In this section, the method represented by Kholodenko *et al.* is used to retrieve the interaction map of the MAPK pathway from computer generated responses of a kinetic model of the cascade to perturbations. These responses simulate experimental interventions.

The mitogen-activated protein kinase (MAPK) cascade is a part of the growth-factor/Ras pathway in eucaryotic cells. The cascade is highly conserved, which means that the same principles can be observed in a variety of organisms and cell types. This shows the importance of the cascade. However, although some regulatory feedbacks are well documented, the complete interaction map of the MAPK pathway is unknown. Moreover, both stimulatory and inhibitory feedbacks may differ in various cell types.

To retrieve the MAPK cascade interaction map, we go through the steps in section 7.1.3.

- We consider each of the three tires of the mitogen-activated protein kinase (MAPK) cascade as a functional module that involves unphosphorylated, mono-phosphorylated, and biphosphorylated forms of a protein kinase and the reactions converting these forms. Then, we see that only the doubly phosphorylated forms of the kinases have any interactions with components on other tires/levels of the network. The unphosphorylated and singly phosphorylated forms only interact with components on the same tire/level. Therefore, in this network, the modules correspond to the MKKK, MKK, and MAPK levels and are represented by the communicating intermediates MKKK-PP, MKK-PP, and MAPK-PP, respectively.

- According to the last step, all modules have just one communicating intermediate as output. Then, for each modules, an inhibitor or other perturbation should be used which affects that single network module only. In [K+02] the following perturbations are used. As a perturbation to the first module, the input signal is inhibited by decreasing the Ras-GTP concentration. As a perturbation to the second module, either the maximal activity of the phosphatase, which dephosphorylates MKK-PP and MKK-P, or the kinase that acts on MKK were inhibited. The third module was perturbed by inhibiting either the maximal activity of the MAPK phosphatase or the kinase. 12 different perturbations are used in [K+02] to perturb the whole MAPK cascade.

Before a perturbation to the module j, the steady state level of the communicating intermediate related to this module is measured. After the perturbation, the whole MAPK cascade should be allowed to reach a new steady state. Then, the new steady state level of the communicating intermediate related to the module j is measured. Now, the column j of the matrix $\mathbf{R}_P$ can be calculated using (7.16). You can refer to [K+02] to see the global fractional responses times 100, which shows the changes to communicating intermediates as a percentage of the mean.

- Apply the equation (7.11) to reveal and quantify the network interaction map in terms of the matrix $\mathbf{r}$ of intermodular (local) response coefficients.

A comparison between the retrieved experimental interaction maps and known theoretical interaction map can be seen in [K+02]. It is observed that both different simulated inhibitors and perturbation values, which brought about widely diverse global changes in communicating intermediates, resulted in four nearly identical experimental interaction maps.

Let us consider one of the retrieved interaction maps in [K+02], shown in (7.17), in order to reveal the interactions between MAPK cascade modules.

$$\begin{bmatrix} -1.0 & 0.0 & -1.1 \\ 1.9 & -1.0 & -0.6 \\ -0.0 & 2.0 & -1.0 \end{bmatrix} \tag{7.17}$$

Remember the definition of the network interaction map, $\mathbf{r}$. The ith row of the matrix $\mathbf{r}$ quantifies how module i is affected by each network module through intermediate interaction, whereas the jth column of $\mathbf{r}$ measures how module j directly influences each network module. Therefore, according to the retrieved network interaction map, module 1 directly affects module 2 because $r_{21} > 0$ and but not vice versa, because $r_{12} = 0$. Moreover, the affect of module 1 on module 2 is stimulatory because $r_{21} \succ 0$. In the same manner is module 3 affected by Module 2. From this interaction map, two other direct affects are retrieved which are inhibitory. Module 3 affects module 1 and module 3 affects module 2. The inhibitory nature of these affects is observed from the sign of the corresponding entry in the network interaction map, r_{13}, r_{23}, which is negative. Therefore, the general outline of the MAPK pathway shown in Figure 7.5 can be refined as shown in Figure 7.6 in which the modular framework is used to black box the layers of the cascade.

Unraveling the wiring of a gene network

In this section, the method represented by Kholodenko *et al.* is used to untangle gene network interactions (wiring) by carrying out specially designed gene microarray experiments. To retrieve the gene network interaction map, we go through the steps in section 7.1.3.

- As described before in the definition of the gene networks, the interactions proceed through multiple protein products (for example, transcription factors) and

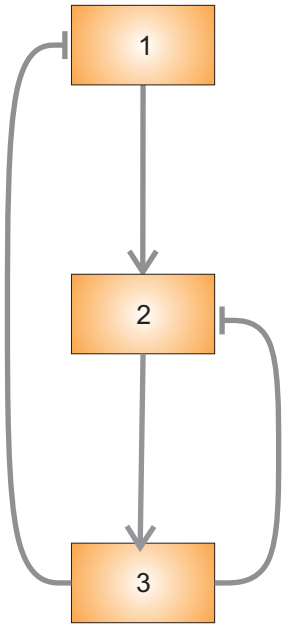


Figure 7.6: Revealed interactions between MAPK cascade modules.

metabolic intermediates. These protein products and metabolic intermediates are not considered explicitly in the method of Kholodenko *et al.* Instead, the mRNAs are considered as communicating intermediates. It is assumed that no knowledge about the interactions is in hand. Actually, Figure **??** shows the only information about the network at this point, which is just the topology of connections. A link represents some association between genes, or more specifically, correlation between the mRNAs.

- According to the last step, all modules have just one communicating intermediate as output. Then, for each modules, an inhibitor or other perturbation should be used which affects that single network module only. In [K+02] two series of four different perturbations to the network. In each set of perturbations, the transcription rate of each gene is perturbed independently by decreasing or increasing its maximal activity.

  Before a perturbation to the module j, the steady state level of the communicating intermediate related to this module is measured. After the perturbation, the whole gene network should be allowed to reach a new steady state. Then, the new steady state level of the communicating intermediate, (mRNA response), related to the module j is measured. Now, the column j of the matrix $\mathbf{R}_P$ can be calculated using (7.16). You can refer to [K+02] to see the two matrices of global fractional responses times 100, which shows the changes to communicating intermediates (mRNAs) as a percentage of the mean.

- Apply the equation (7.11) to reveal and quantify the gene network interaction map

in terms of the matrix **r** of intermodular (local) response coefficients.

A comparison between the retrieved experimental gene interaction maps and known theoretical interaction map can be seen in [K+02]. It is observed that both different simulated inhibitors and perturbation values, which brought about different global changes in communicating intermediates, resulted in two nearly identical experimental interaction maps. Moreover, all the gene interactions were retrieved successfully. Therefore, experimentally obtained network wiring and its quantitation was very near to the known interaction map for this system.

From the application of the method of Kholodenko at al. to the above networks, it is concluded in [K+02] that the method is a powerful tool for unraveling the interactions in signalling and gene networks.

7.2 Method of Sontag et al.

The method of Kholodenko *et al.* represented in the last section has some limitations when to be applied to *in vivo* systems. These limitations appear when for example a biological process is in nature time dependent. The cell cycle is an examples of such a process. In this case, the method of Kholodenko *et al.* which deals only with steady state behaviour can not be applied to such a system. The second case where the method of Kholodenko *et al.* can not be applied is when the biological system includes both information feedbacks and mass flow connections through biochemical conversions. The method of Kholodenko *et al.* implies that network nodes are connected through regulatory interactions that exclude mass flow. Moreover, the perturbations applied to the system in the method of Kholodenko *et al.* should influence just one node at a time. In reality, it happens often in an experiment that an intervention influences more than one module simultaneously [KDMH99, Tys01, SKK04].

In chapter we referred to the realisation of a stochastic process as a time series, which is actually a sequence of observations. For a time series, an observation at time t is modelled as the outcome of a random variable. Compared to stationary data, time series data enable us to understand the dynamics of biological processes. The method represented by Sontag *et al.* in [SKK04] offers a quantitative technique to unravel functional interactions between genes, transcription factors, and metabolites from monitoring time dependent responses of a biological system (in the form of time series data) to perturbations. Using this method, it is possible to untangle functional interactions even when not all the modules can be perturbed directly. In such a case, two or more independent perturbations should be applied to other nodes. In the extreme case, it is possible to deduce all the connections related to a special node in an interaction network from the system responses to perturbations none of which directly influence that particular node.

The method of Kholodenko *et al.* showed how the connection coefficients can be inferred from steady state measurements [K+02]. Method of Sontag *et al.*, published in [SKK04] demonstrates a technique allowing for determination of the dynamics of gene interactions described by the elements of the Jacobian matrix from time dependent network responses to experimental interventions. The inferred interaction dynamics

offers a more valuable and complete description of a cellular network than the less informative steady state data.

7.2.1 Mathematical Derivation

A mathematical description of the method offered by Sontag *et al.* and published in [SKK04] is represented in the following.

Consider again the set of differential equations describing the dynamic behaviour of a system as follows.

$$\frac{dx}{dt} = f\left(x, p\right)$$

Where the vector $x = (x_1, ..., x_n)$ consists of the state variables, each assigned to one network node. Each state variable represents the concentration or the activity level of the corresponding network node. The vector $p = (p_1, ..., p_n)$ consists of the parameters present in the system. Each parameter represents an external or internal condition maintained constant. Examples of internal parameters can be rate constants. External parameters can be PH and temperature. The f functions describe how the rate of change of each state variable depends on all other state variables and parameters present in the system.

Here starts the distinguishing feature of this approach from other network identification techniques. We assume that for each state variable x_i, a set of experimental interventions that do not directly influence x_i exists. Let us assume node i, one of the m nodes present in the network, and call the subset of the parameters which do not affect x_i directly as P_i. From now on, we refer to each member of P_i as p_j. Therefore,

$$\frac{\partial f_i}{\partial p_j}\left(x, p\right) = 0, p_j \in P_i \tag{7.18}$$

For each perturbation, we measure the original and perturbed time series (also called trajectories) describing the time dependence of network variables. These trajectories are solutions to equation (7.4) corresponding to parameter values p_j and $p_j + \Delta p_j$, respectively, and to the same initial condition x^0 (unless a perturbation is a change in the initial condition). The time dependent response $R_{ij}(t)$ of each network variable x_i to a perturbation of p_j is defined as the parameter sensitivity of the solution $x_i\left(t, x^0, p\right)$ to equation (7.4),

$$R_{ij}(t) = \frac{\partial x_i\left(t, x^0, p\right)}{\partial p_j} = \lim_{\Delta p_j \to 0}\left(\frac{x_i\left(t, p_j + \Delta p_j\right) - x_i\left(t, p_j\right)}{\Delta p_j}\right), i = 1, ..., m \tag{7.19}$$

These sensitivities are computed using the variational system along the corresponding trajectory and are routinely employed in differential equation theory and systems analysis for parameter optimisation and identification of dynamical systems. They have been also used in the context of metabolic networks [IS03]. Our objective is to determine dynamic connections, given by the Jacobian elements $J_{ij}(t)$ from the experimental time series that evaluate global response coefficients $R_{ij}(t)$. To this aim, we also need

the second order sensitivities, $\rho_{ij}(t)$, which can be estimated from the measurements of the global responses $R_{ij}(t)$, at two successive time points t and $t + \Delta t$.

$$\rho_{ij}(t) = \frac{\partial^2 x_i\left(t, x^0, p\right)}{\partial p_j \partial t} = \frac{\partial R_{ij}(t)}{\partial t} \approx \frac{\left[R_{ij}\left(t + \Delta t\right) - R_{ij}(t)\right]}{\Delta t} \tag{7.20}$$

A new matrix used in this method is the global response matrix $\mathbf{R}(t, P_i)$, which is composed of the measured time dependent responses, $R_{kj}(t)$. Each jth column of the response matrix $\mathbf{R}(t, P_i)$ corresponds to a single experiment. In this experiment, the parameter p_j is perturbed and the time course of the response of each network node x_k to a change in p_j is evaluated. Therefore, the number of rows of matrix $\mathbf{R}$ equals to the number of nodes, m, in the network. Matrix $\mathbf{R}$ has as many columns as selected parameters p_j that directly influence either a single node different from x_i, or any combination of such nodes. For each perturbation experiment related to p_j, the second order sensitivity $\rho_{ij}(t)$ is evaluated from time series for node x_i. Using the matrix $\mathbf{R}$ and the $\rho_{ij}(t)$ values, the problem can be solved as follows.

The dynamic behaviour of the node i is described by the solution $x_i\left(t, x^0, p\right)$ (x^0 is the initial state of the system) to the following equation,

$$\frac{dx}{dt} = f\left(x, p\right), x = \left(x_1, ..., x_n\right), p = \left(p_1, ..., p_n\right)$$

Therefore,

$$\frac{dx_i\left(t, x^0, p\right)}{dt} = f_i\left(x_1, ..., x_i, ..., x_j, ..., x_n, p\right)$$

Taking the derivative with respect to p_j from both sides and using equations (7.18), (7.19), and (7.20), the unknown elements of the ith row of the Jacobian matrix $\mathbf{J}$ would be found. These elements satisfy the following system of linear equations,

$$\rho_{ij}(t) = \sum_{k=1}^{m} R_{kj}(t) \cdot J_{ik}(t), p_j \in P_i \tag{7.21}$$

The quantities we are looking for, the m Jacobian matrix entries, J_{ik}, can be found from equation (7.21) provided that the the rank of the matrix $\mathbf{R}(t, P_i)$ equals m. These quantities quantify the influence of every node k on node i and can be determined from the measured responses, $R_{kj}(t)$ and $\rho_{ij}(t)$, to the perturbations which do not directly affect node i.

There are noticeable points in the Sontag *et al.* method. To deduce dynamic connections leading to node i, the second order sensitivities ρ_{ij} should be measured only for component x_i. Compare it to the fact that for all other components x_k, the first order sensitivities $R_{kj}(t)$ should be determined. Another point is the fact that the rank of the response matrix $\mathbf{R}(t, P_i)$ generically at any given time equals m. It is demonstrated in the supplementary material of the paper published in [SKK04]. Moreover, it is showed that this rank generically equals m even when only a single network node is directly

affected by m experimental interventions, each of which changes an independent parameter influencing that particular node. Another point can be the interesting fact that perturbations can be changes in the initial conditions.

We close this section with the conclusion that estimates of the time varying sensitivity coefficients allow us to completely infer and quantify the network connections even if experimental interventions can directly perturb only selected network components, the number of which is less than the number of nodes in the network.

7.2.2 Enhancement

There are cases where it is difficult or impossible to measure or estimate the parameter changes. In fact, such measurements would be difficult if not impossible to make *in vivo*. In order to enhance the applicability of the proposed approach in these cases, we need a method to express the network connections in terms of the measured changes in the levels of the intermediates without requiring any knowledge about the values of the parameter changes. To this aim, one can simply consider the global changes Δx_i in network variables caused by a perturbation Δp_j and induce the quantities $\Delta R_{ij}(t)$ and $\Delta \rho_{ij}(t)$ using finite differences as an approximation of mathematically correct infinitesimal changes.

$$\Delta R_{ij}(t) = x_i\left(t, p_j + \Delta p_j\right) - x_i\left(t, p_j\right) \tag{7.22}$$

$$\Delta \rho_{ij}(t) = \frac{\Delta R_{ij}\left(t + \Delta t\right) - \Delta R_{ij}(t)}{\Delta t} \tag{7.23}$$

The same equation as (7.21) is obtained as follows.

$$\Delta \rho_{ij}(t) \approx \sum_{k=1}^{m} \Delta R_{kj}(t) \cdot J_{ik}(t), p_j \in P_i \tag{7.24}$$

The difference between the equation (7.21) and (7.24) is that $\Delta R_{ij}(t)$ and $\Delta \rho_{ij}(t)$ are appeared in the latter equation instead of using the absolute values of $R_{ij}(t)$ and $\rho_{ij}(t)$ in the former one.

7.2.3 Practical Application

According to the method of Sontag *et al.* in [SKK04], the following steps should be applied in practise in order to be able to reconstruct cellular networks from time series of gene expression, signalling and metabolic data.

- Conceptually divide the network under consideration into interacting modules and identify the state variables x_i considering both regulatory interactions and mass flow. Here we assume that there are m modules in the network.
- Apply an experimental setup where the states or activities of nodes can be measured during a transient process. For example, in such an experimental setup, the states of nodes should be measurable throughout a transition from a resting state to an active state of the network. There are two possibilities for the transient

process to get started. Either is the system behaviour inherently transient or the transient process is initiated by stimulation. Examples for the two cases can be the cell cycle for the former and cell stimulation by a ligand for the latter case.

- For the state variable x_i corresponding to the node i, select an experimental intervention that does not directly influence x_i. This perturbation however possibly directly affects one or several nodes different from the node the state variable x_i is associated to. Use the biological information available about the system. Examples of these prior information can be that a certain protein has no direct influence on an unrelated gene, or a certain inhibitor of a membrane kinase has no direct influence on a cytoplasmic phosphatase. It can be a change in external ligand concentration, a change in the initial concentration, activity of a component (node) different from node i, a pharmacological manipulation, or the use of nucleic acid based technologies such as tetracycline inducible expression and small RNA interference. At selected time points, monitor and measure the original (unperturbed) and perturbed values of all m network nodes and determine the differences according to equation (7.22) and equation (7.23). From the equation (7.22), column j of the matrix $\mathbf{R}(t, P_i)$ is filled. Equation (7.23) on the other hand, provides one of the second order sensitivities $\rho_{ij}(t)$. It is important to note that it is often convenient to normalise the differences by the mean values and determine the fractional changes.

- Repeat the last step to do as many perturbation experiments as there are nodes (m). For each experiment, determine the global response matrix $\mathbf{R}(t, P_i)$, and the sensitivity coefficient $\rho_{ij}(t)$. At the end, there would be m sensitivity coefficients and one global response matrix for each selected time point and node.

- At each selected time point t, solve equation (7.21) or (7.24) in order to find the elements $J_{ij}(t)$ of the Jacobian matrix $\mathbf{J}(t)$. In this step, it is important to check if the network approaches a steady state. Steady state condition necessitates perturbations to each network node. The reason is that the rank of the matrix $\mathbf{R}(t, P_i)$ decreases as time approaches infinity. In fact, the maximal possible rank of the steady state response matrix $\mathbf{R}(t, P_i)$ equals $n - 1$ when time goes to infinity. At least $n - 1$ perturbation experiments are required to achieve this rank. All nodes except node i should get perturbed in these perturbation experiments. Each perturbation however, can be to each node or to a combination of the nodes excluding node i. Actually, when quasi-steady state behaviour is displayed by the system under consideration, the Jacobian matrix elements can not be determined. Instead, it is possible to determine the scaled Jacobian matrix elements or in other words, the connection coefficients r_{ij}, using the equation (7.25) represented in [K+02].

$$\sum_{k=1,k\neq i}^{m} \Delta R_{kj} \cdot r_{ik} \approx \Delta \mathbf{R}_{ij} \tag{7.25}$$

This way, the connection coefficients r_{ij} are determined using the finite differences

ΔR_{ij}. These finite differences in turn, are equal to changes in state variables following a transition from an initial state to a new steady state.

- Repeat the third and forth steps for the remaining nodes. An important point to consider in this step is that a specific measurement which forms a column of the global response matrix, can be used for two or more different nodes provided that these nodes are not directly affected by the specific perturbation related to that measurement. The complete interaction network would be obtained at the end of this step.

In [SKK04], two examples of applying the proposed methodology are represented. We show some important aspects of these applications in the following.

Reverse engineering of a signalling network

In this section, the method represented by Sontag *et al.* is used to retrieve the interaction map of the MAPK pathway from computer generated time dependent responses of a kinetic model of the cascade to perturbations. The kinetic model used for this pathway consisting of rate expressions, differential equations and parameter values is published as the supplementary material of [SKK04]. This model is the same model generated in [K$^+$02] and shown in Figure 7.6. The time dependent responses of the network variables in the model are used to retrieve the interaction map or in other words to reverse engineer the signalling network. Let us go through the steps in section 7.2.3.

- The network has six nodes. The reason is the moiety conservation which makes us consider two independent variables at each level of the network. In contrast to the method of Kholodenko *et al.*, we consider the mass flow between unphosphorylated, monosphorylated and biphosphorylated protein forms at each MAPK cascade level besides the information flow between the three layers. Therefore, the nodes of the network correspond to MKKK-P, MKKK-PP, MKK-P, MKK-PP, MAPK-P, MAPK-PP.

- Apply an experimental setup, where the six states, or activities of nodes can be measured during a transition of the MAPK pathway from a resting state to a stable activity state. In all simulations, the initial condition at $t = 0$ corresponded to the steady state of the MAPK pathway with a low Ras activity. When $t > 0$, the RasGTP level increases to a new high value and the transition from the steady state with a low activity to a high activity state is observed. A point to consider is that the responses $R_{ij}(0) = 0$ at time zero, since both perturbed and unperturbed solutions have the same initial condition.

- for each node, six different perturbations are applied to the network, each revealing the connections leading to that node, correspondingly one row of the Jacobian matrix. Each perturbation can affect one or several reactions or nodes in the pathway.

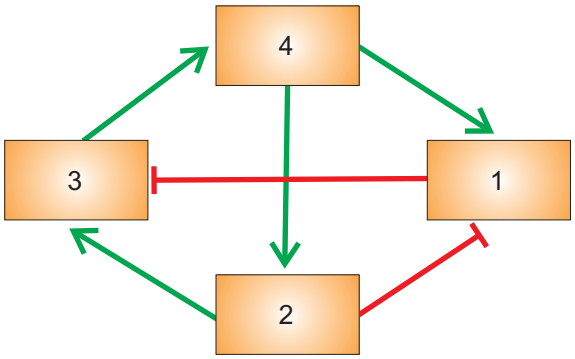


Figure 7.7: Schematic four-gene network.

- The finite differences between the control (unperturbed) and perturbed responses (transitions) of all the nodes of the network are observed and evaluated using equations (7.22), (7.23) and (7.24) for three different perturbation magnitudes.
- At each selected time point t, solve equation (7.24) numerically in order to find the elements $J_{ij}(t)$ of the Jacobian matrix $\mathbf{J}(t)$.

A comparison between the correct (theoretical) Jacobian matrix elements with experimentally retrieved interaction strengths shown in [SKK04] demonstrates that the architecture of the entire MAPK pathway is correctly deduced from the time dependent responses of the model to the perturbations.

Reverse engineering of a gene network

In this section, the method represented by Sontag *et al.* is used to reconstruct or in other words, reverse engineer the four gene network shown in Figure 7.7. The gene network is first modelled in computer using the rate expressions, differential equations and parameter values of the gene network model published as the supplementary material of [SKK04]. The computer generated time dependent responses of the network variables in this model to perturbations are used to retrieve the network interaction map. Let us go through the steps in section 7.2.3.

- The network has $m = 4$ nodes representing the mRNA concentrations. Therefore, the system of equations describing the rate of change of network nodes (their concentrations) consists of 4 differential equations each described by the difference between the transcription and degradation rate of one node as bellow,

$$\frac{d[mRNA_i]}{dt} = v_i^{\text{synth}} - v_i^{\text{degr}}$$

  The transcription and degradation rate equations of all the 4 genes besides the Michaelis constants appearing in them are represented in the supplementary material of [SKK04].

- An experimental setup is applied, where the four states, or activities of nodes can be measured during a transition of the gene network from a resting state to a stable activity state. In all simulations, the initial condition at $t = 0$ corresponds to the steady state of the gene network where all four genes are inactive (the catalytic constants of the transcription rates and therefore the concentration of all genes are zero). When $t > 0$, the constants were assigned the values given in the supplementary material of [SKK04] and the transition to an active state begins.

- for each node, four different perturbations are applied to the network, each revealing the connections leading to that node, correspondingly one row of the Jacobian matrix. Each perturbation can affect one or several reactions or nodes in the gene network through a change in a transcription or degradation rate.

- The finite differences between the control (unperturbed) and perturbed kinetic model generated responses (transitions) of all the nodes of the network is observed and evaluated using equations (7.22), (7.23) for different perturbation magnitudes.

- At each selected time point t, solve equation (7.24) numerically in order to find the elements $J_{ij}(t)$ of the Jacobian matrix $\mathbf{J}(t)$. This way, both the architecture and the strength of functional interactions between genes in the network during the transition of the network from steady state to an stable activity state is inferred. In [SKK04] the inferred dynamics of activation or repression of each gene by another (in terms of the Jacobian matrix element J_{ij}) is schematically illustrated and compared with the correct interaction strengths. A very important point to conclude is that the deviation between the retrieved and correct interaction strengths begins to rise as steady state is approached. The reason is that when quasi-steady state behaviour is displayed by the system under consideration, the Jacobian matrix elements can not be determined. Instead, it is just possible to determine the scaled Jacobian matrix elements or in other words, the connection coefficients r_{ij}, using the equation (7.25) represented in [K$^+$02]. However, in absence of quasi-steady state, the architecture of the entire gene network is deduced from the time dependent responses of the model to the perturbations with a high accuracy.

7.3 Conclusion

One of the main objectives of system biology research is the unravelling of the functional interaction structure of a biomolecular network from a given set of experimental data. If we can estimate the direct influence (or local response) of the variation of one node (a functional unit or module of the bimolecular network) on all other nodes in a biomolecular network, then we can identify the local network structure and from this we can also infer the global network structure **??**. However, carrying out practical experiments in order to quantify the change in the transcription level of one gene, owing to a change in another gene while keeping everything else constant is practically impossible. Instead, what we can typically measure in a practical experience is for example in the case of a

gene network, the global change in the transcription level of a gene owing to some perturbation to the system. The reason is that the variation of one node promptly spreads over the other nodes of the network till the whole network reaches a new steady state. Therefore, a method is required which can be used to indirectly infer the local responses based on information from global changes measured in a practical experiment.

Considering the disadvantages of qualitative and mechanistic modelling methods, known also as bottom-up approaches, such as the necessity of modelling a large number of molecular processes (especially when the values of multiple kinetic parameters may be unknown), and the fact that button-up approaches inevitably miss the interactions and regulatory feedbacks still awaiting discovery, another group of approaches, namely top-down approaches came to existence.

In [K+02] a general methodology was proposed by Kholodenko *et al.* that is applicable to a network of modules (i.e., a combination of genes, proteins, and other species). The assumption made in this method is that each module i contains at least one intrinsic parameter p_k that can be directly perturbed without the intervention of other nodes or parameters and the following equation holds for it. Note that m is the number of modules in the network.

$$\frac{\partial f_i}{\partial p_k} = 0, 1 \leq k \leq m, k \neq i \tag{7.26}$$

The key point of the method is actually using these intrinsic parameters and the stationary experimental data to find the coefficient r_{ij} which are referred to as local response or local coefficient and quantify the sensitivity of module i to module j. If we use the sign of ∞ for the stationary experimental data, the algorithm of the method of kholodenko *et al.* can be summarised as follows. Find r_{ij} satisfying the following equation.

$$\frac{\Delta {x^\infty}_{i,k}}{{x^\infty}_i\left(p_k\right)} = \sum_{1\leq j\leq m, j\neq i} r_{ij} \frac{\Delta {x^\infty}_{j,k}}{{x^\infty}_j\left(p_k\right)} \tag{7.27}$$

The Kholodenko approach represents a distinct and fresh approach to the problem of identification of a gene network based on stationary experimental data. However, there are still some problems with the method. For example, in order to identify the functional interaction structure of a network of m nodes, we need to find the solutions of the experimental algebraic equations (7.27), where each solution implies the strength of the direct influence of one node on the other node. However, it is uncertain how the experimental algebraic equations for determining the solutions are derived and why all the solutions regarding the influence of one node on itself are set to -1. However, the latter point may be for numerical reasons associated with the inversion of a certain matrix required in the solution procedure. However, further explanations are required regarding whether we must calculate the inverse matrix to obtain the solutions and why special algebraic forms are used in the experimental equations. Most importantly, the proposed approach is only applicable to the case when there exists a unique parameter for each node of the network. Furthermore, this parameter should directly and exclusively affect the corresponding node.

Sontag *et al.* have also proposed a technique in [SKK04] that is complementary to Kholodenkos method. The assumption made for applying this method is the existence

of a parameter set $\{p_{ik} \mid 1 \leq k \leq m\}$ independent of the node i. The key point of the method is actually using this parameter set satisfying the assumption of $\frac{\partial f_i}{\partial p_{ik}} = 0, 1 \leq k \leq m$ and the temporal experimental data $x_i\,(t, p_{ik})$ to find the Jacobian matrix elements $J_{ij}(t)$. The algorithm of the method of Sontag *et al.* can be summarised as follows. Find $J_{ij}(t)$ satisfying the following equation.

$$\frac{R_{ip_{ik}}(t + \Delta t) - R_{ip_{ik}}(t)}{\Delta t} = \sum_{1 \leq j \leq m} J_{ij}(t) R_{jp_{ik}}(t) \tag{7.28}$$

The Sontag approach is based on time series measurements. This fact shows the advantage of this method to the Kholodenko method when steady state data are not available. Another advantage of this method is hat the strength of self regulation at each node or module is not known in advance and should also be estimated. The Sontag approach also has the merit that the estimated structure can be repeatedly refined over each sampling time point, thereby compensating in some sense for the effects of experimental errors. Despite these nice features, the method still contains some restrictions. Specifically, it is only applicable in the case when for each node there are as many parameters as the number of overall network nodes, and these parameters do not directly affect the corresponding node. Moreover, the number of parameter perturbations can be increased since we should determine the parameters to be perturbed at each node. Furthermore, there is no guidance for choosing sampling time intervals in the time series measurements. Likewise, the effect of approximating the time derivative used in the experimental algebraic equations in (7.28) is not considered.

A common issue in both Kholodenko and Sontag methods is the fact that the estimation error can become unacceptably large. This issue together with others expounded in the previous paragraphs lead to a number of literature trying to find ways to extend and unify the two approaches. An example can be found in [CCWW05] where an approach based on stationary and/or temporal data obtained from parameter perturbations is represented. The approach unifies the two methods of kholodenko *et al.* and Sontag *et al.* to represent a comprehensive unified framework. A novel experimental design procedure is developed, whereby the estimation error of the network interaction structure can be reduced and the correct estimation of the qualitative structure, such as activation and inhibition between nodes, can be guaranteed.

Glossary

Most biological textbooks have a glossary, for mathematical expressions we refer to Eric Weisstein's MathWorld web-site http://mathworld.wolfram.com/ or a mathematics dictionary (e.g. [BB89]).

abscissa The horizontal axis or x-coordinate of a point in the two-dimensional plane. See also *ordinate*.

absolute value The positive value of a number, disregarding its sign and written $|x|$.

activation loop A segment of the amino acid sequence that contains phosphorylation sites usually at the surface of a protein and accessible by protein kinases.

active site Region of an enzyme surface to which a substrate molecule binds in order to undergo a catalyzed reaction.

active transport Movement of a molecule across a membrane or other barrier driven by energy other than that stored in the electrochemical gradient of the transported molecule.

adaptors Adaptor proteins typically do not posses a catalytic function but bind to other proteins. Adaptors serve to physically connect proteins with each other. See also *exchange factors*.

algebra A branch of mathematics that generalises arithmetic operations with numbers to operations with variables, matrices etc.

amino acid Class of biochemical compounds from which proteins are composed. Around 20 amino acids are present in proteins.

analysis A branch of mathematics concerned primarily with limits of functions, sequences, and series.

analytic function A function possessing derivatives of all orders and agreeing with its Taylor series locally.

antibody A protein molecule produced in response to an antigen.

antigen Molecule that is able to provoke an immune response.

apoptosis Controlled cell death.

argument The argument of a function is the element to which a function applies. Usually the independent variable of the function.

associative A law or operation is called associative if the placement of brackets does not matter: $(a \cdot b) \cdot c \equiv a \cdot (b \cdot c)$.

ATP The principal carrier of chemical energy in cells.

attractor A region of the space describing the temporal solution of a dynamic system towards which trajectories nearby converge, are attracted to. An attractor can be a equilibrium point or a circle. An attractive region that has no individual equilibrium point or cycle is referred to as a chaotic or strange attractor.

autocatalysis Reaction that is catalyzed by one of its produces, creating a positive feedback (self-amplifying) effect on the reaction rate.

autoinhibition Mechanism for inhibiting own activity; e.g., Raf contains an autoregulatory domain that inhibits its own activity by binding to its catalytic domain. The autoregulatory domain is relieved from the catalytic domain by phosphorylation of characteristic residues.

autonomous A system (of differential equations) is said to be autonomous if it does not *explicitly* depend on time.

bifurcation point An instability point in which a single equilibrium condition is split into two. At a bifurcation point the dynamics of a system changes structurally.

bioinformatics The management and analysis of genomic data, most commonly using tools and techniques from computer science.

calculus A branch of mathematics concerned with the rate of change of a dependent variable in a function.

category theory A branch of mathematics that considers mappings and their effect on sets. A category is a structure consisting of a set of objects and a class of maps, which satisfy specific properties.

chain rule A rule used in the context of differential equations and which states that $dy/dx = dy/dt \times dt/dx$.

class Another name for set, especially a finite set.

closed form An expression or solution in terms of well understood quantities.

coefficient A numerical or constant multiplier of a variable in an algebraic term.

continuous function A function for which the value changes gradually.

control Target or set-point tracking, making the system sensitive to changes in the input. See also *regulation* and *homoeostasis*.

control coefficient Relative measure of how much a perturbation of a systems parameter affects a systems variable. The control coefficient is a global property and one of three building blocks in Metabolic Control Analysis.

cytokine Extracellular signal protein or peptide that acts as a local short distance mediator in cell-cell communication. Cytokines are called lymphokines if produced by lymphocytes, interleukines if produced by leucocytes, and monokines if produced by monocytes and macrophages.

cytoplasm Contents of a cell that are contained within its plasma membrane but, in the case of eucaryotic cells, outside the nucleus.

damped oscillations An oscillation in which the amplitude decreases over time.

differentiable A system (usually a process described by differential equations) is called differentiable if its *phase space* has the structure of a differentiable *manifold*, and the change of *state* is described by differentiable functions.

differentiation process by which the cell acquires specialised functional properties.

dimer A protein molecule which consist of two subunits separated polypeptide chains); homodimer: the subunits are identical; heterodimer: the subunits are different; heterotrimer: three subunits, some different.

dimerisation The process by which two molecules of the same chemical composition form a condensation product or polymer.

discretisation An approximation of a continuous object.

dynamic system A system that changes with time.

EGF Epidermal Growth Factor. EGF is expressed by many cells and stimulates the proliferation of many cell types via Ras and the Raf/MEK/ERK pathway.

EGFR EGF Receptor, a prototypical receptor tyrosine kinase.

elasticity coefficient Relative measure of the dependence of the reaction rate of an isolated step in a metabolic pathway on systems variables and internal parameters. The elasticity coefficient is a local or system property and one of three building block of Metabolic Control Analysis.

electrophoresis An experimental technique to separate DNA fragments or proteins from a mixture. The molecules are separated by their mass, size or rate of travel through a medium (typically agarose or gels) and their electrical charge.

enzyme Protein that catalyzes a specific chemical reaction.

epithelial A epithelial is a coherent cell sheet formed from one or more layers of (epithelial) cells covering an external surface or lining a cavity. For example, the epidermis is the epithelial layer covering the outer surface of the body.

equilibrium State where there is no net change in a system. E.g. in a chemical reaction the equilibrium is defined by the state at which the forward and reverse rates are equal.

equilibrium point Point such that the derivatives of a system of differential equations are zero. An equilibrium point may be stable (then called an *attractor*) or unstable (*repellor*).

exchange factors Bind to the activated receptor, i.e., act as an adaptor; facilitate the exchange of bound GDP for GTP on small G-proteins, which are several steps away from the receptor, and thus activate them.

expression Production of a protein which has directly observable consequences.

extended phase space See *phase space*.

feedback inhibition Regulatory mechanism in metabolic pathways - an enzyme further up in the pathway is inhibited by a product further down in that pathway.

finite-dimensional A process is called finite-dimensional if its *phase space* is finite dimensional, i.e., if the number of parameters needed to describe its *states* is finite.

fixed point See *steady state*.

formal system A mathematical framework in which to represent natural systems.

fun What we experience doing mathematics.

function A relation between two sets that describes unique associations among the elements of the two sets. A function is sometimes called a mapping or transformation.

G-proteins Small monomeric GTP-binding proteins (e.g. Ras), molecular switches that modulate the connectivity of a signalling cascade: resting G-proteins are loaded with GDP and inactive, replacement of GDP with GTP by exchange factors means activation.

GAP GTPase Activating Protein. Ras proteins possess intrinsic GTPase activity which hydrolyzes the bound GTP to GDP, i.e., cleaves off a phosphate from GTP. This hydrolysis is a dephosphorylation and as such a phosphatase reaction. A dephosphorylation or phosphatase reaction is a special case of a hydrolysis reaction. Hydrolysis reactions are all reactions where water, H_2O, is used to break a chemical bond. The intrinsic GTPase activity of Ras is weak. However, GAPs can accelerate this activity almost 1000fold. GAPs do not hydrolyze GTP, they bind to Ras and make Ras a more efficient GTPase.

gene expression The process by which the information, coded in the genome, is transcribed into RNA. Expressed genes include those for which the RNA is *not* translated into proteins.

gene product The macromolecules, RNA or proteins, that are the result of *gene expression*.

genome The entirety of genetic material (DNA) of a cell or an organism.

gradient The slope of a line measured as the ratio of its vertical change to its horizontal change.

Grb-2 Growth-factor Receptor Binding protein-2. Grb-2 is an adaptor protein.

group A mathematical group is a set, together with a binary operation on the group elements.

growth factor Extracellular signalling molecule that can stimulate a cell to grow or proliferate.

GTP/GDP Guanosine triphosphate (GTP) refers to three phosphate molecules attached to the sugar, guanosine diphosphate for two (GDP). See also *GAP*.

homoeostasis Regulation to maintain the level of a variable. See also *regulation*.

homologues proteins/genes Have descended from a common ancestor; genes are either homologous or non-homologous, not in between; though, due to multiple genomic rearrangements, the evolutionary history of individual components (domains = evolutionary units) of a gene/protein might be difficult to trace.

hydrolysis See *GAP*.

immunoglobin General expression for antibody molecules.

in vitro Experimental procedures taking place in an isolated cell-free extract. Cells growing in culture, as opposed to an organism.

in vivo In an intact cell or organism.

in silico In a computer, simulation.

infinitesimal Infinitely small. Infinitesimal quantities are used to define integrals and derivatives, and are studied in the branch of maths called analysis.

integral curve A trajectory in extended phase space.

isoforms Closely homologous proteins (from different genes) that perform similar or only slightly different functions, e.g., under tissue-specific control. Two or more RNAs that are produced from the same gene by different transcription and/or differential RNA splicing are referred to as isoforms.

kinase Enzyme which catalyzes the phosphorylation of a protein.

ligand Molecule that bind to a specific site on a protein or other molecule.

linear equation An equation $y = ax + b$ is linear because the graph of y against x is a straight line (with slope a and intercept b. A linear equation should not be confused with a linear system. See also *nonlinearity*.

linear system A system is nonlinear if changes in the output are not proportional to changes in the input.

linearisation Taylor expansion of a dynamical system in the dependent variable about a specific solution, discarding all but the terms linear in the dependent variable.

linearity Linearity is defined in terms of functions that have the property $f(x + y) = f(x) + f(y)$ and $f(ax) = af(x)$. This means that the result f may not be proportional to the input x or y.

locus The position of a gene on a chromosome, the DNA of that position; usually restricted to the main regions of DNA that are expressed.

lysis Rupture of a cell's plasma membrane, leading to the release of cytoplasm and the death of the cell.

manifold A mathematical space in which the local geometry around a point in that space is equivalent to the Euclidean space.

MAP-kinase Mitogen-activated protein kinase that performs a crucial step in transmitting signals from the plasma membrane to the nucleus.

Metabolic Control Analysis (MCA) Method for analyzing variation in fluxes and intermediate concentrations in a metabolic pathway relating to the effects of the different enzymes that constitute the pathway and external parameters. The building blocks of MCA are: control coefficients, elasticity coefficients, and response coefficients.

metabolism The entirety of chemical processes in the cell.

mitogen Substance that stimulates the mitosis of certain cells.

mitosis Process in cell division by which the nucleus divides.

monomer A protein molecule which consist of one subunits separated polypeptide chains; homodimer: the subunits are identical; heterodimer: the subunits are different; heterotrimer: three subunits, some different.

morphism Generalisation of the concepts of relation and function. Often synonymously used with mapping.

multimer A protein molecule which consist more than four subunits separated polypeptide chains); homodimer: the subunits are identical; heterodimer: the subunits are different; heterotrimer: three subunits, some different.

natural system An aspect of the phenomenal world, studied in the natural sciences.

noise A description of real or simulated data for which the behavior is or appears unpredictable.

oncogene An altered gene whose product which takes a dominant role in creating a cancerous cell.

orbit The set of points in phase space through which a trajectory passes.

ordinate The vertical or y-axis of the coordinate system in the plane.

organisation Pattern or configuration of processes.

peptide A small chain of amino acids linked by peptide bonds.

percepts The consequence of cognitive processes or observations.

phase space Phase space is the collection of possible states of a dynamical system, i.e., the mathematical space formed by the dependent variables of a system. An extended phase space is the cartesian product of the phase space with the independent variable, which is often time.

phenomena A collection of percepts to which relationships are assigned.

phosphatase Enzyme that removes phosphate groups from a molecule.

phosphorylation Important regulatory process, one third of mammalian proteins are regulated by reversible phosphorylation; phosphate groups P from ATP molecules are transferred to the -OH groups of serine, threonine or tyrosine residues by protein kinases; phosphate groups are two times negatively charged, their addition will change the protein's local conformational characteristics and can thus activate a protein. See also *GAP* and *protein phosphorylation*.

polymer Large molecule made be linking monomers together.

protein A linear polymer of linked amino acids, referred to as a macromolecule and major constituent component of the cell.

protein kinase Enzyme that transfers the terminal phosphate group of ATP to a specific amino acid of a target protein.

protein phosphorylation The covalent addition of a phosphate group to a side chain of a protein catalyzed by a protein kinase.

proteinase, protease Enzymes that are degrading proteins by splitting internal peptide bonds to produce peptides.

proteinase inhibitor small proteins that inhibit various proteinase enzymes. An example is antitrypsin.

random process A description of real or simulated data for which the behavior is or appears unpredictable.

RAS protein Member of a large family of GTP-binding proteins that helps transmit signals from cell-surface receptors to the nucleus. Ras-GDP is the inactive form of Ras, which is bound to Guanosine-Di-Phosphate. Ras-GTP is the active form of Ras, which is bound to Guanosine-Tri-Phosphate. This form of Ras undergoes a conformational change that enables it to bind with high affinity to other proteins such as Raf.

receptor tyrosine kinase Receptor tyrosine kinases play an important role in the regulation of cell proliferation, survival and differentiation. The binding of the ligand (including growth factors, hormones etc.) to the extracellular portion of the receptor typically activates the kinase activity of the intracellular portion of the receptor, resulting in autophosphorylation on several tyrosine residues. the phosphorylated tyrosines serve as docking sites for adaptor proteins such as Grb-2 resulting in the assembly of a multiprotein complex at the receptor. This complex is a platform that typically mediates the specific biological responses by activating several intracellular signalling pathways.

regulation The maintenance of a regular or desirable state, making a system robust against perturbations. See also *homoeostasis* and *control*.

repressor Protein that binds to a specific region of DNA to prevent transcription of an adjacent gene.

residue Proteins are built of amino acids by forming peptide bonds under removal of water; what remains of the amino acids are the amino acid residues.

response coefficient Relative Measure of the dependence of a system variable of a pathway on an external parameter. The response coefficient is one of three building blocks in Metabolic Control Analysis.

sample space The set of possible outcomes in a statistical experiments.

scaffold protein Protein that organises groups of interacting intracellular signalling proteins into signalling complexes.

sigma algebra A σ-algebra is a collection of subsets of a set that contains the set itself, the empty set, the complements in the set of all members of the collection, and all countable unions of members.

signalling, signal transduction A process by which signals are relayed through biochemical reactions.

SOS Son of Sevenless. SOS is the prototypic GDP/GTP Exchange Factor, GEF. There are many GEFs, but SOS is ubiquitously expressed. GEFs cause Ras to release GDP. Since the cell contains much higher concentrations of GTP than GDP, per

default a GTP molecule will bind to Ras in place of the released GDP. Oncogenic Ras mutants cannot release GDP. Therefore, they are always in the active (GTP bound) form.

steady state A system state in which the system remains. A steady state is associated with a *fixed point*, i.e., the point in the state-space in which the system remains.

stochastic process A mathematical concepts defined as a sequence of random variables.

system A collection of objects and a relation among these objects.

tangent bundle The set of tangent vectors to a manifold.

terminal domain N-terminal domain, C-terminal domain chain of amino acid residues leaves an amino group free at one end, and a carboxyl group at the other end; by convention a protein chain starts at the N-terminus, i.e., the N-terminal domain is the first domain near the amino terminus; the C-terminal domain the last near the carboxyl terminus.

tetramer A protein molecule which consist of four subunits separated polypeptide chains; homodimer: the subunits are identical; heterodimer: the subunits are different; heterotrimer: three subunits, some different.

TNF Tumor necrosis factor, protein produced by macrophages in the presence of an endotoxin.

trajectory The solution of a set of differential equations, synonymous with the phrase phase curve.

tyrosine kinase See *receptor tyrosine kinase*.

vector A mathematical vector is an ordered set of elements, e.g., (a, c, b). An unordered list is denoted $\{a, b, c\}$, where the position of the elements in the list does not matter.

Notation

The notation used in this text was one of the biggest challenges. Since we are dealing with various aspects of mathematics and different application areas, there are conflicting customary uses of symbols. For example, in stochastic modelling we use n to denote the state vector, i.e., the number of molecules at any particular time. In modelling with differential equations, n is a constant used to denote the number of equations, $\dot{x}_i$, $i = 1, \ldots, n$. The letter x refers to a variable, random variable $x(t)$, vector $x = (x_1, \ldots, x_n)$, ... An effort is made to introduce notation and symbols where they appear first. According to convention in biological textbooks, acronyms printed in lower case indicate genes (e.g. ras), capitalised acronyms indicate their protein products (Ras or RAS).

Units

L	liter.
Da	Dalton.
mol	moles, molar mass.
M	molarity, molar concentration.
sec	seconds.
min	minutes.
g	grams.

Mathematical Symbols

$\rightarrow$	mapping, function, morphism, arrow.
$\mapsto$	"maps to".
:	"for which", "such that".
$\vert$	"conditional on".
$\forall$	"for all".
$\in$	"element of".
$\doteq$	"by definition".
$\exists$	"there exists".
$\equiv$	"equivalent", "identical".
$\propto$	"proportional to".
$\approx$	"approximately".
$\Rightarrow$	"implies", material implication.
$\Leftrightarrow$	"if and only if" (iff).
$\therefore$	"therefore".

N_A	Avogadro number.
$\{\ \}$	set, list.
$(\ )$	ordered set, sequence, vector.
$\mathbb{Z}$	set of integers $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$.
$\mathbb{Z}_+$	set of nonnegative integers $\{0, 1, 2, \ldots\}$.
$\mathbb{N}$	set of natural numbers $\{1, 2, \ldots\}$.
$\mathbb{R}$	set of real numbers.
$\mathbb{Q}$	set of rational numbers.
$\mathbb{C}$	set of complex numbers.
$\mathbb{R}^{p\times m}$	set of real $p \times m$ matrices.
$\mathbb{B}$	σ-algebra.
$\emptyset$	empty set.
$\subseteq$	subset.
$\subset$	proper subset.
$\cap$	intersection.
$\cup$	union.
$\preceq$	partial or semi-ordering.
$\vee$	disjunction, "or".
$\wedge$	conjunction, "and".
$\circ$	composition.
$1(\cdot)$	identity map.
d/dt	differential operator in an ODE.
$\dot{x}$	short form of the differential dx/dt.
$\partial/\partial t$	partial differential operator.
$\mathcal{N}(\bar{x}, \sigma_x^2)$	normal or Gaussian probability distribution/density function.
$\bar{x}$	mean value.
σ^2	variance.
ρ	Euclidean distance.
$n!$	factorial, $n! = 1 \times 2 \times 3 \times \cdots \times n$.
∞	infinity.
$o(x)$	"litte-o" a negligible quantity that vanish faster than x as x approaches zero.
$\mathcal{O}(x)$	"big-O" terms of the same order of x.

Abbreviations

ADP	adenosine diphosphate.
ATP	adenosine triphosphate.
CME	chemical master equation.
EGF	epidermal growth factor.
ERK	extracellular signal-regulated kinase.
GDP	guanosine diphosphate.
GTP	guanosine triphosphate.
GMA	generalised mass action.

JAK	janus kinase.
LMA	law of mass action.
MAP	mitogen-activated protein.
MAPK	mitogen-activated protein kinase.
MAPKK	mitogen-activated protein kinase kinase.
MEK	MAPK/ERK kinase.
MEKK	MEK kinase.
ODE	ordinary differential equation.
pgf	probability generating function.
mgf	moment generating function.
cgf	cumulant generating function.
CV	coefficient of variation.
Var	variance.
Std	standard deviation.
lim	limes, in the limit.
det	determinant.
w.r.t.	with respect to.
iff	if and only if.
SOS	son of sevenless.
STAT	signal transducers and activators of transcription.
TNF	tumor necrosis factor.
TPA	12-O-tetracecanoyl-phorbol-12-acetate.

Chapter 1

$\mathcal{S}$	system.
O	object(s).
R	relation.
$A \times B$	Cartesian product.
T	time set.
I	index set.
U, Y	input, output objects/spaces.
ϕ	state mapping.
g, h	input, output mapping.
u, y, x	input-, output-, and state-variable/vector.
Ω	sample space of a random variable.
$\mathbb{B}$	σ-algebra.
$P(\cdot)$	probability measure/function.
Prob$\{A\}$	probability of event A.
$\omega \in \Omega$	elementary event.
$w(\omega)$	random variable.
$w_t(\omega)$	stochastic process.
n	number of state variables/ODEs.

m	number of dependent variables.
K_{eq}	equilibrium constant.
K_d	dissociation constant.

Chapter 2

n	number of molecules, state-vector.
Δ	small but not infinitesimal change.
$\#S$	number of molecules.
n_T	total number of molecules.
k	rate constant.
R_μ	reaction channel (irreversible reaction).
M	number of reaction channels.
$\tilde{x}$	steady state.
$\langle S(t) \rangle$	mean or average of the process $S(t)$.
K_m	Michaelis-Menten constant.
V	volume, or velocity.
$V_{\max}$	limiting rate in a kinetic reaction.
S_j	chemical species.
N	number of chemical species.
$[S]$	concentration of S.
S	state (vector) of the system.
R_μ	reaction channel.
c_μ	stochastic reaction constant (stochastic simulation).
a_μ	propensity of reaction R_μ.
a^*	propensity for any of the R_μ to occur.
h_μ	number of distinct combinations of R_μ reactant molecules.
K_μ	molecularity of reaction R_μ.
$l_{\mu j}$	stoichiometric coefficient.
L_μ	number of reactant species.
$\nu_{\mu j}$	change in the population of molecular species S_j in reaction R_μ.
$P(\cdot)$	probability measure.
$F(\cdot)$	cumulative distribution function.
$p_{m,n}$	transition probability.
Π	probability transition matrix.
v_μ	rate of reaction.
$\mathfrak{P}$	probability generating function (pgf).
$\mathfrak{P}'$	derivative of the pgf.
$\mathfrak{M}$	moment generating function (mgf).
$\mathfrak{C}$	cumulant generating function (cgf).

Chapter 4

θ	parameter(s).
n, m	number of dependent, independent variables.
x, X	state variable, state space or fiber.
u, y	input and output variable.
ϕ	state mapping.
h	output mapping.
J	Jacobian matrix.
$H(A, B)$	set of all mappings from A to B.
φ	flow.
$\mathcal{G}$	group.
$\mathcal{C}$	category.
$T_x X$	tangent space to domain X.
TX	tangent bundle.
$\mathbb{M}$	family of models.
$(\mathbb{P}, \pi)$	parametrisation, $\pi \colon \mathbb{P} \to \mathbb{M}$.
$\mathbb{P}$	parameter space, base space.
$(\mathbb{P}, X)$	fiber bundle.
B^A	exponential of maps from A to B.
e_f	evaluation map.

Bibliography

[A+04] M.I. Aladjem et al. Molecular interaction maps–a diagrammatic graphical language for bioregulatory networks. *Sci. STKE*, 222:pe8, 2004. http://www.stke.org.

[AdP02] P.W. Atkins and J. de Paula. *Atkins' Physical Chemistry*. Oxford University Press, seventh edition, 2002.

[AFJS04] D. Angeli, J.E. Ferrell Jr., and E.D. Sontag. Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proc. Natl. Acad. Sci. USA*, 101(7):1822–1827, February 2004.

[AJL+02] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland Science, fourth edition, 2002.

[AKLS05] Michael Andrec, Boris N Kholodenko, Ronald M Levy, and Eduardo Sontag. Inference of signaling and gene regulatory networks by steady-state perturbation experiments: structure and accuracy. *J. Theor. Biol.*, 232(3):427–441, Feb 2005.

[AL01] A.R. Asthagiri and D.A. Lauffenburger. A Computational Study of Feedback Effects on Signal Dynamics in a Mitogen-Activated Protein Kinase (MAPK) Pathway Model. *Biotechnol. Prog.*, 17:227–239, 2001.

[Alo06] Uri Alon. *An Introduction to Systems Biology*. Mathematical and Computational Biology Series. Chapman & Hall/CRC, first edition, 2006.

[AP09] E. August and A Papachristodoulou. A New Computational Tool for Establishing Model Parameter Identifiability. *J. Comput. Biol.*, 16(6):875–885, 2009.

[ARM98] Adam Arkin, John Ross, and Harley H. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected escherichia coli cells. *Genetics*, 149(4):1633–1648, 1998.

[AS72] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions*. Number 55 in Applied Mathematics Series. National Bureau of Standards, tenth edition, 1972.

[AS04] P. Albertos and A. Sala. *Multivariable Control Systems*. Springer, 2004.

[ASK89] L. Acerenza, H.M. Sauro, and H. Kacser. Control analysis of time-dependent metabolic systems. *J. Theor. Biol.*, 137:423–444, 1989.

[BADG08] Laurier Boulianne, Sevin A. Assaad, Michel Dumontier, and Warren Gross. GridCell: a stochastic particle-based biological system simulator. *BMC Syst. Biol.*, 2(1):66, 2008.

[Bat79] E. Batschelet. *Introduction to Mathematics for Life Scientists.* Springer, third edition, 1979.

[BB89] E.J. Borowski and J.M. Borwein. *Collins Dictionary of Mathematics.* Harper Collins Publishers, 1989.

[BB01] J.M. Bower and H. Bolouri. *Computational Modeling of Genetic and Biochemical Networks.* MIT Press, 2001.

[BCBG+85] J.A. Burns, A. Cornish-Bowden, A.K. Groen, R. Heinrich, J.W. Porteus, S.M. Rapoport, T.A. Rapoport, J.W. Stucki, J.M. Tager, R.J.A. Wanders, and H.V. Westerhoff. Control analysis of metabolic systems. *Trends Biochem. Sci.*, 10(1):16, January 1985.

[BCF+07] Massimo Bernaschi, Filippo Castiglione, Alessandra Ferranti, Caius Gavrila, Michele Tinti, and Gianni Cesareni. ProtNet: a tool for stochastic simulations of protein interaction networks dynamics. *BMC Bioinformatics*, 8(Suppl 1):S4, 2007.

[BD01] W.E. Boyce and R.C. DiPrima. *Elementary differential equations and boundary value problems.* John Wiley and Sons, seventh edition, 2001.

[Bec05] M. Beckerman. *Molecular and Cellular Signaling.* Springer-Verlag (Berlin), 2005.

[Bel90] D. Bell. *Mathematics of Linear and Nonlinear Systems.* Oxford University Press, 1990.

[Ber68] L. von Bertalanffy. *General Systems Theory.* George Braziller, 1968.

[Bet01] D. Betounes. *Differential Equations. Theory and Applications with Maple.* Springer, 2001.

[BF00] F.A. Brightman and D.A. Fell. Differential feedback regulation of the MAPK cascade underlies the quantitative differences in EGF and NGF signalling in PC12 cells. *FEBS Lett.*, 482(3):169–174, October 2000.

[BH25] G.E. Briggs and J.B.S. Haldane. A note on the kinetics of enzyme action. *Biochem. J.*, 19:338–339, 1925.

[BI99] U.S. Bhalla and R. Iyengar. Emergent properties of networks of biological signaling pathways. *Science*, 283:381–387, 1999.

[Bla05] F.F. Blackman. Optima and Limiting Factors. *Ann. Bot.*, 19(2):281–296, April 1905.

[BMCM98] J. Benito, C. Martin-Castellanos, and S. Moreno. Regulation of the G_1 phase of the cell cycle by periodic stabilization and degradation of the $p25^{rum1}$ CDK inhibitor. *EMBO J.*, 17(2):4882–497, January 1998.

[BPE00] Otto G. Berg, Johan Paulsson, and Måns Ehrenberg. Fluctuations and Quality of Control in Biological Cells: Zero-Order Ultrasensitivity Reinvestigated. *Biophys. J.*, 79(3):1228–1236, 2000.

[BPM82] R.D. Bliss, P.R. Painter, and A.G. Marr. Role of Feedback Inhibition in Stabilizing the Classical Operon. *J. Theor. Biol.*, 97:177–193, 1982.

[Bra96] M. D. Brand. Top down metabolic control analysis. *J. Theor. Biol.*, 182(3):351–360, Oct 1996.

[Bro99] T.A. Brown. *Genomes*. βIOS Scientific Publishers, 1999.

[BS03] K.S. Brown and J.P. Sethna. Statistical mechanical approaches to models with many poorly known parameters. *Phys. Rev. E*, 68:021904, 2003.

[BSDdPM00] M.A. Blanco, A. Sanchez-Daz, J.M. de Prada, and S. Moreno. $APC^{s}te9/srw1$ promotes degradation of mitotic cyclins in G_1 and is inhibited by cdc2 phosphorylation. *EMBO J.*, 19(15):3945–3955, August 2000.

[CB95a] A. Cornish-Bowden. *Fundamentals of Enzyme Kinetics*. Portland Press, 1995.

[CB95b] A. Cornish-Bowden. Metabolic Control Analysis in Theory and Practice. *Adv. Mol. Cell Biol.*, 11:21–64, 1995.

[CB99] A. Cornish-Bowden. *Basic Mathematics for Biochemists*. Oxford University Press, second edition, 1999.

[CB04] A. Cornish-Bowden. *Fundamentals of Enzyme Kinetics*. Portland Press, third edition, 2004.

[CBN95] J. Correa-Bordes and P. Nurse. $p25^{r}um1$ orders S phase and mitosis by acting as an inhibitor of the $p34^{c}dc2$ mitotic kinase. *Cell*, 83(6):1001–1009, December 1995.

[CCWW05] Kwang-Hyun Cho, Sang-Mok Choo, Peter Wellstead, and Olaf Wolkenhauer. A unified framework for unraveling the functional interaction structure of a biomolecular network based on stimulus-response experimental data. *FEBS Lett.*, 579(20):4520–4528, Aug 2005.

[Dan08] B.C. Daniels. Sloppiness, Robustness and Evolvability in Systems Biology. *Curr. Opin. Biotechnol.*, 19(4):389–395, 2008.

[Dav01] E.H. Davidson. *Genomic Regulatory Systems*. Academic Press, 2001.

[DGKS04] A. Dhooge, W. Govaerts, Y.A. Kuznetsov, and B. Sautois. Limit cycles and their bifurcations in MatCont, 2004.

[DGKS06] A. Dhooge, W. Govaerts, YA Kuznetsov, and B. Sautois. Matcont: A Matlab Package for Dynamical Systems with Applications to Neural Activity, 2006.

[dlFBM02] Alberto de la Fuente, Paul Brazhnik, and Pedro Mendes. Linking the genes: inferring quantitative gene networks from microarray data. *Trends Genet.*, 18(8):395–398, Aug 2002.

[Dow01] J. Downward. The ins and outs of signalling. *Nature*, 411(6839):759–762, June 2001.

[DTB97] K. Dutton, S. Thompson, and B. Barraclough. *The Art of Control Engineering*. Addison-Wesley, 1997.

[ECG80] G. Eason, C.W. Coles, and G. Gettinby. *Mathematics and Statistics for the Bio-Sciences*. Pearson Education, 1980.

[EE03] Johan Elf and Måns Ehrenberg. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Res.*, 13(11):2475–2484, November 2003.

[Elf04] Johan Elf. *Intracellular Flows and Fluctuations*. PhD thesis, Uppsala University, Teknisk-naturvetenskapliga vetenskapsomrdet, Biology, Department of Cell and Molecular Biology, September 2004.

[ELS01] J. Elf, P. Lötstedt, and P. Sjöberg. Problems of high dimension in molecular biology. In *17th GAMM-Seminar*, pages 1–10, Leipzig, 2001.

[ELSS02] M.B. Elowitz, A.J. Levine, E.D. Siggia, and P.S. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, August 2002.

[EP01] C.H. Edwards and D.E. Penney. *Differential Equations and Linear Algebra*. Prentice Hall, 2001.

[Fel97] D. Fell. *Understanding the Control of Metabolism*. Portland Press, 1997.

[FFE+06] M. Farina, R. Findeisen, Bullinger E., Bittanti S, Allgöwer F., and P. Wellstead. Results towards identifiability properties of biochemical reaction networks. In *Proc. 45th IEEE Conf. Decision Contr CDC'06*, 2006.

[FHC+04] X. Feng, S. Hooshangi, D. Chen, G. Li, R. Weiss, and H. Rabitz. Optimizing Genetic Circuits by Global Sensitivity Analysis. *Biophys. J.*, 87(4):2195–2202, 2004.

[FJ96] J.E. Ferrell Jr. Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. *TIBS*, 21(12):460–466, December 1996.

[FMWT02] C.P. Fall, E.S. Marland, J.M. Wagner, and J.J. Tyson. *Computational Cell Biology*. Springer Verlag (Berlin), 3rd edition, 2002.

[FN95] D. Fisher and P. Nurse. Cyclins of the fission yeast Schizosaccharomyces pombe. *Semn. Cell Biol.*, 6(2):73–78, April 1995.

[FPEN05] G. Franklin, J.D. Powell, and A. Emami-Naeini. *Feedback Control of Dynamic Systems*. Prentice Hall, 5th edition, 2005.

[Fre97] D. Freedman. From association to causation: some remarks on the history of statistics. *Statist. Sci.*, 14(3):243–258, 1997.

[Gar85] C.W. Gardiner. *Handbook of Stochastic Models*. Springer, second edition, 1985.

[GB00] M.A. Gibson and J. Bruck. Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *J. Phys. Chem.*, A104:1876–1888, 2000.

[GF06] T. Gross and U. Feudel. Generalized models as a universal approach to the analysis of nonlinear dynamical systems. *Phys. Rev. E*, 73:016205, 2006.

[GGD05] K. Gadkar, R. Gunawan, and F. Doyle. Iterative approach to model identification of biological networks. *BMC Bioinformatics*, 6(1):155, 2005.

[GH83] J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer, 1983.

[Gie88] C. Giersch. Control analysis of matabolic networks: 1. Homogeneous functions and the summation theorems for control coefficients. *Eur. J. Biochem.*, 174:509–513, 1988.

[Gil76] D.T. Gillespie. General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *J. Comput. Phys.*, 22:403–434, 1976.

[Gil77] D.T. Gillespie. Exact Stochastic Simulation of Coupled Chemical Reactions. *J. Phys. Chem.*, 81(25):2340–2361, 1977.

[Gil92] D.T. Gillespie. *Markov Processes*. Academic Press, 1992.

[Gil96] D.T. Gillespie. The multivariate Langevin and Fokker-Planck equations. *American Journal of Physics*, 64(10):1246–1257, 1996.

[Gil01] D.T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting system. *J. Chem. Phys.*, 115(4):1716–1733, July 2001.

[GKJ81] A. Goldbeter and D.E. Koshland Jr. An amplified sensitivity arising from covalent modification in biological systems. *Proc. Natl. Acad. Sci. USA*, 78(11):6840–6844, November 1981.

[Gol96] A. Goldbeter. *Biochemical Oscillations and Cellular Rythms*. Cambridge University Press, 1996.

[Gom03] B. Gomperts. *Signal Transductions*. Academic Press, 2003.

[Gri68] J.S. Griffith. Mathematics of Cellular Control Processes. *J. Theor. Biol.*, 20:202–208, 1968.

[GS94] P. Gray and S. Scott. *Chemical Oscillations and Instabilities: Nonlinear Chemical Kinetics*. Oxford University Press, 1994.

[GSB+07] S. Grimbs, J. Selbig, S. Bulik, H.-G. Holzhütter, and R. Steuer. The stability and robustness of metabolic states: identifying stabilizing sites in metabolic networks. *Mol. Syst. Biol.*, 3:146, 2007.

[GUV07] Carlos A. Gmez-Uribe and George C. Verghese. Mass fluctuation kinetics: capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations. *J. Chem. Phys.*, 126(2):024109, January 2007.

[GWC+07] R.N. Gutenkunst, J.J. Waterfall, F.P. Casey, K.S. Brown, C.R. Myers, and J.P. Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.*, 3(10):1871–1878, 2007.

[GWW+82] A.K. Groen, R.J.A. Wanders, H.V. Westerhoff, R. van der Meer, and J.M. Tager. Quantification of the Contribution of Various Steps to the Control of Mitochondrial Respiration. *J. Biol. Chem.*, 257(6):2754–2757, March 1982.

[Han97] J.T. Hancock. *Cell Signalling*. Prentice Hall, 1997.

[Har01] F.M. Harold. *The Way of the Cell*. Oxford University Press, 2001.

[HBB+05] J.J. Hornberg, F.J. Bruggeman, B. Binder, C.R. Geest, A.J.M.B. de Vaate, J. Lankelma, R. Heinrich, and H.V. Westerhoff. Principles behind the multifarious control of signal transduction. ERK phosphorylation and kinase/phosphatase control. *FEBS J.*, 272(1):244–258, January 2005.

[HCB91] J.-H. S. Hofmeyr and A. Cornish-Bowden. Quantitative assessment of regulation in metabolic systems. *Eur. J. Biochem.*, 200:223–236, 1991.

[HCB97] J.-H. S. Hofmeyr and A. Cornish-Bowden. The reversible Hill equation: how to incorporate cooperative enzymes into metabolic models. *Comput. Appl. Biosci.*, 13(4):377–385, 1997.

[HCBR93] Jan-Hendrik S. Hofmeyr, A. Corish-Bowden, and JM. Rohwer. Taking enzyme kinetics out of control; putting control into regulation. *Eur. J. Biochem.*, 212:833–837, 1993.

[Hel05] E.J.M Helmreich. *The Biochemistry of Cell Signalling*. Oxford University Press, 2005.

[HFJ96] Chi-Ying F. Huang and J. E. Ferrell Jr. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc. Natl. Acad. Sci. USA*, 93(19):10078–10083, September 1996.

[HFWN94] J. Hayles, D. Fisher, A. Woollard, and P Nurse. Temporal order of S phase and mitosis in fission yeast is determined by the state of the p34cdc2-mitotic B cyclin complex. *Cell*, 78(5):813–822, September 1994.

[HHLM99] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray. From molecular to modular cell biology. *Nature*, 402(6761 Suppl):C47–C52, Dec 1999.

[HJ04] F. Hayot and C. Jayaprakash. The linear noise approximation for molecular fluctuations within cells. *Physical Biology*, 1(4):205–210, December 2004.

[HMIC01] J. Hasti, D. McMillen, F. Isaacs, and J.J. Collins. Computational studies of gene regulatory networks: *in numero* molecular biology. *Nature Reviews*, 2:268–279, April 2001.

[HNR02] R. Heinrich, B.G. Neel, and T.A. Rapoport. Mathematical Models of Protein Kinase Signal Transduction. *Mol. Cell*, 9(5):957–970, May 2002.

[Hou07] James House. *Principles of Chemical Kinetics*. Academic Press, San Diego, California, 2007.

[HR73] R. Heinrich and T.A. Rapoport. Linear theory of enzymatic chains; its application for the analysis of the crossover theorem and of the glycolysis of human erythrocytes. *Acta Biol. Med. Germ.*, 31:479–494, 1973.

[HR74] R. Heinrich and T.A. Rapoport. A Linear Steady-State Treatment of Enzymatic Chains. General Properties, Control and Effector Strength. *Eur. J. Biochem.*, 42(1):89–95, January 1974.

[HR75] r. Heinrich and T.A. Rapoport. Mathematical Analysis of multienzyme systems. II. Steady-state and transient control. *BioSystems*, 7:130–136, 1975.

[HR91] R. Heinrich and C. Reder. Metabolic control analysis of relaxation processes. *J. Theor. Biol.*, 151:343–350, 1991.

[HS96] R. Heinrich and S. Schuster. *The Regulation of Cellular Systems.* Chapmann and Hall, 1996.

[HSD04] M.W. Hirsch, S. Smale, and R.L. Devaney. *Differential Equations, Dynamical Systems and an Introduction to Chaos.* Elsevier - Academic Press, 2004.

[Hun09] P. Hunter. Robust yet flexible. *EMBO Reports*, 10:949–952, 2009.

[HW01] J.-H.S. Hofmeyr and H.V. Westerhoff. Building the Cellular Puzzle — Control in Multi-level Reaction Networks. *J. Theor. Biol.*, 208(3):261–285, February 2001.

[IS03] B.P. Ingalls and H.M. Sauro. Sensitivity analysis of stoichiometric networks: an extension of metabolic control analysis to non-steady state trajectories. *J. Theor. Biol.*, 222(1):23–36, May 2003.

[Isi89] A. Isidori. *Nonlinear Control Systems.* Springer-Verlag, 1989.

[Jef93] A. Jeffrey. *Linear Algebra and Ordinary Differential Equations.* CRC Press, 1993.

[JM61] F. Jacob and F. Monod. Genetic regulatory mechanisms in the synthesis of protein. *J. Mol. Biol.*, 3:318, 1961.

[JS07] D.W. Jordan and P. Smith. *Nonlinear Ordinary Differential Equations.* Oxford University Press, fourth edition, 2007.

[K+02] B.N. Kholodenko et al. Untangling the wires: A strategy to trace functional interactions in signaling and gene networks. *Proc. Natl. Acad. Sci. USA*, 99(20):12841–12846, October 2002.

[KAWP05] K.W. Kohn, M.I. Aladjem, J.N. Weinstein, and Y. Pommier. Molecular Interaction Maps of Bioregulatory Networks: A General Rubric for Systems Biology. *Mol. Biol. Cell*, 17(1):1–13, 2005.

[KB73] H. Kacser and J.A. Burns. The control of flux. *Symp. Soc. Exp. Biol.*, 27:65–104, 1973.

[KC82] M.C. Kohn and E. Chiang. Metabolic network sensitivity analysis. *J. Theor. Biol.*, 98:109–126, 1982.

[KDMH99] B.N. Kholodenko, O.V. Demin, G. Moehren, and J.B. Hoek. Quantification of Short Term Signaling by the Epidermal Growth Factor Receptor. *J. Biol. Chem.*, 274(42):30169–30181, October 1999.

[KEBC05] Mads Kaern, Timothy C. Elston, William J. Blake, and James J. Collins. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.*, 6(6):451–464, June 2005.

[KFA69] R.E. Kalman, P.L. Falb, and M.A. Arbib. *Topics in Mathematical Systems Theory*. McGraw-Hill, 1969.

[KFMO05] H. Kitano, A. Funahashi, Y. Matsuoka, and K. Oda. Using process diagrams for the graphical representation of biological networks. *Nat. Biotechnol.*, 23:961–966, 2005.

[KG95] D. Kaplan and L. Glass. *Understanding Nonlinear Dynamics*. Springer Verlag, 1995.

[KH04] R. Krüger and R. Heinrich. Model Reduction and Analysis of Robustness for the Wnt/β-Catenin Signal Transduction Pathway. *Genone Inform.*, 15:138–148, 2004.

[KHK+05] E. Klipp, R. Herwig, A. Kowald, C. Wierling, and H. Lehrach. *Systems Biology in Practice*. John Wiley & Sons, 2005.

[Kho00] B.N. Kholodenko. Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur. J. Biochem.*, 267(6):1583–1588, March 2000.

[Kie02] A.M. Kierzek. STOCKS STOchastic Kinetic Simulations of biochemical systems with Gillespie algorithm. *Bioinformatics*, 18:470–481, 2002.

[Kit02] H. Kitano. Standards for modeling. *Nat. Biotechnol.*, 20(4):337, 2002.

[Kit07] H. Kitano. Towards a Theory of Biological Robustness. *Mol. Syst. Biol.*, 3:137, 2007.

[Kli91] G.J. Klir. *Facets of Systems Theory*. Plenum Press, 1991.

[KMS98] K. Kitamura, H. Maekawa, and C. Shimoda. Fission Yeast Ste9, a Homolog of Hct1/Cdh1 and Fizzy-related, Is a Novel Negative Regulator of Cell Cycle Progression during G_1-Phase. *Mol. Biol. Cell*, 9(5):1065–1080, May 1998.

[Koh99] K.W. Kohn. Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Mol. Biol. Cell*, 10(8):2703–2734, August 1999.

[Koh01] K.W. Kohn. Molecular interaction maps as information organizers and simulation guides. *Chaos*, 11(1):84–97, 2001.

[Kol00] W. Kolch. Meaningful relationships: the regulation of the Ras/Raf/MEK/ERK pathway by protein interactions. *Biochem. J.*, 351:289–305, 2000.

[Kre93] E. Kreyszig. *Advanced Engineering Mathematics*. John Wiley & Sons, 1993.

[KS01] J. Keener and J. Sneyd. *Mathematical Physiology*. Springer Verlag, 2001.

[KS03] H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, second edition, 2003.

[KSA90] H. Kacser, H.M. Sauro, and L. Acerenza. Enzyme-enzyme interactions and control analysis. 1. The case of non-additivity : monomer-oligomer associations. *Eur. J. Biochem.*, 187(3):481–491, February 1990.

[KST07] M Kaufman, C. Soule, and R. Thomas. A new necessary condition on interaction graphs for multistationarity. *J. Theor. Biol.*, 248(4):675–685, October 2007.

[Lan04] A.D. Lander. A Calculus of Purpose. *PLoS Biol.*, 2:0712–0714, June 2004.

[Lau00] D. A. Lauffenburger. Cell signaling pathways as control modules: complexity for simplicity? *Proc Natl Acad Sci U S A*, 97(10):5031–5033, May 2000.

[Laz02] Y. Lazebnik. Can a biologist fix a radio? - Or what I learned while studying apoptosis. *Cancer Cell*, 2:179–182, September 2002.

[LC03] S. Lorkowski and P. Cullen. *Analysing Gene Expression*. Wiley-VCH, 2003.

[LD93] J.C. Liao and J. Delgado. Advances in Metabolic Control Analysis. *Biotechnol. Prog.*, 9(3):221–233, 1993.

[LFB89] R.A. Laskey, M.P. Fairman, and J.J. Blow. S Phase of the Cell Cycle. *Science*, 246(4930):609–614, November 1989.

[LFS98] S. Liang, S. Fuhrman, and R. Somogyi. Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pac Symp Biocomput*, pages 18–29, 1998.

[LKM07] Joseph Levine, Hao Yuan Kueh, and Leonid Mirny. Intrinsic fluctuations, robustness, and tunability in signaling cycles. *Biophys. J.*, 92(12):4473–4481, 2007.

[LL93] D.A. Lauffenburger and J.J. Linderman. *Receptors: Models for binding, trafficing, and signalling*. Oxford University Press, 1993.

[Lot20] Alfred J. Lotka. Undamped oscillations derived from the law of mass action. *Journal of the American Chemical Society*, 42(8):1595–1599, 1920.

[LQ03] G. Li and H. Qian. Sensitivity and Specificity Amplification in Signal Transduction. *Cell Biochem. Biophys.*, 39:45–59, 2003.

[Lüt91] H. Lütkepohl. *Introduction to Multiple Time Series Analysis*. Prentice Hall, second edition, 1991.

[LWB+91] K. Lundgren, N. Walworth, R. Booher, M Dembski, M. Kirschner, and D. Beach. mik1 and wee1 cooperate in the inhibitory tyrosine phosphorylation of cdc2. *Cell*, 64(6):1111–1122, March 1991.

[MA97] H.H. McAdams and A.P. Arkin. Stochastic Mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA*, 94:814–819, 1997.

[Mag97] B. Magee. *Confessions of a Philosopher*. Phoenix, Orion Books, 1997.

[Mat] The MathWorks. Matlab.

[MCLM96] C. Martin-Castellanos, K. Labib, and S. Moreno. B-type cyclins regulate G1 progression in fission yeast in opposition to the p25$^{r}um1$ cdk inhibitor. *EMBO J.*, 15(4):839–849, February 1996.

[Mes68] M.D. Mesarović. Systems Theory and Biology - View of a Theoretician. In M.D. Mesarović et al., editors, *System Theory and Biology*, volume 351, pages 59–87. Springer-Verlag, 1968.

[MHTSK90] E. Melendez-Hevia, N.V. Torres, J. Sicilia, and H. Kacser. Control analysis of transition times in metabolic systems. *Biochem. J.*, 265(1):195–202, January 1990.

[MK89] J.R. McIntosh and M.P. Koonce. Mitosis. *Science*, 246(4930):622–628, November 1989.

[MK06] B. Munsky and M. Khammash. A reduced model solution for the chemical master equation arising in stochastic analyses of biological networks. In *Decision and Control, 2006 45th IEEE Conference on*, pages 25–30, 2006.

[MM13] L. Michaelis and M.L. Menten. Die Kinetik der Invertinwirkung. *Biochem. Z.*, 19:333–369, 1913.

[MMD+02] Gisela Moehren, Nick Markevich, Oleg Demin, Anatoly Kiyatkin, Igor Goryanin, Jan B Hoek, and Boris N Kholodenko. Temperature dependence of the epidermal growth factor receptor signaling network can be accounted for by a kinetic model. *Biochemistry*, 41(1):306–320, Jan 2002.

[MMP00] M.T. Madigan, J.M. Martinko, and J. Parker. *Brock: Biology of Microorganisms*. Prentice Hall, ninth edition, 2000.

[MN94] S. Moreno and P. Nurse. Regulation of progression through the Gl phase of the cell cycle by the rum1$^+$ gene. *Nature*, 367(6460):236–242, January 1994.

[Mor08] Robert G. Mortimer. *Physical Chemistry*. Elsevier Academic Press, San Diego, California, 3rd edition, 2008.

[MR92] J.B.A. Millar and P. Russell. The cdc25 M-phase inducer: An unconventional protein phosphatase. *Cell*, 68(3):407–410, February 1992.

[MR07] M. Mincheva and M. Roussel. Graph-theoretic methods for the analysis of chemical and biochemical networks. I. Multistability and oscillations in ordinary differential equation models. *J. Math. Biol.*, 55(1):61–86, July 2007.

[MS99] D.D. Mooney and R.J. Swift. *A Course in Mathematical Modeling*. The Mathematical Association of America, 1999.

[MT75] M. D. Mesarović and Y. Takahara. *General Systems Theory: Mathematical Foundations*. Academic Press., 1975.

[Mur02] J.D. Murray. *Mathematical Biology*. Springer Verlag, third edition, 2002.

[MVHA99] C.K. Mathews, K.E. Van Holde, and K.G. Ahern. *Biochemistry*. Benjamin-Cummings, third edition, 1999.

[Nas96a] K. Nasmyth. At the heart of the budding yeast cell cycle. *Trends Genet.*, 12(10):405–412, October 1996.

[Nas96b] K. Nasmyth. Putting the Cell Cycle in Order. *Science*, 274(5293):1643–1645, October 1996.

[NCNB+98] B. Novak, A. Csikasz-Nagy, Gyorffy B., K. Nasmyth, and J.J. Tyson. Model scenarios for evolution of the eukaryotic cell cycle. *Philos. T. Roy. Soc. B*, 353:2063–2076, 1998.

[NPCT01] B. Novak, Z. Pataki, A. Ciliberto, and J.J. Tyson. Mathematical model of the cell division cycle of fission yeast. *Chaos*, 11(1):277–286, March 2001.

[Nur75] P. Nurse. Genetic control of cell size at cell division in yeast. *Nature*, 256(5518):547–551, August 1975.

[Nur90] P. Nurse. Universal control mechanism regulating onset of M-phase. *Nature*, 344(6266):503–508, April 1990.

[Nv90] H. Nijmeijer and A.J. van der Schaft. *Nonlinear Dynamical Control Systems*. Springer, 1990.

[ol09a] on line. AUTO: Software force Continuation and Bifurcation Problems in Ordinary Differential Equations, 2009.

[ol09b] on line. CANDYS/QA, 2009.

[ol09c] on line. CANDYS/QA, 2009.

[OMFK05] K. Oda, M. Matsuoka, A. Funahashi, and H. Kitano. A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol. Syst. Biol.*, 10:1038, 2005.

[P+01] W.K. Purves et al. *Life: The Science of Biology.* Sinauer Associates, sixth edition, 2001.

[Pah08] Jrgen Pahle. *Stochastic Simulation and Analysis Of Biochemical Networks.* PhD thesis, Humboldt-Universitt zu Berlin, Fach Biophysik, 2008.

[Pal06] B.O. Palsson. *Systems Biology.* Cambridge University Press, 2006.

[Pau05] J. Paulsson. Models of stochastic gene expression. *Phys. Life Rev.*, 2:157–75, June 2005.

[PE06] Johan Paulsson and Johan Elf. Stochastic Modeling of Intracellular Kinetics. In Zoltan Szallasi, Jrg Stelling, and Vipul Periwal, editors, *System Modeling in Cellular Biology*, pages 149–176. The MIT Press, Cambridge, Massachusetts, 2006.

[Per98] L. Perko. *Differential equations and Dynamical Systems.* Springer Verlag, 2nd edition, 1998.

[Pho97] D. Phoenix. *Introductory Mathematics for the Life Sciences.* Taylor and Francis, 1997.

[Pol01] G.H. Pollack. *Cells, Gels and the Engines of Life.* Ebner and Sons Pub, 2001.

[PP01] Athanasios Papoulis and S. Unnikrishna Pillai. *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, fourth edition, 2001.

[PTVF93] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C – The Art of Scientific Computing.* Cambridge University Press, second edition, 1993.

[QK02] G.P. Quinn and M.J. Keough. *Experimental Design and Data Analysis for Biologists.* Cambridge University Press, 2002.

[RBFS09] A. Rizk, G. Batt, F. Fages, and S. Soliman. A general computational method for robustness analysis with application to synthetic gene networks. *Bioinformatics*, 25(12):i169–i178, 2009.

[Red88] C. Reder. Metabolic Control Theory: A structural approach. *J. Theor. Biol.*, 135:175–201, 1988.

[RLP07] Andre S Ribeiro and Jason Lloyd-Price. SGN Sim, a stochastic genetic networks simulator. *Bioinformatics*, 23(6):777–779, March 2007.

[Rn53] Alfrd Rnyi. On the theory of order statistics. *Acta Mathematica Hungarica*, 4(3):191–231, September 1953.

[RN87] P. Russell and P. Nurse. Negative regulation of mitosis by wee1+, a gene encoding a protein kinase homolog. *Cell*, 49(4):559–567, May 1987.

[RO05] Jonathan M. Raser and Erin K. O'Shea. Noise in gene expression: Origins, consequences, and control. *Science*, 309(5743):2010–2013, September 2005.

[RS02] A. Regev and E. Shapiro. Cells as Computation. *Nature*, 419:343, 2002.

[Rub75] S.I. Rubinow. *Introduction to Mathematical Biology.* John Wiley, 1975.

[RWA02] C.V. Rao, D.M. Wolf, and A.P. Arkin. Control, exploitation and tolerance of intracelluar noise. *Nature*, 420(6912):231–237, November 2002.

[Sas99] S. Sastry. *Nonlinear Systems.* Springer Verlag, 1999.

[Sav69a] Michael A. Savageau. Biochemical systems analysis: I. Some mathematical properties of the rate law for the component enzymatic reactions. *J. Theor. Biol.*, 25:365–369, 1969.

[Sav69b] Michael A. Savageau. Biochemical systems analysis: II. Steady state solutions for an n- poll system using a power-law approximation. *J. Theor. Biol.*, 25:370–379, 1969.

[Sav70] M. A. Savageau. Biochemical systems analysis: III. Dynamic solutions using a power-law approximation. *J. Theor. Biol.*, 26:215–226, 1970.

[SBB98] P. Smolen, D. A. Baxter, and J. H. Byrne. Frequenz selectivity, multistablilty, and oscillations emerge from models of genetic regulatory systems. *Am. J. Physiol.*, 274:531–542, 1998.

[SBT02] L. Stryer, J.M. Berg, and J.L. Tymoczko. *Biochemistry.* W H Freeman, 2002.

[Sch72] F. Schlögl. Chemical Reaction Models for Non-Equilibrium Phase Transitions. *Z. Physik*, 253(2):147–161, April 1972.

[SCS00] A. Saltelli, K. Chan, and E.M. Scott. *Sensitivity Analysis.* Wiley, 2000.

[SDPS73] K. Schwetlik, H. Dunken, G. Pretzschner, and H.-J. Scherzer, K. Tiller. *Chemische Kinetik*, volume LB 6 of *Lehrwerk Chemie*. VEB Deutscher Verlag für Grundstoffindustrie, Leipzig, 1973.

[Seg93] I.H. Segel. *Enzyme Kinetics*. John Wiley and Sons, 1993.

[SF89] J.R. Small and D.A. Fell. The Matrix Method of Metabolic Contro Analysis: Its Validity for Complex Pathway Structures. *J. Theor. Biol.*, 136:181–197, 1989.

[SHY+02] S. Y. Shvartsman, M. P. Hagan, A. Yacoub, P. Dent, H. S. Wiley, and D. A. Lauffenburger. Autocrine loops with positive feedback enable context-dependent cell signaling. *Am J Physiol Cell Physiol*, 282(3):C545–C559, Mar 2002.

[SJ04] H. Schmidt and E.W. Jacobsen. Identifying Feddback Mechanisms behind Complex Cell Behavior. *IEEE Contr. Syst. Mag.*, pages 91–102, August 2004.

[SK90] H.M. Sauro and H. Kacser. Enzyme-enzyme interactions and control analysis. 2. The case of non-independence : heterologous associations. *Eur. J. Biochem.*, 187(3):493–500, February 1990.

[SK04] E. Sontag, E. Kiyatkin and B.N. Kholodenko. Inferring dynamic architecture of cellular networks using time series of gene expression, protein and metabolite data. *Bioinformatics*, 20:1877–1886, 2004.

[SKK04] Eduardo Sontag, Anatoly Kiyatkin, and Boris N Kholodenko. Inferring dynamic architecture of cellular networks using time series of gene expression, protein and metabolite data. *Bioinformatics*, 20(12):1877–1886, Aug 2004.

[SM97] S. Schnell and C. Mendoza. Closed Form Solution for Time-dependent Enzyme Kinetics. *J. Theor. Biol.*, 187(2):207–212, July 1997.

[SMDC08] H. Schmidt, M.F. Madsen, S. Dano, and G. Cedersund. Complexity reduction of biochemical rate expressions. *Bioinformatics*, 24(6):848–854, 2008.

[SMT+03] I. Swameye, T.G. Müller, J. Timmer, O. Sandra, and U. Klingmüller. Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by data-based modeling. *Proc. Natl. Acad. Sci. USA*, 100(3):1028–1033, February 2003.

[SN96] B. Stern and P. Nurse. A quantitative model for the cdc2 control of S phase and mitosis in fission yeast. *Trends Genet.*, 12(9):345–350, September 1996.

[SNM99] A. Sveiczer, B. Novak, and J.M. Mitchison. Mitotic control in the absence of cdc25 mitotic inducer in fission yeast. *J. Cell Sci.*, 112(7):1085–1092, April 1999.

[Son98] E.D. Sontag. *Mathematical Control Theory*. Springer, 1998.

[SR94] R.R. Sokal and F.J. Rohlf. *Biometry*. Palgrave Macmillan, third edition, 1994.

[SSF87] H.M. Sauro, J.R. Small, and D.A. Fell. Metabolic control and its analysis - Extensions to the theory and matrix method. *Eur. J. Biochem.*, 165:215–221, 1987.

[SSS+04] J. Stelling, U. Sauer, Z. Szallasi, F.J. III Doyle, and J. Doyle. Robustness of Cellular Functions. *Cell*, 118(6):675–685, September 2004.

[Ste04] P. Sterling. *Allostasis, Homeostasis and the Cost of Physiological Adaptation*, chapter Principles of Allostasis: Optimal Design, Predictive Regulation, Pathophysiology, and Rational Therapeutics, pages 17–64. Cambridge University Press, 2004.

[Str00a] S.H. Strogatz. From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D*, 143:1–20, 2000.

[Str00b] S.H. Strogatz. *Nonlinear Dynamics and Chaos*. Westview Press, 2000.

[STSB06] R. Steuer, Gross T., J. Selbig, and B. Blasius. Structural kinetic modeling of metabolic networks. *Proc. Natl. Acad. Sci. USA*, 103(32):11868–11873, August 2006.

[SVK07] W.K. Smits, J.W. Veening, and O. Kuipers. *Bacterial Physiology. A Molecular Approach*, chapter Phenotypic Variation and Bistable Switching in Bacteria. Springer, 2007.

[TB04] A. Tözeren and S.W. Byers. *New Biology for Engineers and Computer Scientists*. Pearson Prentice Hall, 2004.

[TCN03] J.J. Tyson, K.C. Chen, and B. Novak. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.*, 15(2):221–231, April 2003.

[TD90] R. Thomas and R. D'Ari. *Biological Feedback*. CRC Press, 1990.

[Tho94] R. Thomas. The role of feedback circuits: positive feedback circuits are a necessary condition for positive real eigenvalues of the Jacobian matrix. *Berichte der Bunsen-Gesellschaft*, 98(9):1148–1151, 1994.

[TJD05] Yi Tao, Yuting Jia, and T. Gregory Dewey. Stochastic fluctuations in gene expression far from equilibrium: Omega expansion and linear noise approximation. *J. Chem. Phys.*, 122(12):124108, March 2005.

[232] The Mathworks (TM). Matlab 2009b, 2009.

[TN01] J.J. Tyson and B. Novak. Regulation of the Eucaryotic Cell Cycle: Molecular Antagonism, Hysteresis, and Irreversible Transitions. *J. Theor. Biol.*, 210:249–263, 2001.

[TO78] J.J. Tyson and H.G. Othmer. The dynamics of feedback control circuits in biochemical pathways. *Prog. Theor. Biol.*, 5:1–62, 1978.

[Ton90] H. Tong. *Non-linear Time Series*. Oxford University Press, 1990.

[TSB04] T.E. Turner, S. Schnell, and K. Burrage. Stochastic approaches for modelling in vivo reactions. *Comput. Biol. Chem.*, 28(3):165–178, July 2004.

[Tys75] J.J. Tyson. Classification of instabilities in chemical reaction systems. *J. Chem. Phys.*, 62(3):1010–1015, February 1975.

[Tys79] J.J. Tyson. Periodic Enzyme Synthesis: Reconsideration of the Theory of Oscillatory Repression. *J. Theor. Biol.*, 80:27–38, 1979.

[Tys83] J.J. Tyson. Periodic Enzyme Synthesis and Oscillatory Repression: Why is the period of oscillation close to the cell cycle time? *J. Theor. Biol.*, 103:313–328, 1983.

[Tys01] J.J. Tyson. Network Dynamics and Cell Physiology. *Nat. Rev. Mol. Cell Biol.*, 2:908–916, 2001.

[UIYS06] Tomohiro Ushikubo, Wataru Inoue, Mitsumasa Yoda, and Masaki Sasai. Testing the transition state theory in stochastic dynamics of a genetic switch. *Chem. Phys. Lett.*, 430(1-3):139–143, October 2006.

[UW09] Mukhtar Ullah and Olaf Wolkenhauer. Stochastic approaches in systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 9999(9999), 2009.

[vDMMO00] G. von Dassow, E. Meir, E. M. Munro, and G. M. Odell. The segment polarity network is a robust developmental module. *Nature*, 406(6792):188–192, Jul 2000.

[vGK01] C. van Gend and U. Kummer. STODE - Automatic Stochastic Simulation of Systems Described by Differential Equations. In Yi and Hucka, editors, *Proceedings of the 2nd International Conference on Systems Biology*, pages 326–333, Pasadena, 2001. Omnipress.

[VH02] Diana Visser and Joseph J Heijnen. The mathematics of metabolic control analysis revisited. *Metab. Eng.*, 4(2):114–123, Apr 2002.

[vK92] N.G. van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland, 1992.

[Voi00] E.O. Voit. *Computational Analysis of Biochemical Systems*. Cambrigde University Press, 2000.

[Vol26] Vito Volterra. Fluctuations in the abundance of a species considered mathematically. *Nature*, 118:558–560, 1926.

[VR94] S. Vajda and H. Rabitz. Identifiability and Distinguishability of General Reaction Systems. *J. Phys. Chem.*, 98(20):5265–5271, May 1994.

[VV04] D. Voet and J.G. Voet. *Biochemistry*. John Wiley and Sons, 3rd edition, 2004.

[WC84] H.V. Westerhoff and Y.-D. Chen. How do enzyme activities control metabolite concentrations? An additional theorem in the theory of metabolic control. *Eur. J. Biochem.*, 142(2):425–430, 1984.

[WGW84] H.V. Westerhoff, A.K. Groen, and R.J.A. Wanders. Modern theories of metabolic control and their applications. *Bioscience Rep.*, 4(1):1–22, January 1984.

[WH00] J. Wolf and R. Heinrich. Effect of cellular interaction on glycolytic oscillations in yeast: a theoretical investigation. *Biochem. J.*, 345:321–334, 2000.

[Wig03] S. Wiggins. *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. Springer Verlag, 2003.

[Wil06] Darren James Wilkinson. *Stochastic Modelling for Systems Biology*. Mathematical & Computational Biology. Chapman & Hall/CRC, London, April 2006.

[Wol01] O. Wolkenhauer. Systems Biology: The reincarnation of systems theory applied in biology? *Brief. Bioinform.*, 2(3):258–270, 2001.

[WP96] E. Walter and L. Pronzato. On the identifiability and distinguishability of nonlinear parametric models. *Math. Comput. Simul.*, 42(2-3):125–134, 1996.

[WP97] E. Walter and L. Pronzato. *Identification of parametric models from experimental data*. Springer, 1997.

[Wri04] Margaret Robson Wright. *Introduction to Chemical Kinetics*. Wiley-Interscience, West Sussex, 2004.

[WUKC04] O. Wolkenhauer, M. Ullah, W. Kolch, and K.-H. Cho. Modelling and Simulation of IntraCellular Dynamics: Choosing an Appropriate Framework. *IEEE T. Nanobiosci.*, 3(3):200–207, September 2004.

[WW05] K. Wilson and J. Walker. *Principles and Techniques of Practical Biochemistry.* Cambridge University Press, sixth edition, 2005.

[YHS00] T.M. Yi, Y. Huang, and J. Simon, M.I. Dolye. Robust Perfect Adaption in Bacterial Chemotaxis through Integral Feedback Control. *Proc. Natl. Acad. Sci. USA*, 97:4649–4653, 2000.

[YJM+00] K. Yeung, P. Janosch, B. McFerran, D. W. Rose, H. Mischak, J. M. Sedivy, and W. Kolch. Mechanism of Suppression of the Raf/MEK/Extracellular Signal-Regulated Kinase Pathway by the Raf Kinase Inhibitor Protein. *Mol. Cell Biol.*, 20:3079–3085, May 2000.

[YMO97] S. Yamaguchi, H. Murakami, and H. Okayama. A WD Repeat Protein Controls the Cell Cycle and Differentiation by Negatively Regulating Cdc2/B-Type Cyclin Complexes. *Mol. Biol. Cell*, 8(12):2475–2486, December 1997.

[YON00] S. Yamaguchi, H. Okayama, and P. Nurse. Fission yeast Fizzy-related protein srw1p is a G_1-specific promoter of mitotic cyclin B degradation. *EMBO J.*, 19(15):3968–3997, August 2000.

[YRD+01] K. C. Yeung, D. W. Rose, A. S. Dhillon, D. Yaros, M. Gustafsson, D. Chatterjee, B. McFerran, J. Wyche, W. Kolch, and J. M. Sedivy. Raf Kinase Inhibitor Protein Interacts with NF-κB-Inducing Kinase and TAK1 and Inhibits NF-κB Activation. *Mol. Cell. Biol.*, 21:7207–7217, November 2001.

[YSL+99] K. Yeung, T. Seitz, S. Li, P. Janosch, B. McFerran, C. Kaiser, F. Fee, K. D. Katsanakis, D. W. Rose, H. Mischak, J. M. Sedivy, and W. Kolch. Suppression of Raf-1 kinase activity and MAP kinase signalling by RKIP. *Nature*, 401:173–177, September 1999.

[ZYDQ06] Yuping Zhang, Huan Yu, Minghua Deng, and Minping Qian. Nonequilibrium Model for Yeast Cell Cycle. In *Computational Intelligence and Bioinformatics*, pages 786–791. Springer Berlin, 2006.