Molecular Systems Biology

Application of mathematical models and theoretical analyses in cell biology for students in the life sciences, engineers and physicists

F.J. Bruggeman, J.L. Snoep & B. Teusink
VU Universiteit, Amsterdam, The Netherlands
Stellenbosch University, South Africa
2012

Contents

Preface **v**

1 Cell biology 1.0.1 **1**
1.1 Cells are systems . . . 1
1.2 Examples of molecular networks . . . 2
1.2.1 Metabolic networks . . . 2
1.2.2 Signaling networks . . . 5
1.3 Examples of functional network properties and important network findings . . . 7
1.3.1 Cells are dynamic! . . . 7
1.3.2 Cells and their networks are organized in space and time! 9
1.3.3 Isogenic cells can display large cell-to-cell heterogeneity! . 10
1.3.4 Evolutionary network adaptations can be tracked and predicted! . . . 11
1.3.5 Network principles exist that apply to many species! . . . 13
1.3.6 Predictive mathematical modeling of molecular network dynamics is feasible! . . . 13
1.3.7 Foremost biology but also engineering, physics and mathematics in *SYSTEMS* BIOLOGY! . . . 13

2 Diffusion and reactions **15**
2.1 Introduction . . . 15
2.2 The intimate relationship between diffusion and reactions . . . 15
2.2.1 Diffusion of molecules precedes their reactive collisions . . 16
2.3 Diffusion of molecules . . . 19
2.4 Diffusion-limited reactions . . . 23
2.5 Single molecule location searches in the cytosol . . . 24
2.6 Different kinds of single-molecule searches . . . 25

3 Kinetic description of reactions between molecules **29**
3.1 Reactions between molecules are the basic processes of life . . . 29
3.2 The quantitative description of molecular reactions: mass-action kinetics . . . 29
3.3 Rate characteristics, thermodynamic equilibrium and steady state 33

3.4 Binding equilibria, association and dissociation constants 35
3.5 A number of biological examples 37
3.5.1 Protein complex formation (different perspective) 37
3.5.2 Cooperative transcription factor binding to DNA targets . 39
3.5.3 Negative autoregulation of a gene 40

4 Enzyme kinetics **45**
4.1 Enzymes 45
4.2 Irreversible Michaelis-Menten kinetics 45
4.2.1 Derivation of enzyme kinetics: quasi-steady state assumption 46
4.2.2 Derivation of enzyme kinetics: equilibrium-binding assumption 47
4.2.3 Consideration of inhibitors and activators: equilibrium-binding assumption 48
4.2.4 Exercises 49
4.2.5 Sensitivity of the enzyme rate to reactants and effectors . 50
4.3 Reversible Michaelis-Menten kinetics 50
4.4 Enzyme action and thermodynamics (advanced material) 55
4.5 Enzyme inhibition: quasi-steady state approach 58
4.6 Equilibrium binding models and convenience kinetics 60
4.7 Cooperative enzymes 64
4.7.1 The regulatory potential of cooperative enzymes 64
4.7.2 The Monod Wyman Changeux model for cooperative enzyme kinetics 65
4.7.3 The reversible Hill Equation 70

5 Stoichiometric network analysis **77**
5.1 Introduction 77
5.2 The stoichiometric matrix 78
5.3 Conservation relations and the link matrix L 78
5.4 Independent fluxes and the kernel matrix K 82
5.4.1 Exercise 84
5.5 Derivation of the K and L matrix using linear algebra for any N matrix 85
5.6 Steady-state flux space and constraint-based modeling 87
5.7 Applications and pitfalls 96
5.8 Sensitivity analysis 97

6 Biological control theory **103**
6.1 Introduction 103
6.2 System properties 104
6.3 The machinery of MCA: response, control and elasticity coefficients 107
6.4 Control coefficients for a linear pathway 110
6.4.1 Summation and connectivity theorems for linear pathways 113

6.5 A branched pathway: one robust branch and the other highly sensitive . . . 114
6.5.1 Control coefficients . . . 114
6.5.2 Low-flux branches and branch point control . . . 115
6.6 A metabolic pathway with negative feedback: homeostasis . . . 116
6.7 Ultrasensitivity of signaling networks . . . 118
6.7.1 A single covalent modification cycle . . . 118
6.7.2 A signaling cascade: sensitivity amplification . . . 120
6.7.3 Signaling cascades with feedback . . . 120

7 Dynamics of molecular systems 123
7.1 Stability of steady states . . . 123
7.2 Bistable dynamics of single-variable systems . . . 126
7.2.1 Emergence and disappearance of bistability as function of a parameter . . . 128
7.2.2 Bistability as a mechanism for memory . . . 130
7.3 Stability of two variable systems . . . 131
7.4 Phase plane analysis for two variable dynamic systems . . . 132
7.5 Stability of steady states of dynamic systems with two variables . 134
7.5.1 Analysis of the 2x2 jacobian matrix . . . 139
7.5.2 Exercises . . . 142

Preface

Cells are the fundamental units of life. Whether they exist as free-living organisms or as building blocks for multicellular organisms, cells are truly amazing molecular systems, consisting of up to several thousands of components whose integrated activities lead to the functional behavior characteristic of life. Growth and reproduction are of the most fundamental characteristics, but in addition cells can adapt to various and varying conditions by changing their composition. Short term adaptations are made via regulation of gene expression, and on longer time scales cells evolve via selection of beneficial mutations in the genetic material.

The cells' functional behavior emerges from the concerted activities of its underlying reaction network; linking small molecules (metabolites) to macromolecules (enzymes, proteins, genetic material) to function (e.g. pathway fluxes, signal transduction, growth rate, cellular adaptation). The cell is a dynamic system, needing a continuous free-energy influx for maintenance and growth delivered via the metabolic network, and adapting itself to changes in the environment, sensed by signal transduction networks. These networks cannot be understood on the basis of individual reactions or metabolites; its dynamics are crucially dependent on interactions between the components of the network. Ultimately network behavior can be understood on the basis of the network components and its interactions. This is the main aim of molecular systems biologists; to reach a quantitative understanding of cellular behavior on the basis of the characteristics of the cellular components.

The multitude of components, reactions and interactions in cellular networks and the non-linearity of many of its kinetics make cellular behavior both complicated and complex. Whereas classic treatment of cellular networks is qualitative (e.g. listing of components, reaction stoichiometries and interactions), or semi-quantitative (e.g. positive or negative regulatory interactions, reactions close or far away from chemical equilibrium), this cannot lead to the quantitative understanding for which we aim.

Mathematical (modeling) approaches are essential for quantitative studies of cellular systems, only via the integration of the activities of the network reactions can our hypotheses and predictions be tested critically against experimental data. Here is where the inter-disciplinary character of system biology approaches comes into play as the iteration between experiment, model and theory. These are extraordinary times for biological research with quantitative

experimental data sets for complete systems becoming available. Such datasets, in combination with knowledge of the systems' components can be used to construct and validate detailed kinetic models. Unlocking the enormous potential of mathematical models that allow quantitative predictions of biological system is one of the great challenges of our time!

In this book we focus on modeling and theoretical aspects of molecular systems biology approaches. We make use of simple systems for introducing the concepts and apply them to detailed systems and make frequent references to existing scientific literature to maintain a strong link with experimental data. We have written the book for final year undergraduate and postgraduate students in the life sciences, physics, and engineering. In short introductory chapters we fill potential gaps in biology or mathematics, and throughout the text we explain concepts starting from generic principles all the way to examples in the scientific literature. Topics include diffusion, chemical reaction kinetics, enzyme kinetics, reaction networks, the kinetic model, metabolic control analysis, and dynamic systems analysis. We hope you will enjoy reading and studying this book and that it may enrich your understanding of cellular systems and help in your research approaches.

Chapter 1

Cell biology 1.0.1

1.1 Cells are systems

Whereas the diversity of biological systems is often dazzling, there are some general characteristics that hold for all organisms. In analogy to the statement in physics; all material consists of atoms, for biology we can state that all organisms consist of cells. Of course not all cells are the same, basic differences exist between prokaryotic and eukaryotic cells, but from uni-cellular to mammalian species, they all consist of cells. The central dogma to which all these cells adhere is that DNA is transcribed to RNA which in turn is translated to proteins, linking the genetic material to the working horse in the cell.

Proteins function as the 'workers' inside cells, they perform all the tasks within cells to support their repair, environmental sensing, nutrient acquisition, and macromolecular turnover. Typically, proteins function as enzymes to catalyze reactions such as metabolite conversions, protein phosphorylation in signal transduction, DNA modifications to regulate gene expression and transport of molecules over cellular membranes. Other proteins have structural roles such as acting as components of the cytoskeleton, flagella or histones, which are involved in DNA organization. The activity of each protein depends on its abundance.

The amount of a given protein per cell depends on its stability and the abundance of its mRNA and translation rate. Each of these determinants may change with environmental conditions. The level of any mRNA results from transcription and degradation. The transcriptional activity of the corresponding gene depends typically on a set of transcription factors, all proteins. Transcription factors may be under the control of a signaling pathway, which may receive its signal at a ligand-sensitive receptor embedded in the cellular membrane. This example immediately indicates that we are quickly dealing with networks of molecules inside cells when we aim to understand cellular behavior.

As networks are so intertwined, cellular functions are carried out concertedly by whole segments of metabolism, signaling and gene circuitry. Any change at the level of metabolism or signal transduction will then tend to ripple through

the entire cellular network, with some segments responding strongly while others remain robust. Feedback loops are all over molecular networks, linking distant segments of the network, contributing to sensitization or robustness upon changes in the extracellular or intracellular environment. Feedback loops may make cellular dynamics very counterintuitive to understand. This complicated dynamics is often vital for the cell. We shall see that we quickly have to resort to mathematical models and theory to understand the behavior of molecular networks. This indicates that understanding of cell biology is not only about biology but also about mathematical models and concepts from mathematics, engineering, and physics.

In some abstract fashion, cells are 'just' molecular networks. Even though many properties at the cellular level are not always directly expressed or observed in molecular terms, they do all derive from it, such as growth rate or cell movement. Because cellular properties are so different from molecular properties, they are sometimes said to be emergent. For a scientist interested in a particular biological phenomenon, it is then always a challenge to figure out how emergent phenomenon arise out of the molecular interactions. Such a search for a molecular mechanism involves identifying the molecular components, their interactions, and key properties that contribute to the biological phenomenon to be explained.

Even though, in principle, all molecules inside cells are linked to each other through interactions, a particular phenomenon can be nearly always explained in terms of a molecular mechanism that only refers to a subset of all cellular molecules. Those explanations will often be in terms of mathematical models of the molecular mechanism couched in terms of the kinetic properties of molecular interactions. These models are central to this entire book.

In this chapter some examples of networks will be discussed - to give you some insight into the sort of networks this book is all about. Some of the recurrent properties of networks that are key to understand cell biology better will be briefly introduced. All of them will return at some point in this book where they will be explained in more depth.

1.2 Examples of molecular networks

1.2.1 Metabolic networks

Metabolic networks assimilate and convert nutrients into building blocks for cellular components, such as lipids, nucleic acids and amino acids. Those building blocks are converted further or polymerised by enzymes (or polymerize spontaneously) to yield proteins, RNA, DNA and membranes. For all these processes, energy is required. Energy is generated in a segment of metabolism called catabolism. This energy is required to make compounds in quantities and rates that would otherwise not spontaneously occur. In other words, cells operate in conditions out of thermodynamic equilibrium where an energy flux is required for cell function and maintenance. Cells are organized systems kept contin-

uously in a state out of thermodynamic equilibrium by processes that extract energy from nutrients. Anabolism is responsible for the usage of energy to make macromolecular components out of building blocks. Energy is stored in the displacement of metabolite ratios from their thermodynamic equilibrium values as we shall later on. The metabolites that act as the main energy carriers are ATP, NADPH and NADH. In figure 1.1 an example of a well-studied and important metabolic pathway, glycolysis, is presented. A segment of trehalose synthesis is shown as well.

Glycolysis is involved in the conversion of sugars into building blocks and energy metabolites. Many organisms rely on glycolysis; including yeast, most bacteria, and us. It is composed out of a large number of enzyme-catalyzed reactions in a sequence. Most of the reactions have multiple substrates and products. Some of the enzymes are regulated in their activity through metabolites that act as effectors, they are not consumed or produced by the reaction they regulate. For instance, often pyruvate kinase (PYK) is strongly activated by fructose-1,6-bisphosphate (F16P). This is indicated by the dashed arrows in figure 1.1.

The organization of glycolysis is as follows. Sugars, such as glucose, are composed out of six carbon atoms. Upto fructose-1,6-bisphosphate (F16P) all intermediates in glycolysis are composed out of six carbon atoms. Aldolase (ALD) then splits it into two molecules of composed out of three carbon atoms. Upto F16P two ATP molecules have been invested. If no glycerol is formed then 4 ATP molecules are generated by the lower part of glycolysis out of 1 molecule of glucose. If only ethanol is produced then at most two ethanol molecules can be formed per glucose molecule and no net synthesis or degradation of NADH will occur. The pathway will then only produce CO_2 in addition to ethanol (plus protons and water if the reactions are written in a higher level of detail). The production of ethanol is called fermentation, which is an important process in the rising of dough, beer brewing and wine making. In glycolysis a few branches occur. Two of them are shown in figure 1.1. They lead to excretion products or the construction of building blocks, for the synthesis of nucleic acids, amino acids, storage molecules, or lipids.

The response of *Saccharomyces ceresiae* to a glucose pulse is shown in figure 1.2. Five gram of this yeast per liter medium consumed $\approx$ 52 mM of glucose within 2 hrs and about 20 minutes. It excretes ethanol, which it consumed again after glucose has been consumed. This is known as the short-term Crabtree effect. The Crabtree effect refers to the behavior of *S. cerevisiae* to ferment - produce ethanol - under aerobic conditions. Many organisms ferment only in the absence of oxygen. The subsequent consumption of the produced ethanol is known as diauxi.

The structure of the pathway has been known for nearly one hundred years. Still biotechnologists are struggling with the metabolic engineering of this pathway to make better yeast strains for beer breweries and wine making. Often such studies aim at increasing the flux through glycolysis or the synthesis of ethanol. In the field of metabolic engineering mathematical modeling of metabolic pathways is a growing activity to get more insight into which proteins should be

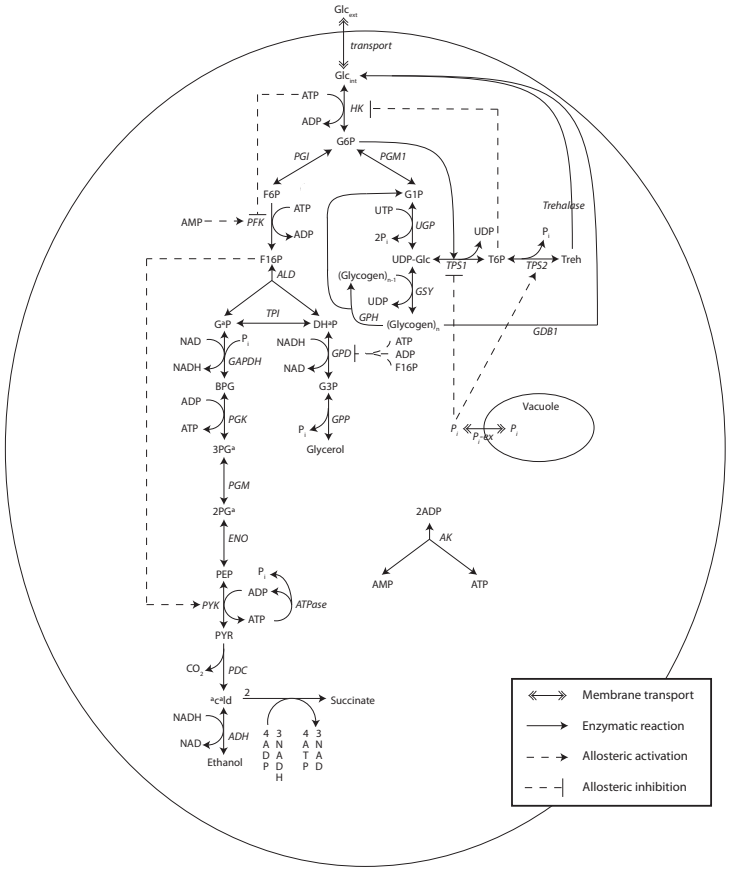

Figure 1.1: Network of diagram of glycolysis as it occurs in *Saccharomyces cerevisiae*, a yeast species of interest to biotechnology. It is also often used as a eukaryotic model organism in cell biology and systems biology. The arrows denote reactions, every reaction has a dedicated enzyme as a catalyst. The enzyme name is written next to the reaction, e.g. *HK*, for the reaction $GLC_{INT} + ATP \rightleftharpoons ADP + G6P$. Double-headed arrows indicate membrane transport reactions. Dotted lines with arrows denote an activating influence, lines ending in perpendicular lines denote inhibition of a reaction.

enhanced or decreased in activity to get the desired effect. Detailed mathematical models of glycolysis on the basis of extensive amounts of experimental data exist [33]. Often more than one enzyme needs to be changed in levels. This can be easily grasped: as soon as one of the major limiting enzymes has been enhanced in level another set of enzymes will become limiting. This delicate interplay between enzyme level, activity and importance for determining the

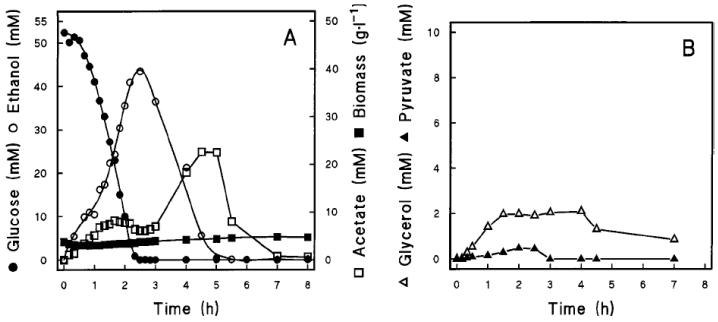

Figure 1.2: The dynamics of extracellular products of glycolysis upon a glucose pulse to *Saccharomyces cereviae* [10].

synthesis and consumption of industrially relevant compounds is partly due to glycolysis' complicated regulation. Glycolysis will return a couple of times in this book.

1.2.2 Signaling networks

Cells perseive their immediate environment through the action of membrane-embedded receptors. Those proteins, often dimers, transmit the presence of an external signal to proteins in the cytosol. Typically, upon ligand binding the conformation of the intracellular side of the receptor alters or the receptor modifies itself there, for instance by autophosphorylation. Downstream signaling proteins have a high-affinity for such alterations in the receptor structure. They form complexes with activated receptors to transmit the signal downstream. An example of a receptor driven signaling network is shown in figure 1.3. This signaling network is involved in regulating gene expression as function of an external signal, transforming growth factor β (TFGβ), to alter cell growth, adhesion, differentiation, and controlled cell death [28].

A number of proteins are involved in this process of TGFβ-induced gene regulation. Together they form a network with specific signaling properties that benefit the cell, such as a high ligand specificity, sigmoidal or hyperbolic ligand-dose transcription-factor-response relation, a response time, and the ability to integrate additional (intracellular) signals. The function of the network for the cell is determined by these network properties. They cannot be attributed to one single signaling protein; they are determined by all proteins in the network to varying extends. Hence, the entire network needs to be appreciated to understand how the cell uses this network for important decisions. This is what makes the identification of anti-cancer or diabetes drugs complicated.

In the Smad network, multiple proteins form complexes and alter each other's activity by phosphorylation and dephosphorylation. Some of the processes involve transport between the cytoplasm and nucleus over the nuclear membrane. Nucleocytoplasmic transport requires a dedicated protein, called the nuclear

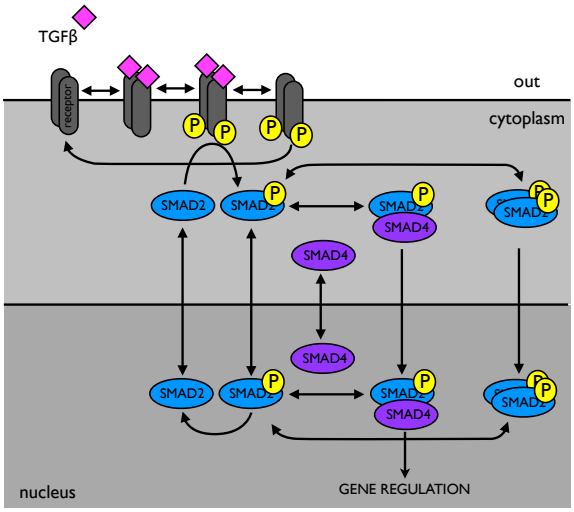

Figure 1.3: Network diagram of the Smad signaling network that responds to external TGFβ (TGF=transforming growth factor) levels [28]. This signaling pathway, as many others, is involved in regulation of cell growth, adhesion, migration, cell-fate determination and apoptosis.

pore complex. In addition, nuclear import and export requires a myriad of other proteins that assist in attaining net accumulations of SMAD's in the nucleus or cytoplasm depending on the external level of TFGβ. All of those are not shown in the diagram. The action of the phophatase in the nucleus leads to the dephosporylation of SMAD2 in the nucleus and hereby SMAD4 is released from the SMAD2P-SMAD4 complex. Representative dynamics of this signaling network measured at the level of single cells is shown in figure 1.4. This network has also been studied using mathematical models in tight interaction with experiments [6, 29].

Signaling networks often have a design where multiple phosphorylations of proteins occur in cascade. A prokaryotic and eukaryotic example are shown in figure 1.5. In early days, the advantage of such designs were elusive. Mathematical models and theory have improved our understanding of the benefits and trade-offs of signaling transduction cascades. Some of those aspects are discussed later in this book.

A well-studied signal transduction cascade is the EGF-induced MAPK cascade composed out of three MAPK proteins, MAPK, MAPKK, and MAPKKK. In the early days of mathematical studies on this system, it was hypothesized that this system could display high sensitivities of it's output, ERKPP, to the signal EGF, [18] and the system could display oscillations [20]. Even though, ultrasensitivity remains slightly controversial and is perhaps condition dependent, oscillations have now been observed experimentally at the level of single cells [31] (see figure 1.6).

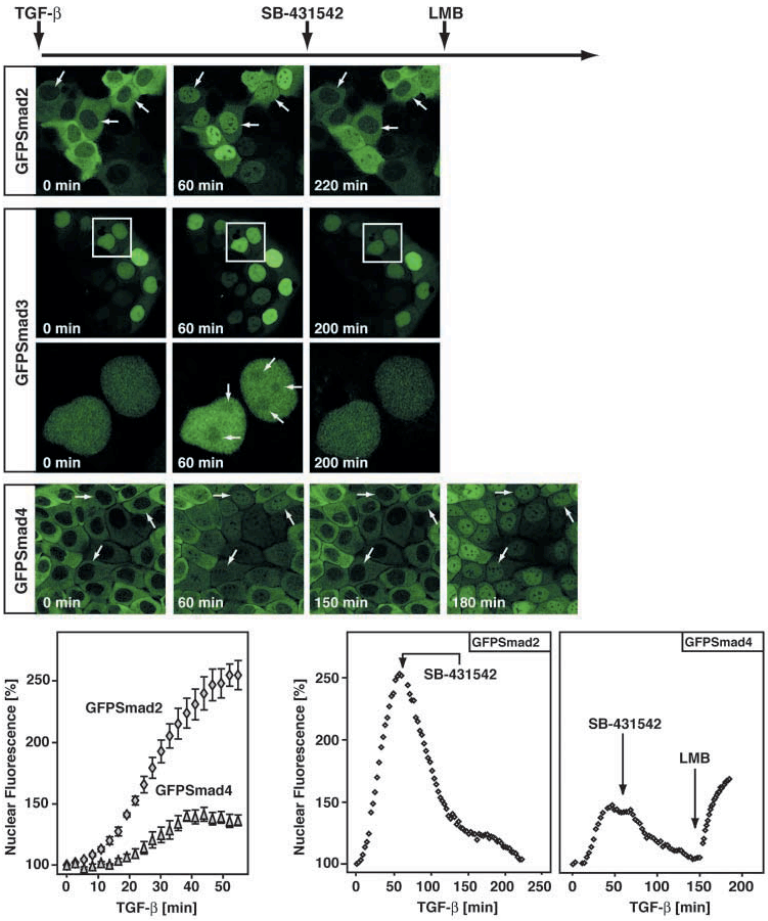

Figure 1.4: Single-cell dynamics of fluorescently labelled Smad2, GFPSmad2, and Smad4, GFPSmad4. Within 60 minutes both Smads translocate to the nucleus upon addition of TGFβ. The addition of the receptor inhibitor, SB-431542, causes Smad2 to return to the nucleus. Smad2 is then no longer phosphorylated causing unphosphorylated Smad2 to accumulate in the cytoplasm. Smad4 is no longer transported to the nucleus by phosphorylated Smad2. Smad4 export from the nucleus is mediated by a transport protein CRM1. The activity of CRM1 can be inhibited by LMB, which causes Smad4 to stay in the nucleus. These data were taken from [26].

1.3 Examples of functional network properties and important network findings

1.3.1 Cells are dynamic!

The dynamic responses of networks upon perturbation of cells, upon addition of nutrients, toxins, or changes in temperature, are rarely isolated to a few

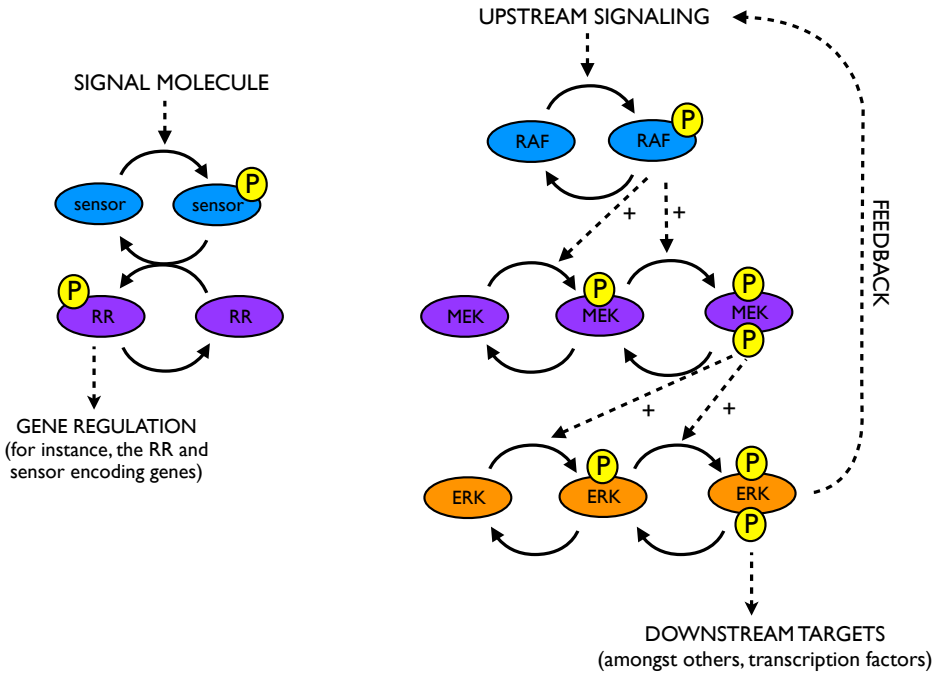

Figure 1.5: Protein phosphorylation is a recurrent mechanism for signal transduction (see also figure 1.3). Upon phosphorylation proteins change their affinity for signaling partners. Two component signaling as often observed in prokaryotes (and plants) and a mitogen activation protein kinase (MAPK) cascade. MAPK signaling has been studied in great depth in the last decade using mathematical models and quantitative experimentation [21, 31, 18].

proteins. Most of the time a large part of network responds; ranging in time scales from seconds to minutes to hours. Signal transduction and metabolism are typically fast but gene expression and protein turnover may take tens of minutes in bacteria to several hours in metazoans.

Even under steady external conditions, cells can display complicated dynamics. For instance, many well-known biological phenomena are periodic, such as the cell cycle and the circadian rhythm. Dedicated protein interaction networks maintain these oscillations and adjust progression depending on intracellular cues, such as spindle formation or DNA replication. Compared to the time scale of the cell cycle, which depending on the organism may range from 20 minutes to 24 hours in duration, many subnetworks are fast and may attain a quasi-stationary state. This is for instance likely the case for yeast and human glycolysis.

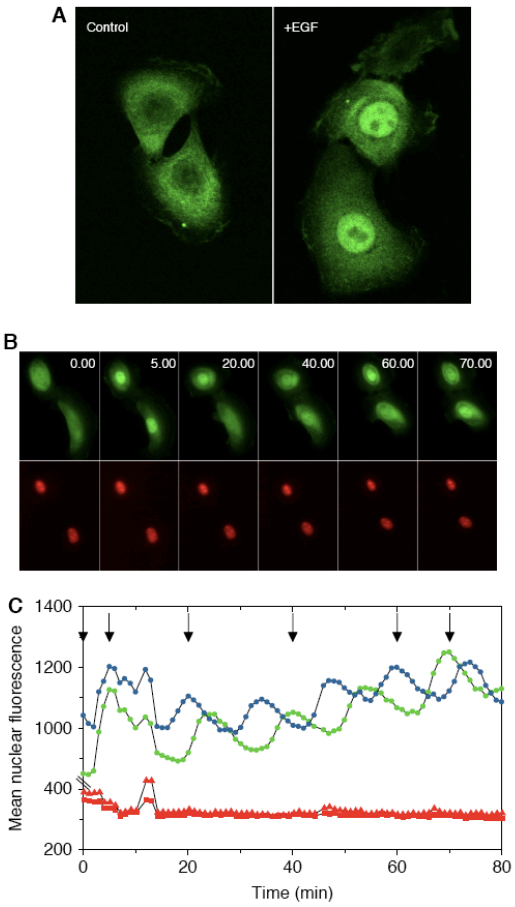

Figure 1.6: Single-cell monitoring of oscillatory dynamics of doubly phosphorylated ERK in both the cytoplasm and the nucleus [31]. Panel A shows that the total fluorescence increases upon addition of epidermal growth factor (EGF) - the signal of the signaling network. This indicates activation. In addition, $ERKPP$ accumulates in the nucleus. On top this assumulation oscillations occur (panel B (numbers indicate minutes after EGF addition) and figure C). The cells stained in red and the red data points indicate a control protein that only resides in the nucleus. In earlier experiments this dynamics was hard to observe as those often dealt with population studies. Then, oscillatory dynamics would only have been spotted if the cells would oscillate in synchrony.

1.3.2 Cells and their networks are organized in space and time!

Eukaryotic cells have intracellular compartments, organelles, that separate internal processes from the cytoplasm. Prokaryotes lack compartments but do

show dynamics induced spatiotemporal organization. For instance, in the regulation of the cell cycle where periodically varying gradients of signaling proteins allow the cell to identify its poles and middle. In eukaryotes, gradient formation of signaling proteins is for instance used in the cell movement where the cell needs to perform qualitatively different phenomena at its front and back. Those locations are separated in space through a gradient of several signaling molecules. This gradient is maintained by localized covalent modification reactions of signaling proteins and tightly interacts with the cell's machinery to extend its body forward and restructure its cytoskeleton.

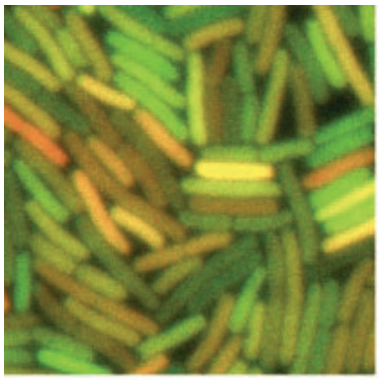

Figure 1.7: Heterogeneity of the expression of a fluorescent protein in a population of isogenic *Escherichia coli* cells [9].

1.3.3 Isogenic cells can display large cell-to-cell heterogeneity!

The introduction of fluorescent proteins and genetic engineering allow the observation of the dynamics of single cells. Cells with the same genetic make up - so-called isogenic cells - with the same growth history and current environment have been shown to display large cell-to-cell variability. An example of a snapshot of a population of isogenic cells expressing the same fluorescent protein in shown in figure 1.7. Those cells differ remarkably in the level of this protein! Representative distributions of fluorescence intensity across a cell population of *Saccharomyces cerevisiae* (yeast) cells are shown in figure 1.8. Figure 1.10 shows the dynamics of the variability of a fluorescently-labelled mRNA for four different *E. coli* cells. mRNA is produced in bursts of several molecules during "on" and "off" periods.

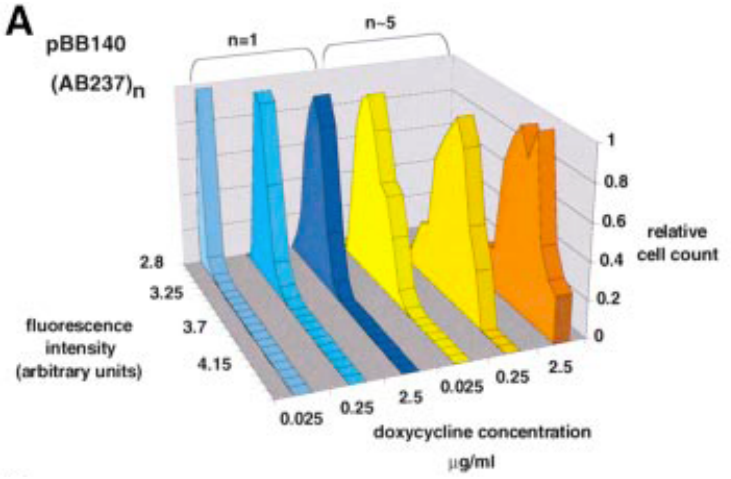

Figure 1.8: Distribution of fluorescence intensity per cell as function of a gene regulator [1].

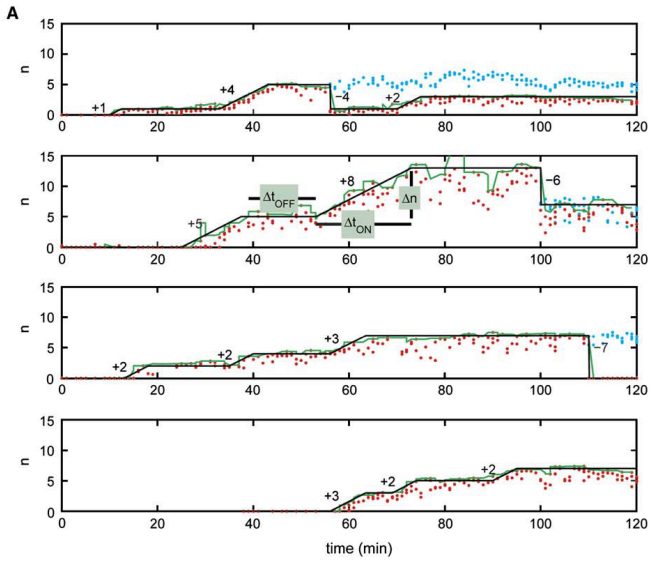

Figure 1.9: Variability of mRNA levels per cell as function of time for four cells [13].

1.3.4 Evolutionary network adaptations can be tracked and predicted!

Cells in a population accumulate mutations that gradually or in large jumps alter their physiological properties and fitness. Hereby some cells become by

chance better adapted to the current environment and will have a growth benefit. In this way, natural selection sieves better adapted mutants. By laboratory micro-evolutionary experiments this evolution can be tracked over time and in some cases theoretical predictions of optimal adaptation can be confirmed [19]. Mutations alter the properties of proteins and hereby whole network properties. The benefits of particular regulatory mechanisms inside networks, such as feedback loops and specific kinetic parameters, can be experimentally verified using this approach or by competition experiments of variants.

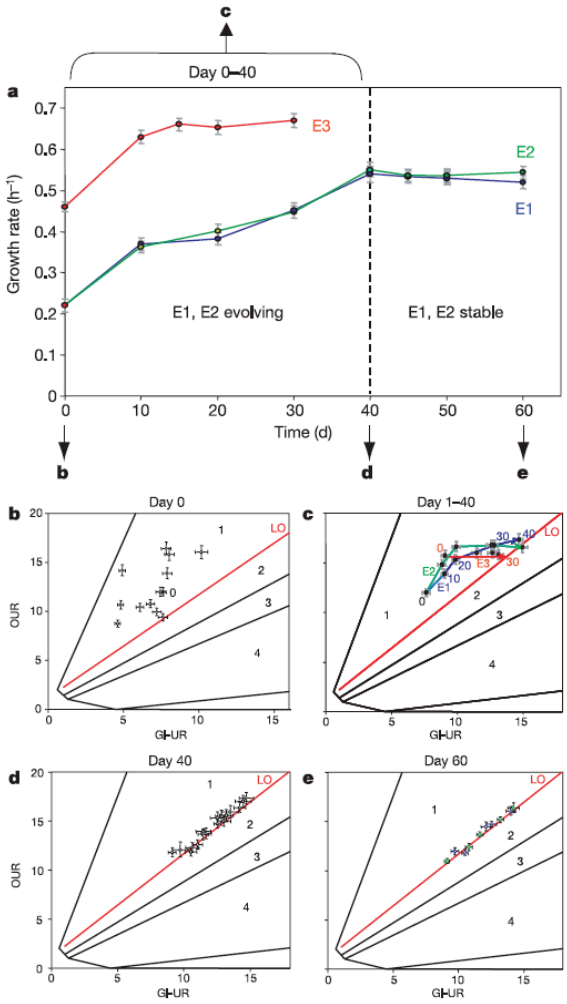

Figure 1.10: Adaptive evolution of *E. coli* growing on glycerol [19]. As function of time the population reaches the red line of optimality (figs b-e) as predicted by flux analysis of *E.coli*'s metabolic network.

1.3.5 Network principles exist that apply to many species!

As all organisms rely on molecular networks and use similar molecular regulatory mechanisms quite a few recurrent network designs have been identified that are used by distinct species. Similar signal transduction cascades, patterns of feedback and feedforward circuitry have been found. Negative feedback has been shown to be important for fast responses, robustness and giving rise to oscillations. Positive feedback turns out to be important for discrete switching in physiological states. Feedforward loops have shown to underlie sign-sensitive delays and pulse generators. Similar concepts and principles underlie the activity of networks with diverse functions in different species.

1.3.6 Predictive mathematical modeling of molecular network dynamics is feasible!

Besides the development of models and theory for the illustration of qualitative properties of networks to compare alternative network designs for their pros and cons, detailed mathematical models of molecular networks can be developed. Those models are useful in medicine and biotechnology as they allow for the prediction of systemic consequences of molecular perturbations. This facilitates metabolic engineering and drug target identification. Such approaches rely on quantitative experimental data on system behavior and/or kinetic properties of the molecular components of the network. Such data allows for model parameterization and validation prior to model usage as a predictive tool.

1.3.7 Foremost biology but also engineering, physics and mathematics in *SYSTEMS* BIOLOGY!

The development of advanced and quantitative techniques for the monitoring of the dynamics of molecular networks has shifted the emphasis in cell biology from molecule to network. The growing realization that a molecule centered approaches to medicine and biotechnology are limited due to the intricate functioning of molecular networks as molecular systems has made this shift an important one. A systems perspective brings with it different questions and challenges and requires usage of new techniques. Whereas many cell biologists could do without approaches from engineering, physics and mathematics before the network era, they are becoming more and more aware of the added benefit of multidisciplinary approaches. In the chapters that follow some of the basics for such a *systems biology* are explained.

Chapter 2

Diffusion and reactions

2.1 Introduction

The association of proteins into complexes is a basic process in cell biology: it is frequently occurring in signal transduction and gene expression regulation. In metabolism it is not found often. There, most of the reactions are enzymatic conversions of low-molecular weight organic molecules, such as pyruvate, glucose, amino acids, lipids, etc. Only when metabolic enzymes are covalently-modified, e.g. by phosphorylation, does protein-protein association play a role. Because protein-protein interactions are so prevalent, we will look into these processes in a bit more detail. Often, we can estimate the association rate constant for complex formation by considering the rate of this process as limited by the diffusional search of the two proteins for each other. In other words, the time that it takes before the complex is formed, i.e. the reaction time, which is the sum of the diffusional search time plus the time to form the complex (after the collision), is by more than 90% determined by the diffusional search time. This means we have to understand how the diffusional search time depends on properties of the proteins and the cytoplasm.

2.2 The intimate relationship between diffusion and reactions

A cellular compartment, such as the cytoplasm of a cell or the interior of an organelle, is a volume packed with (macro-) molecules. (Figure 2.1). It is often said to be a crowded with macromolecules environment, because the average distance between macromolecules is roughly the dimensions of a single macromolecule (average radius: $\approx 5\ nm$).

Cells contain numerous molecules of different types having different physicochemical properties and functions. All of this makes the intracellular milieu an environment difficult to mimic *in vitro*. This does not necessarily mean that

Property	Bacterium	Eukaryotic cell
Cell volume	1 μm^3	10000 μm^3
Proteins/cell	$4 \cdot 10^6$	$4 \cdot 10^{10}$
1 Protein/cell	1 nM	0.1 pM
Mean size of protein	5 nm	5 nm
Diffusion time of protein across cell	0.1 s (D=10 $\mu m^2/s$)	100 s
Diffusion time of small molecule across cell	1 ms (D=$10^3 \mu m^2/s$)	0.1 s

Table 2.1: Properties (estimates) of a bacterium, such as *E. coli*, compared to the properties of a mammalian cell (an eukaryote).

the *in vivo* milieu will be critical for all molecular properties but certainly for some. Macromolecular crowding favors for instance the association of proteins.

It is instructive to envision cells as shown in Figure 2.1 but then with all the molecules moving erratically. This mental picture and a number of other physicochemical properties of molecules and cells will turn out to be useful in understanding the fundamental basis of molecular reactions. This is the topic of this chapter.

Table 2.1 summarizes some properties we will need throughout this chapter.

Exercise

1. The volume of *E. coli* is approximately 1 μm^3.

   (a) Calculate the concentration of one molecule per cell in nM.

   (b) How many (spherical) receptors fit in the membrane of *E. coli* if the diameter of a receptor is 10 nm? Assume *E. coli* to be spherical; in reality, it is cigar shaped in most growth conditions.

   (c) How many macromolecules of similar dimensions as the receptors would fit in *E. coli*'s cytoplasm?

2. All nutrients need to be transported over the membrane. This means that the total transport capacity depends on total membrane area. Show that it is advantageous for a spherical cell to be small. What would determine the minimal cell size? Take into consideration that a cell with an increased volume would have to more protein to keep all the protein concentrations homeostatic.

2.2.1 Diffusion of molecules precedes their reactive collisions

Molecules inside cells engage in all kinds of molecular processes. Molecules have to be in close enough proximity and in the right relative orientation for a reaction to take place. Molecules find each other mostly through diffusion. Some

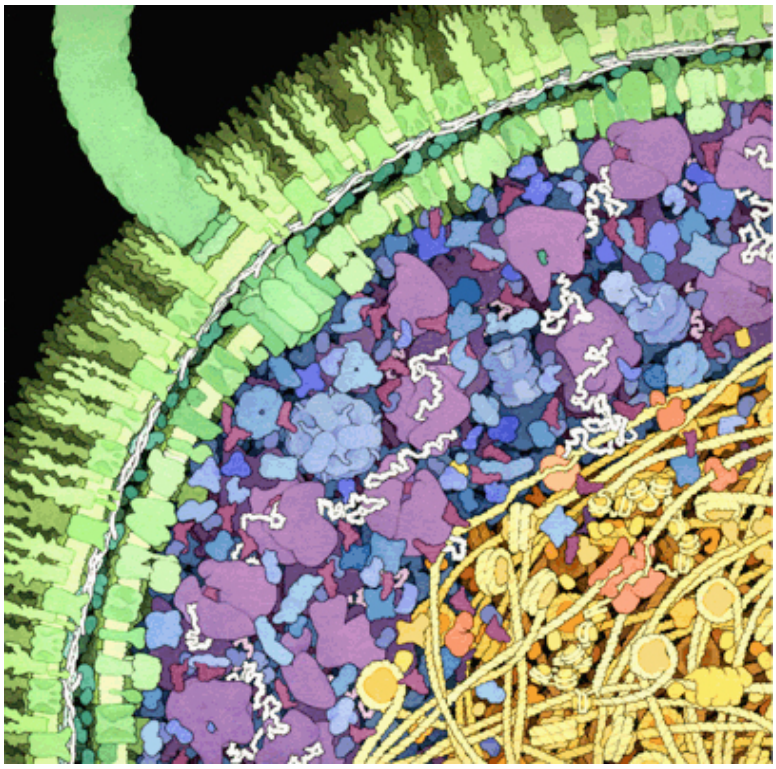

Figure 2.1: **An impression of a segment of a bacterium's membrane, cytoplasm and flagellum with realistic relative dimensions: an environment crowded with macromolecules.** In vacuum, these molecules would be moving really fast, with speeds of approximately $v = \sqrt{kT/m}$ (with k as Boltzmann's constant, T as temperature in Kelvin, and m as mass) in all three spatial dimensions. This means that at 27 degrees Celcius, an average protein with molecular weight $1.4 \cdot 10^4$ g/mol moves at $1.3 \cdot 10^3$ cm/sec! However, this considers the enzyme in a vacuum. In reality, it is emerged in a crowded aqueous environment where the motion is better described with a diffusion coefficient as we shall see below. The particle is continuously bumping (exchanging momentum) with the molecules in its vicinity. This leads to an erratic motion of the particle with very small jumps (of the size of the mean free path), a so-called random walk or drunken mans walk (see below). The mean free path for a macromolecule, is roughly the radius of the macromolecule: this is how packed - *crowded* - cells are!

molecules are transported by cargo proteins, such as dynein. They carry macromolecules or vesicles along the cytoskeleton to particular intracellular locations. But most protein diffuse through the cell's interior in three dimensions.

Diffusion is a "blind" processes, molecules move as a drunken man walks, as

they are bumping into other molecules (including water) all the time (Figure 2.2).

Figure 2.2: **Examples of four 2D random walks of 10000 diffusive steps.** The start position is indicated with a rate dot. The net distance travelled is given by $\sqrt{dN\delta^2}$ for a random walk of N diffusive steps of size δ in dimension d. The net distance travelled is the square root of the length of the diffusion trajectory. Because molecules change their direction of movement all the time, due to collisions, their travelled distance is much shorter than the length of the diffusion trajectory. Consider a drunken man. If he travels 1 m per step and has to walk 100 m in distance then he will likely have taken much more than 100 steps when he has reached his destination because he does not walk to the target point in a straight line.

We will now relate the diffusion of molecules with the reaction rate of association. The rate equation that describes the rate of an association reaction between two molecules A and B, i.e. $A + B \rightarrow AB$, is given by $k \cdot a \cdot b$. Here AB denotes the stable complex. The rate constant k is a second-order rate constant

and a realistic unit is $mM^{-1}min^{-1}$. Before the molecules A and B have formed the stable complex AB, two processes have to occur: i. the molecules have to bump into each other while they are each diffusing, and ii. after the collision they have to find the right relative orientation to form a stable complex; or, if the orientation does not matter, they just to form the stable complex. Thus the association rate $k \cdot a \cdot b$ depends on: (i) the number of their diffusive collisions per time unit and (ii) the fraction of collisions that lead to complex formation. This means that the time takes it for a complex to form, i.e. the association time, is the sum of the diffusion time plus the stable complex formation time. In the next section, we will consider the diffusion time.

2.3 Diffusion of molecules

Diffusive motion (without drift) is unbiased in direction. A molecule moving in one dimension at a particular location along an x-axis has probability $1/2$ to move to the right and probability $1/2$ to move to the left. In $3D$, at an instant of a diffusive step, the particle moves in the $+$ or $-$ direction of the x, y, and z axes; it moves diagonally. Examples are shown in Figure 2.2.

To get some insight into the statistics of diffusing molecules, we will consider 1D diffusion. Many of its properties can easily transferred to higher dimensions. Consider a molecule that has had sufficient time to make N diffusive steps, n_L to the left and n_R $(= N - n_L)$ steps to the right. The probability to travel n_L steps to the left out of N is given by the binomial distribution,

$$p(N, n_L) = \left(\begin{array}{c} N \\ n_L \end{array} \right) p^{n_L} (1-p)^{N-n_L} \tag{2.1}$$

This equation should look familiar from your elementary probability classes. It is a discrete probability distribution (a so-called probability mass function), which just means that n_L can only be chosen from 0,1,2,3,... . The probability p corresponds to the probability to move to the left and $1-p$ is the probability to move to the right. For normal diffusion those probabilities equal $1/2$. It is instructive to plot this equation as function of n_L at different values for N and p to get some intuition.

The mean and variance of a binomial distribution are Np and $Np(1-p)$. This means that on average $Np = 0.5N$ steps are taken to the left and therefore the same number of steps to the right. Thus on average the molecules do not move! This does not describe the process completely however. Over time, i.e. when the number of diffusive steps N increases, some molecules will of course have moved a longer distance. Since, always some molecules will exist that have had most of their steps to the left (or the right) the distribution of molecules will become broader over time. This is captured by the variance, which is a measured for the spread of the molecules over the medium or the width of the binomial distribution, because the variance increases with the number of steps N. If this sounds puzzling to you then imagine a drop of blue dye in the middle of the petri dish filled with water. Over time, the dye will diffusive out of the

centre to the edge of the dish. While this happens, the blueness will become less because the number of molecules at one location on the dish (the centre for instance) reduces. How the blue color spreads over the petri dish is described by the binomial distribution for different values of N.

When we consider movement to the right relative to the origin as a positive distance, the travelled distance Δ equals $n_R - n_L$. Often we are interested in situations where the number of diffusive steps is large such that N and Np are large. Under those conditions, a binomial distribution can approximated by a continuous Gaussian distribution; which is sometimes called a normal distribution or a bell curve (see: figure 2.3). The probability for a molecule to move distance d after time t then obeys the following Gaussian distribution,[1]

$$p(\Delta, t) = \frac{1}{\sqrt{4\pi D t}} e^{\frac{-\Delta^2}{4Dt}} \tag{2.2}$$

Here we have defined time as $t = N\tau$ and the diffusion coefficient as $D = \delta^2/(2\tau)$ with τ and δ as the (average) time and distance per one diffusion step, respectively. Note that the unit of a diffusion coefficient is a bit peculiar, e.g. $\mu m^2/s$ (because it equals $\delta^2/(2\tau)$), this will become clear at the moment. The spread of the probability as prescribed by this equation is shown in figure 2.3.

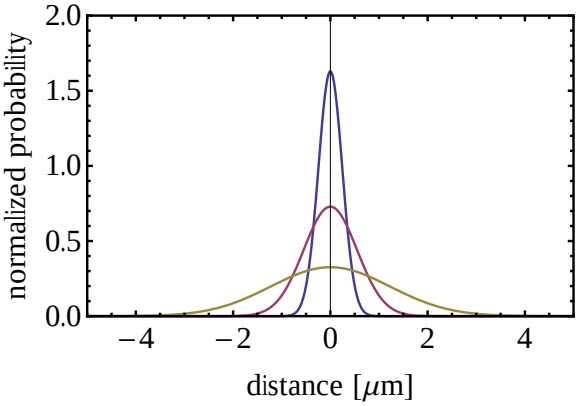

Figure 2.3: Examples of the probability density (equation 2.2) at three different times ($3{\cdot}10^{-3}$, $15{\cdot}10^{-3}$ and $75{\cdot}10^{-3}$ seconds; the diffusion coefficient corresponds to a realistic value (10 $\mu m^2/s$). The radius of an *E. coli* cell is about $1\mu m$.

One result should not surprise you by now given the unbiased nature of the random walk; the mean travelled distance, denoted by $\langle\Delta\rangle$, is zero. When we

[1]The notation for a mean or average of a stochastic variable x is $\langle x\rangle$. If x is a discrete variable and it comes in n values, i.e. $x_1, ..., x_i, ..., x_n$ then $\langle x\rangle = \sum_i^n x_i \cdot p(x_i)$ with $p(x_i)$ as the probability for x_i. One can envision $p(x_i)$ as given by number of occurrences of x_i in a large enough sample of x. In physics such a sample is called an ensemble. If x is a continuous variable with bounds x_L and x_H then $\langle x\rangle = \int_{x_L}^{x_H} x \cdot p(x)dx$. Here $p(x)$ is defined as the continuous probability distribution for x.

calculate this we get,[2]

$$\langle \Delta \rangle = \int_{-\infty}^{\infty} \Delta \cdot p(\Delta, t) d\Delta = 0 \tag{2.3}$$

In the previous example, with the binomial distribution, we found that the spread of the distribution increases with the number of steps N, which was a proxy for time. Intuitively this is in agreement with the slow spreading of the blue dye in a petri dish. This is spreading we can quantify again with the The results that we will use most often relates the travelled distance to the diffusion coefficient through the variance of Δ, denoted by $\langle \delta \Delta^2 \rangle$,

$$\langle \delta \Delta^2 \rangle = \int_{-\infty}^{\infty} \Delta^2 \cdot p(\Delta, t) d\Delta - \langle \Delta \rangle^2 = \langle \Delta^2 \rangle - \langle \Delta \rangle^2 = 2Dt \tag{2.4}$$

(Note that $\int_{-\infty}^{\infty} p(\Delta, t) d\Delta = 1$, the probability that a molecule has travelled any distance equals 1.) $\langle \delta \Delta^2 \rangle$ is often referred to as the mean squared displacement or, simply, the variance. The units of this measure is distance squared, e.g. μm^2. When we speak of the travelled distance, we should then consider $\sqrt{\langle \delta \Delta^2 \rangle}$ such that the units make sense. $\sqrt{\langle \delta \Delta^2 \rangle}$ is of course the standard deviation of the Gaussian distribution we are considering.

The probability that a particle has moved farther than $\sqrt{\langle \Delta^2 \rangle}$ after time t is given by,

$$P(\Delta > \sqrt{2Dt}, t) = 1 - \int_{-\sqrt{2Dt}}^{\sqrt{2Dt}} P(\Delta, t) d\Delta = 0.32 \quad (\approx 1/3) \tag{2.5}$$

This last result shows the applicability of equation 2.4 as it can be used to assess the minimal distance that 1/3 of an ensemble of molecules with diffusion coefficient D has moved after time t. The probability for molecules to move farther than twice the root mean square displacement in distance is 0.045. In other words, 95% of all the molecules will not have moved further than 2 $\sqrt{\langle \Delta^2 \rangle}$ in distance after t time for diffusion.

To extend the previous 1D result to three dimensions, we use the additivity rule for variances: as the diffusive motion in the x, y and z dimension are independent we can sum the variances such that,

$$\langle \Delta^2 \rangle = \langle \Delta_x^2 \rangle + \langle \Delta_y^2 \rangle + \langle \Delta_z^2 \rangle = 6Dt \tag{2.6}$$

This amounts to the following conclusion: a molecule with a diffusion coefficient D has travelled a distance farther than $\sqrt{6Dt}$ with probability 1/3 after time t or, equivalently, 2/3 of the molecules have travelled less than that distance. This is a useful relationship as it tells you something about how fast proteins move inside cells. This sets a limit to the rate of association reactions! Because the association of two proteins can of course not be faster than their diffusion speed.

[2]How to perform these calculations you do not need to know at this stage of the course.

With this knowledge we can address problems related to the duration of diffusive phenomena. For instance, considering that the diffusion coefficient of green fluorescent protein (GFP) is 25 $\mu m^2/s$ in mammalian cytoplasm (in water 87 $\mu m^2/s$ and in bacterial cytoplasm 7.7 $\mu m^2/s$) it can travel in 10 s a distance of $\sqrt{6 * 25 * 10} = 38$ μm. The length of *E. coli* is about 1 μm so this means a single molecule travels the length of *E. coli* in about $1^2/(6 * 7.7) = 0.02$ s on average! Below we will shall see that the diffusive searches of molecules to find another molecule in *E. coli* or a regulatory site on the DNA will take orders of magnitudes longer, in fact 10s of seconds.

So far, we have defined the diffusion coefficient of a particle in terms of the stochastic properties of its random walk. But intuitively, this parameter should depend on the size of the molecule (big things move slower), the viscosity of the medium (consider maple syrup versus water, diffusion in syrup is slower), and the temperature (molecules move quicker at higher temperatures). The physicochemical properties of the particle influence the diffusion coefficient according to the following (Einstein-Smoluchowski) relation,

$$D = \frac{kT}{f} \tag{2.7}$$

With k as the Boltzmann constant (J/K; $kg\ m^2 s^{-2} K^{-1}$), T as the absolute temperature (K) and f as the friction drag coefficient. For a spherical particle, f equals $6\pi\eta a$ with η as the (dynamic) viscosity ($kg\ m^{-1} s^{-1}$) and a as the radius of the particle (e.g., m); therefore,

$$D = \frac{kT}{6\pi\eta a} \tag{2.8}$$

Exercise

1. Diffusion of GFP
   (a) Plot the diffusion coefficient as function of the radius of a spherical particle. Take $30°C$ and express this temperature in units Kelvin.
   (b) Double the temperature and make the same plot.
   (c) How quick does GFP travel the radius of *E. coli* at $15°C$ and $30°C$?
   (d) Suppose we consider a dimer of GFP what happens to these search times?
2. What do you think kT means? (Check its units).
3. What is the unit of f? What do you think this quantity means?
4. The dimension of an average macromolecule is 5 nM in diameter. Calculate the diffusion coefficient using $\eta = 10^{-3} Pa\ s$ and $kT = 4 * 10^{-21}\ J$. Is this a realistic value?

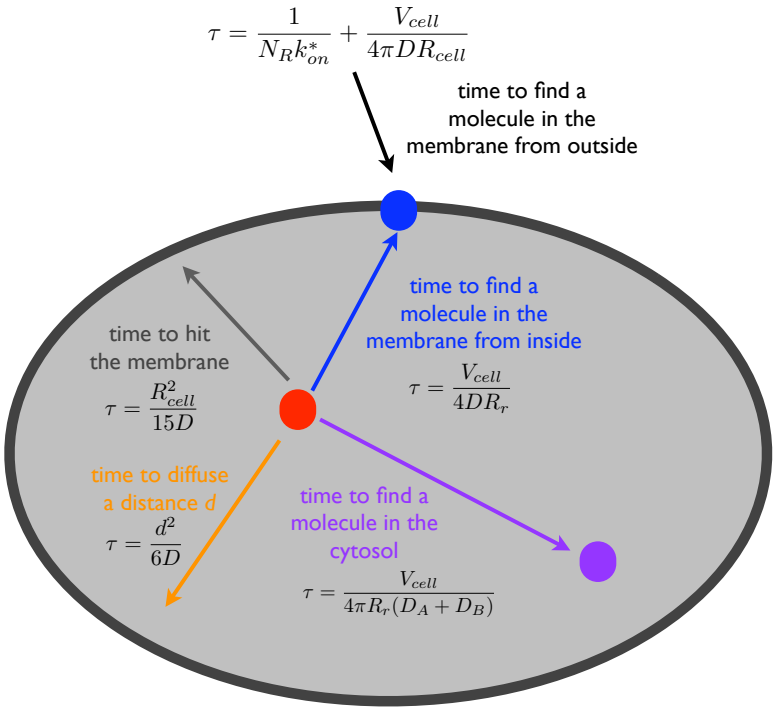

Figure 2.4: **Diffusion times for a single molecule to find the membrane (Dobrzynski and Bruggeman, unpublished), a molecule in the membrane (from inside and outside) [32, 2], to travel a certain distance d, and to find a (diffusing) molecule in the cytosol [14].** The D, D_A, and D_B denote a diffusion coefficient (in $\mu m^2/s$), R_{cell} the cell radius, R_r the reaction radius (often the sum of the radii of the reacting molecules), k^*_{on} the reaction-limited rate constant, and V_{cell} the cellular volume. A typical cell radius is 1 μm for a prokaryote, a diffusion coefficient is typically 5 $\mu m^2/s$ and a radius of a molecule is roughly 2.5 nm. All these diffusion times hold for single molecules, i.e. one molecule diffusing in the cytosol to find another single molecule in the membrane or in the cytosol. To take the reduction in time into account when multiple molecules are diffusing the search times can simply be divided by the number of molecules, so to find one receptor out of N_R with N_C cytosolic proteins reduces the time by $N_C N_R$ or if N_T transcription factors search for a single non-diffusing promotor on the DNA the search time becomes $V_{cell}/(4\pi R_r D_T N_T)$ with D_T as the diffusion coefficient of the transcription factor.

2.4 Diffusion-limited reactions

We are now ready to make the leap from diffusion to reactions. We consider two molecules A and B (with radii r_A and r_B, and diffusion coefficient D_A

and D_B, resp.) that form a complex AB with a second-order rate constant k_a ($molecules^{-1}\ s^{-1}$). The rate of the reaction is given by $v = k_a \cdot a \cdot b$. Given 1 molecule of A and B a characteristic time for this reaction is $1/k_a$. This time τ_a is the sum of a diffusion time τ_d and a reaction time τ_r. The diffusion (collision) time is given by the Smoluchowski equation (Figure 2.4),

$$\tau_d = \frac{V}{4\pi(D_A + D_B)(r_A + r_B))} \tag{2.9}$$

(Check what the units are. You have to realize that this is the time for two single molecules to find each other.) So the total time for the reaction is given by,

$$\tau = 1/k_a = \tau_d + \tau_r = \frac{V}{4\pi(D_A + D_B)(r_A + r_B)} + 1/k_r \tag{2.10}$$

If $1/k_a \approx \tau_d$ then the reaction is said to be diffusion limited and the rate constant for association becomes,

$$k_a = \frac{4\pi(D_A + D_B)(r_A + r_B)}{V} \tag{2.11}$$

Note that the unit of this equation is $(molecule/V_{cell})^{-1}s^{-1}$. This equation was first derived by Smoluchowski and represents the simplest expression for this type of rate constant. With D in cm^2/s and radius in cm multiplication of the latter equation by Avogadros number and division by 1000 gives the rate constant in $M^{-1}s^{-1}$.

If the reaction is reaction limited then $k_a \approx k_r$ and only the processes that occur after A and B have encountered each other determine the reaction time. This may have to do with finding the proper relative orientation or molecular state. In principle, this rate constant can be understood in terms of intramolecular dynamics using quantum mechanics (albeit, only in principle at the moment) or, more phenomenologically, with Eyrings rate theory, but this will not be considered here.

Exercises

1. Can a second-order rate constant be higher than the diffusion limit?

2. Use the diffusion-limited second order rate constant relationship to estimate the binding of the lac repressor to its DNA target site in $M^{-1}s^{-1}$. The diffusion coefficient of the receptor is $5 * 10^{-7} cm^2/s$. The average radius of the receptor is $40 * 10^{-10} m$ and that of the DNA site $10 * 10^{-10} m$.

2.5 Single molecule location searches in the cytosol

Equation 2.9 expresses the time for two single molecules A and B, to find each other in a cellular compartment given their radius and diffusion coefficients.

Let's look at the rate of association a bit more carefully,

$$\begin{aligned} v &= k_a \frac{n_a}{V}\frac{n_b}{V} \approx \frac{1}{\tau_d}\frac{n_a}{V}\frac{n_b}{V} \\ &= \frac{V}{4\pi(D_A+D_B)(r_A+r_B)}\frac{n_a}{V}\frac{n_b}{V} \end{aligned} \tag{2.12}$$

The unit of rate is now *number of molecules* $V^{-1}s^{-1}$. Clearly, if we have only one molecule of A and one of B the association rate is k_a and it takes then $1/k_a$ seconds between associations (on average). The search time (association time) reduces with the number of molecules as $\tau_d/(n_a/V \cdot n_b/V)$. Here n_a and n_b denote the number of molecules of A and B per cell. Let's convert the association rate is more a familiar unit, i.e. concentration per time,

$$v = N_A k_a \frac{n_a}{N_A V}\frac{n_b}{N_A V} \tag{2.13}$$

Here N_A equals Avogadro's constant (number of molecules per mol). $\frac{n_a}{N_A V}$ and $\frac{n_b}{N_A V}$ now equal the concentration of A and B in $mol/liter$. The rate v is now in $mol/(liter \cdot second)$, and $N_A k_a$ is $concentration^{-1}s^{-1}$. $N_A k_a$ is the association rate constant we have been using in previous chapters.

2.6 Different kinds of single-molecule searches

Many more search times, besides the time for two molecules in the cytosol to encounter each other, are relevant in cell biology. All those times can be used to estimate rate constants that we can use in kinetic models. Other expressions for search times are shown in figure 2.4. All these times, scale inversely proportional with molecule numbers. For instance, as the time to hit the membrane involves a single molecule the corresponding time scales as τ/n with the number of molecules per cell, n. The time to hit a molecule in the membrane involves two molecules and will scale as $\tau/(n_1 n_2)$. This makes sense because the rate of reactions increase with the substrate concentrations and, hence, the time between association events then needs to become shorter.

Let's study these equations a bit closer. The ratio over the time to find a molecule in the membrane versus the time to hit the membrane starting from a random position in the cytosol equals,

$$\frac{\text{time to find a molecule in the membrane}}{\text{time to find the membrane}} = 5\pi\frac{R_{cell}}{R_r} \tag{2.14}$$

This ratio is about 8000 for *E.coli* and increases linearly with the dimension of the cell. For a eukaryotic cell this ratio is about $8 * 10^5$. The time to find a molecule in the membrane divided by the time to find a molecule in the cytosol gives,

$$\frac{\text{time to find a molecule in the membrane}}{\text{time to find a molecule in the cytosol}} = 8\pi \tag{2.15}$$

The time to find a molecule in the membrane is always shorter than the time for finding a molecule in the cytosol, which continues to surprise me.

Exercise

1. Study Figure 2.4 and calculate the different times for a molecule to find the membrane, a single molecule in the membrane, a single moving molecule in the cytosol and to traverse the radius of the cell. Take *E. coli* parameters for these calculations. How how much longer do these processes take for a bigger cell with 1000 times the radius of *E. coli*.

2. Calculate the rate of complex formation between a cytosolic signaling protein and a membrane receptor when their concentrations are 300 and a 1000 molecules per cell, respectively. Those are realistic molecule numbers. Assume this rate to be diffusion limited. This process occurs in a eukaryote with cell radius of 80 μm. If a single complex exists for 30 seconds what is the dissociation rate constant. After some time, the association and dissociation process have reached equilibrium. In equilibrium the rates of association and dissociation are the same. What is the fraction of the receptor that is in a complex in equilibrium?

3. Can an average reaction time be shorter than an average diffusion time?

4. Prokaryotic signal transduction often takes place by way of two-component signal transduction. The fastest response time of such a signaling network, in the diffusion limit, is given by the sum of the time for the response regulator (RR; often a transcription factor) to find an active membrane sensor to become activated by phosphotransfer and the time for the active response regulator to find the DNA target site. This time can be approximated by (with N_{RR} and N_{RE} as the number of response regulators and receptors (membrane sensors)) respectively,

$$\tau_r = \frac{1}{N_{RR}} \left(\frac{V_{cell}}{4 D_{RR} R_r} \cdot \frac{1}{N_{RE}} + \frac{V_{cell}}{4\pi D_{RR} R_r} \right). \tag{2.16}$$

   Explain the structure of this equation by studying Figure 2.4 and it's legend. Take representative parameters for *E.coli* and calculate the response time for 25 sensor and 25 response regulator molecules. Is this a fast time? (The time to transcribe a gene and translate the resulting mRNA into protein takes about one minute.) Keep the total number of sensor and response regulator molecules fixed at 50 per cell. Plot the response time as function of the fraction of response regulators. Is there a bias towards having more response regulators than sensors? Can you explain why this is the case? Can you find the analytical expression for the optimal number of response regulators to minimize the response time given a fixed total number of molecules? If you can then plot the minimal response time and the optimal number of response regulators as function of the total number of signaling molecules (change this from 5 to 100). Is having 100 signaling molecules much better than having 50? Is having 50 signaling molecules much better than having 10? How many molecules would be enough for the system? You should realize that at some total

number of molecules, having more molecules will reduce the time so little such that the cost for protein synthesis exceeds the benefit of having one more signaling molecule.

5. A mammalian cell has typically a dimension of 10000 μm^3. Assume such a cell to have 25000 androgen receptors, which are transcription factors which upon binding androgen can regulate gene expression. Calculate the cellular concentration of this receptor. The diffusion coefficient of an androgen receptor is 2 $\mu m^2/s$. Calculate the time it takes for the androgen receptor to travel the radius of the cell (assume the cell to be spherical). The androgen receptor has a diameter of about 10 nm (assume it to be spherical) how many receptors fit inside the cell? Assume that the nucleus takes up 10% of the cell volume and that the nucleus and the cell are spherical. Androgen receptors typically reside in the cytosol when they are not bound to androgen. When active they can move through the nuclear pore complex to enter the nucleus and exit the cytosol. Say a cell has 10000 pore complexes. What is the diffusion-limited rate constant for transport from the cytosol to the nucleus? When androgen receptors are in the nucleus they have to find their targets on the DNA, say there are 500 of such targets, which are 10 nm in dimension. How much time does it take for a single receptor to find one of those targets when it starts in the nucleus and when it starts in the cytosol? The androgen receptor sits on the DNA for 50 seconds. What is the dissociation constant?

Chapter 3

Kinetic description of reactions between molecules

3.1 Reactions between molecules are the basic processes of life

To understand how the molecules inside cells bring about cellular behavior requires understanding of signaling, metabolism, and gene expression in molecular terms. Typically, tens to hundreds of proteins are involved in those cellular activities. Those proteins may act as enzymes and catalyze reactions or may have constructive role, for instance, actin plays a role as a monomer in the cytoskeleton, nucleosomes wrap DNA, or proteins making up a flagellum (the propellor that microorganisms use to move through fluids). In this chapter, we will study how we can quantitatively understand how the reactions between molecules, e.g. proteins, underlie changes in the concentrations of these molecules - dynamics. We will limit ourselves to uncatalyzed reactions and postpone enzyme kinetics to a later chapter.

3.2 The quantitative description of molecular reactions: mass-action kinetics

We will assume throughout this chapter that we can describe reactions between molecules without having to consider diffusion of molecules and stochastic aspects of reactions. Diffusion will be discussed shortly in the next chapter.

What kind of reactions exist between molecules? Well, two molecules can associate and form a complex. Alternatively, molecules can fall apart. In both cases, the concentrations of the substrates and the products of the reaction have changed after the occurrence of the reaction. How fast those concentrations changes depend on the rate of the reaction. If multiple reactions occur the

change in the concentration of a specific molecule depends on the net synthesis rate and the net degradation rate of this molecule. This means that if we account for the rate of all reactions that a given molecules plays a role in, as substrate or product, we can determine the change in the concentration of this molecule. This resembles molecular accounting. A natural approach to accounting is to make us of balances - as you do when managing your bank account. Here we done deal with the amount of money but with numbers of molecules, e.g. expressed as a concentration.

Thus, setting up a mass balance resembles molecular accounting and the same principles apply as when you manage your bank account. You keep track of the number of molecules (analogue: "euros" or "dollars") produced and consumed of a given species (analogue: "currency"), and the difference between those rates gives the net rate of change in the concentration of the molecule at a given moment in time. Let's write this down in mathematical terms. We consider a molecule, X, with concentration, x, which is for instance expressed in terms of mM. The rate of change in the concentration of this molecule X is denoted by dx/dt. One can think of dx/dt as the slope in a figure where the concentration x is plotted as function of time, t. If at a certain moment in time dx/dt is positive then the concentration rises, if it is negative the concentration drops and if is zero the concentration remains constant. The value of dx/dt at a certain time t equals the difference between the net rates of synthesis and degradation, $v_{synth}(t)$ and $v_{deg}(t)$, of this molecule X with concentration x,

$$\frac{d}{dt}x(t) = v_{synth}(t) - v_{deg}(t) = \sum_i v_{i,synth}(t) - \sum_j v_{j,deg}(t) \tag{3.1}$$

For every variable molecule concentration in the system of interest such an equation can be defined. Here we have explicitly indicated that the concentration and the reaction rates depend on time but we will often omit this notation. The symbol $\sum$ means that we take a sum of values; for instance,

$$\begin{aligned} 1+2+3+4+5 &= \sum_{i=1}^{5} i \\ y_1+y_2+y_3+y_4+y_5 &= \sum_{i=1}^{5} y_i \end{aligned}$$

Therefore, $\sum_i v_{i,synth}(t)$ means the sum of all the synthesis rates of X at time t. For the moment we will assume that we only have a single synthesis and degradation rate. If we choose concentration units in mM and time units in minutes, the units of rates are defined. The units of the two rates then necessarily have to be mM/min as the units at the right and left hand side of the equation always have to match. The two rates can depend on concentrations of other molecules in addition to X. The rate equations are given by either mass action or enzyme kinetics. Mass action kinetics will be studied in this section and enzyme kinetics in the next chapter.

Mass action kinetics applies to uncatalyzed reactions. Setting up a rate equation for a reaction involves very intuitive rules. For instance, for the isomerization reaction, $s \rightleftharpoons x$, the net rate of synthesis of X depends on the concentration of s, x, and its intrinsic rate constant for isomerization, k^+ and a similar rate constant for the isomerization of x, i.e. the backward reaction,

$$v = k^+ s - k^- x \tag{3.2}$$

If the unit of the reaction is expressed in terms of mM/min the unit of the concentration needs to mM and the unit of the rate constants are min^{-1}. The reversibility of the reaction dictates that the rate can also be negative, i.e. then S produced from X. The terms $k^+ s$ and $k^- x$ are referred to as the forward and the backward rate of the reaction. The rate constants k^+ and k^- are sometimes called elementary rate constants. They are first-order rate constants because the rate depend to first-order on the concentration, i.e. on x and not on x^2.

Now suppose that the molecules X and Y for a complex: $X + Y \rightleftharpoons XY$. The rate of the reaction is now described by,

$$v = k^+ \cdot x \cdot y - k^- xy \tag{3.3}$$

The unit of the rate constant k^+ should now be $min^{-1} mM^{-1}$. This rate constant is an example of a second-order rate constant, as its associated elementary rate depends on the concentration to second order, i.e. $x \cdot y$. Following this logic: a third order rate constant is then involved in $X + Y + Z \rightleftharpoons XYZ$.

For the reaction, $X + X \rightleftharpoons X_2$, we would obtain for the rate of synthesis of the complex the following rate equation,

$$v = k^+ x^2 - k^- x_2 \tag{3.4}$$

The dissociation rate is given by $-v$.

Some of you may have spotted the logic be now: in general, we obtain for reactions such as,

$$n_1 X_1 + n_2 X_2 + ... + n_s X_s \rightleftharpoons m_1 Y_1 + m_2 Y_2 + ... + m_p Y_p \tag{3.5}$$

the following rate equation for the reaction,

$$v = k^+ \prod_i^s x_i^{n_i} - k^- \prod_j^p y_j^{m_j} \tag{3.6}$$

The symbol $\prod$ means product,

$$\begin{aligned} 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 &= \prod_{i=1}^{5} i \\ Z_1 \cdot Z_2 \cdot Z_3 \cdot Z_4 \cdot Z_5 &= \prod_{i=1}^{5} Z_i \end{aligned} \tag{3.7}$$

Returning to equation 3.6, this means that per unit time $m_1 v$ molecules of Y_1 are made, and $m_i v$ molecules of Y_i.

There is one more thing to remember. Whenever a molecule is consumed or produced multiple times in a single reaction, such as $2X \rightleftharpoons X_2$, then the '2' in front of X is called a stoichiometry coefficient and needs to be taken into account in the mass balance for x. This is easy to understand as per unit rate more than molecule of x is consumed. We would get in this case,

$$\begin{aligned} \frac{dx}{dt} &= -2(k^+x^2 - k^-x_2) = -2v \\ \frac{dx_2}{dt} &= k^+x^2 - k^-x_2 = v \end{aligned}$$

as two molecules of x are consumed per unit rate, which occurs at speed $v = k^+x^2 - k^-x_2$. Here the rate is defined as the association rate. In addition, the total amount of molecules of X remains fixed in this case: no molecules are lost only interconverted. Thus we have the following relationship for the total concentration of X: $x_T = x + 2x_2$. This means that the consumption rate of x equals twice the production rate of x_2: thus $0 = dx/dt + 2dx_2/dt$ and $-dx/dt = 2dx_2/dt$ and this is true because $dx/dt + 2dx_2/dt = -2v + 2v$!

Exercises

1. Determine the mass balances and mass action kinetics for the following molecules and reactions. An underlined molecule indicates that it has a fixed concentration.

   (a) $S \rightleftharpoons X \rightleftharpoons P$

   (b) $\underline{S} \rightleftharpoons X \rightleftharpoons P$

   (c) $3A \rightleftharpoons 2B + C$, $B \rightleftharpoons 2D$, $2C \rightleftharpoons 3E$

   (d) $XY + Z \rightleftharpoons XYZ$, $XYZ \rightleftharpoons X + YZ$, $YZ \rightleftharpoons Y + Z$

2. Determine from these sets of mass balances the reactions,

   (a) $\frac{de}{dt} = -k_1^+ e \cdot s + k_1^- es + k_2^+ es - k_2^- e \cdot p$, $\frac{des}{dt} = k_1^+ e \cdot s - k_1^- es - k_2^+ es + k_2^- e \cdot p$, $\frac{ds}{dt} = -k_1^+ e \cdot s + k_1^- es$, $\frac{dp}{dt} = k_2^+ es - k_2^- e \cdot p$

   (b) $\frac{dx}{dt} = k_1^+ a \cdot x^2 - k_1^- x^3 - k_2^+ x + k_2^- b$

   (c) $\frac{dx}{dt} = k_1^+ a - k_1^- x + k_3 x^2 \cdot y$, $\frac{dy}{dt} = k_2 b - k_3 x^2 \cdot y$

   (d) $\frac{dx}{dt} = v_1 - v_2$, $\frac{dy}{dt} = v_2 - v_3$, $\frac{dz}{dt} = 4v_3 - v_1 - v_2 - v_4$ This is fact a simplified representation of glycolysis with X glucose-6p, Y as fructose1,6-phophate and Z as ATP. What is should be the substrate of reaction 1 and the product of reaction 3?

3.3 Rate characteristics, thermodynamic equilibrium and steady state

In the previous section, we have learned how to set up mass balances and rate equations for processes following mass action kinetics. This is the first step in making a kinetic model of a molecular network. These models are very useful to study basic properties of molecular systems that have to do with their dynamics, their control, and the importance of individual molecules and reactions for system behavior. Those models are central to this book.

Let's analyze the kinetic model of the following system, composed out of two reversible reactions and one variable intermediate X,

$$\underline{S} \rightleftharpoons X \rightleftharpoons \underline{P} \tag{3.8}$$

Remember that the underline of S and P indicates that their concentrations are kept fixed. We are therefore only dealing with a single mass balance. If we assume the rates to follow mass action kinetics, we arrive at,

$$\frac{dx}{dt} = \underbrace{k_1^+ s - k_1^- x}_{v_1} - \underbrace{(k_2^+ x - k_2^- p)}_{v_2} \tag{3.9}$$

Both of the rates of the processes depend on the concentration of molecule X, denoted by x. For a given concentration x these rates have a certain value and depending on their difference x may rise or fall, steeply or only slightly. Alternatively, the rate balance and x remains fixed. This is shown in Figure 3.1 where the rate characteristics of this system is displayed. A rate characteristic is a plot of rates as function of the concentration of a single molecular reactant.

The two lines in Figure 3.1 depict the rates of the reactions as function of x. When x equals 0 the rate of reaction 1 equals $k_1^+ s$ and $-k_2^- p$ for reaction 2. The two rates equal zero at different concentration of x; reaction 1 at $k_1^+ s/k_1^-$ and 2 at $k_2^- p/k_1^+$. This you can conclude by setting each of the rates of zero and solving for x.

Suppose you supply an initial amount of x slightly larger than the intersection of the rate characteristic of the first reaction with the x-axis. At that concentration of X, $v_2 > v_1$ and the concentration of X will decrease because $dX/dt < 0$. The rate with which x decreases becomes smaller as it approaches the intersection between the two rate curves because dX/dt gets smaller. This allows a sketch of the dynamics of X, in a plot of the X as function of time: it reduces from its initial concentration to its value at the intersection between the two rate characteristics. When the two rates eventually become equal, X is given by,

$$x_S = \frac{k_1^+ s + k_2^- p}{k_1^- + k_2^+} \tag{3.10}$$

This equation was obtained by setting the mass balance for x to zero, and solving for its stationary concentration x_s. This stationary state is referred

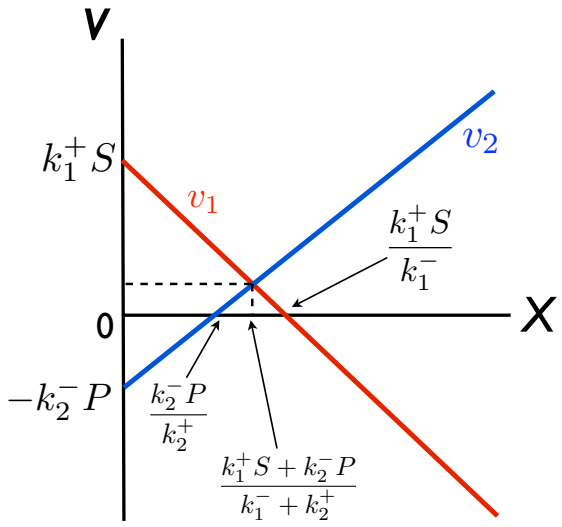

Figure 3.1: Rate characteristics of the reaction $\underline{s} \rightleftharpoons x \rightleftharpoons \underline{p}$ with reversible mass action kinetics.

to as a steady state. The system will reach this steady state from any initial concentration for molecule X. Do you understand why? (This can be concluded from the rate characteristic.) A steady state is defined as the stationary state in which at least one of the reactions is unequal to zero. A stationary state means that all the mass balances equal zero, which in this example will always correspond to the state reached after some time.

The previous expression for the steady-state concentration of X depends on the complete description of the system, all the kinetic constants and the characterization of the environment, the concentrations of S and P. The profound consequence is that already in this simple and biologically too simplistic example the entire system description determines system properties. It is not one molecule or process that is most important, but they all contribute! This fundamental property of molecular systems, i.e. their nonlinear nature and dependence on all molecular properties, makes biology so complicated and forces us to use mathematics and physics to better understand biology! Only the initial condition does not matter for the steady-state concentration of X. In a next chapter, we will study cases where stationary states do depend on the initial condition.

If we would consider the rate characteristics of the system, $\underline{S} \rightleftharpoons X$, the only feasible stationary state is a state where the net rate of reaction equals zero. Such a state is called a state of thermodynamic equilibrium. Its relation to thermodynamics will become clear later.

Note that the stationary state in Figure 3.1 can become a state of ther-

modynamic equilibrium when the values of s or p are chosen appropriately. Thermodynamic equilibrium will be the final state if P/S is chosen equal to,

$$\frac{p}{s} = \frac{k_1^+ k_2^+}{k_1^- k_2^-} \tag{3.11}$$

Only for this concentration ratio of P over S are the rates v_1 and v_2 both equal to zero in the state where x is constant.

You should realize that the rate constants, the 'k's', are properties of the reactants and the reaction conditions. An experimentalist can therefore only change the stationary rate by altering s or p.

Exercise

1. Sketch the dynamics of X as function of time on the basis of the rate characteristic; take $k_1^+ = 5, k_1^- = 1, k_2^+ = 3, k_2^- = 2$. Show that equation 3.11 indeed causes the system to settle to an equilibrium state where all reactions rate equal zero. Show that X then has the same stationary concentration as for the system $\underline{s} \rightleftharpoons x$. Show that the time to reach half the steady-state concentration is halved when all rate constants are doubled in value.

2. Plot the rate characteristic for $dx/dt = v_1 - v_2$ with $v_1 = 1/(1+x)$ and $v_2 = x/(1+x)$. For which concentration of X does v_1 equals v_2. Is this state, a steady state or an equilibrium state? What happens to x as function of time if the initial concentration of x lies below the concentration of X where $v_1 = v_2$? And what if it lies above this value?

3. Plot the rate characteristic for $dx/dt = v_1 - v_2$ with $v_1 = 1/(1+x)$ and $v_2 = V_2 x/(1+x)$ for different values of V_2 what happens to the concentration of x where $v_1 = v_2$? Does it increase or decrease? Why? How would you call the kinetic parameter V_2?

4. Consider the following reactions $\underline{A} \rightleftharpoons B, B \rightleftharpoons C, C \rightleftharpoons \underline{D}$. All these reactions follow reversible mass-action kinetics. Express the concentration ratio of D over A such that the system reaches thermodynamic equilibrium in terms of the rate constants of the reactions.

5. Do the same for:
$$\underline{A} \rightleftharpoons B, B \rightleftharpoons C, B \rightleftharpoons \underline{D}$$

3.4 Binding equilibria, association and dissociation constants

Complex formation between molecules is a fundamental process. It occurs in signaling where proteins dock onto receptors, in transcription where transcription factors bind to DNA, and in molecular machines, such as the ribosome,

where multiple protein together carry out a task. Binding events are often quantified in terms of a dissociation constant, which is a very useful parameter to assess the concentration of the proteins where a significant fraction of the protein exists in a complexed form.

Consider protein A and B, for instance a G-protein and a membrane receptor, that can form a complex,

$$A + B \rightleftharpoons AB \tag{3.12}$$

One of the relevant questions to ask is: what is the fraction of the molecules of A that exists in the complex? When is it 10%? When is it 90%? We will first assume that B is in excess. This means we only have to deal with the conservation of A molecules: $a_T = a + ab$. This equation tells you that if you start with 100 molecules of A in total that over time this amount will not change. This means that we can write for the mass balance of A,

$$\frac{da}{dt} = k_1^-(a_T - a) - k_1^+ a \cdot b \tag{3.13}$$

And this you can solve for the equilibrium concentrations using the information of the last section. Here we will achieve the same outcome but in a different manner. In the equilibrium state, the association rate and dissociate rate are equal and the total amount of A is fixed,

$$\begin{aligned} k_1^+ a \cdot b &= k_1^- ab \\ a_T &= a + ab \end{aligned}$$

We can eliminate ab to obtain,

$$a_T = a + \frac{k_1^+ a \cdot b}{k_1^-} = a\left(1 + \frac{b}{K_D}\right) \tag{3.14}$$

Here we have defined the *dissociation constant* K_D, which equals k_1^-/k_1^+. This means that unbound concentration of A equals,

$$a = \frac{a_T}{1 + \frac{b}{K_D}} \tag{3.15}$$

With the definition of the dissociation constant we can rewrite the equilibrium condition $k_1^- ab = k_1^+ a \cdot b$ as $ab = a \cdot b/K_D$ and we obtain for the bound concentration of A,

$$ab = \frac{a_T \frac{b}{K_D}}{1 + \frac{b}{K_D}} \tag{3.16}$$

The bound fraction is then ab/a_T. The dissociation constant has unit concentration. It indicates the concentration of b where the 50% of the molecules of A are in the complex because when $b = K_D$ the concentration ab equals $a_T/2$. So the measurement of the dissociation constant is useful exercise. Sometimes the association constant is considered, which is defined as $1/K_D$.

Exercise

1. Plot ab as function of b. What type of relationship do you find? What is the ratio of b/K_D where 10% and 90% of A is in the complex?

2. The K_D of a transcription factor for a DNA binding site is 1 nM. What is the concentration of the transcription factor such that bound fraction of binding sites is by 10%, 50% and 90%?

3. Consider the following reactions:

$$A + B \rightleftharpoons AB$$
$$A + AB \rightleftharpoons A_2B$$

   Define a K_D for the first reaction and the second reaction. Do you understand that those can indeed be different? Assume again that the total concentration of B is fixed and that A is in excess. Use the same procedure as explained in the last section to determine the expression of a_2b in terms of b, a_T, K_{D1} and K_{D2}.

4. The same as the previous question but now for:

$$\begin{aligned} A + B &\rightleftharpoons AB \\ A + AB &\rightleftharpoons A_2B \\ A + A_2B &\rightleftharpoons A_3B \end{aligned} \tag{3.17}$$

   (a) At what concentration of A is 50% of B in the A_3B complex?

   (b) At what concentration of A is 50% of B in the A_2B complex?

   (c) What is then the fraction of B in the AB and the A_3B complex?

3.5 A number of biological examples

3.5.1 Protein complex formation (different perspective)

The formation of macromolecular complexes composed out of multiple proteins is a recurrent phenomenon in signal transduction and gene expression. Let's consider the case where two proteins, A and B, form a complex,

$$A + B \rightleftharpoons AB \tag{3.18}$$

Both proteins now occur in a free form and in the complex. The forward rate constant is a second order rate constant. Let's consider, for simplicity, that B is in excess, such that $ab << b$. This means that the free concentration of the B, remains effectively constant, i.e. the free concentration equals the total concentration, $b \approx b_T$. Thus, we are dealing with,

$$A + \underline{B} \rightleftharpoons AB \tag{3.19}$$

As the total amount of A, denoted by a_T, is distributed over a and ab, we only have to consider the following mass balance,

$$\frac{dab}{dt} = k^+(a_T - ab)b - k^- ab \tag{3.20}$$

This is a linear ordinary differential equation, as it depends on ab in a linear manner. As a consequence, it can be solved analytically by hand or by using, for instance, Mathematica,

$$ab(t) = \left(1 - e^{-(k^+ b + k^-)t}\right) ab(\infty) = \left(1 - \frac{1}{e^{(k^+ b + k^-)t}}\right) ab(\infty) \tag{3.21}$$

Here $ab(\infty)$ equal the complex concentration at infinity, respectively. We assumed that that the initial concentration of AB equals 0. The term $1/e^{(k^+ b + k^-)t}$ converges to zero when time becomes large enough. This means that the approach to the equilibrium state can be sped up when any of the kinetic terms in or the concentration b in $k^+ b + k^-$ is increased. At time $t_{1/2} = \ln(2)/(k_b + bk_f)$ the concentration of ab equals half the equilibrium concentration. This definition of half-time is sometimes used as a measure for the characteristic time of the equilibration process. The state of thermodynamic equilibrium that is finally reached, when time goes to infinity, obeys two equations,

$$\begin{aligned} \frac{ab(\infty)}{a(\infty) \cdot b} &= \frac{k^+}{k^-} \\ a_T &= a(\infty) + ab(\infty) \end{aligned} \tag{3.22}$$

The first equation derives from the stationarity condition for the mass balance and the second expresses the conservation of the amount of molecule a. When we solve this for the complex concentration, we obtain

$$ab(\infty) = ab_{EQ} = \frac{a_T \cdot b}{K_D + b} \tag{3.23}$$

Here we have defined the dissociation constant K_D as k^+/k^-. This constant will have as its unit concentration, which you can verify easily. The stationary concentration of the complex increases in a hyperbolic fashion with the concentration of molecule b, B.

The characteristic life time of the complex is given by the time constant, $1/k^-$; indeed with time as its unit. This constant tells you how a complex lives on average before it dissociates.

Exercise

We consider the binding of a transcription factor, A, to a DNA site, B. Assume that the experimentally determined value for the dissociation constant is 1 nM. This is a characteristic affinity for regulatory sites on the DNA. Determine the life time of the transcription factor DNA complex when the association rate constant is diffusion limited and equals 1 $nM^{-1}s^{-1}$ Assume 10 transcription factors and 1 DNA site per cell and take *E. coli*'s cell volume (1 fl).

3.5.2 Cooperative transcription factor binding to DNA targets

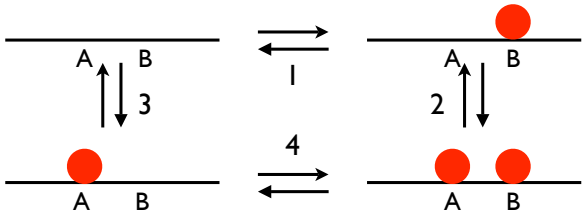

Figure 3.2: A state diagram of a gene promotor regulated by a transcription factor, the red ball. Two sites are present on the DNA, A and B.

We shall now consider the case of two transcription factors binding to two regulatory sites in the promotor region of a target gene. Those transcription factors are repressors and compete for binding with the RNA polymerase, which we do not consider here. As the activity of the gene depends on whether 0, 1, or 2 transcription factors are bound, we have to determine the fraction of promotors in these states. We assume the transcription factor, TF, to be in excess. We have to deal with four reactions,

$$\begin{aligned} P + \underline{TF} &\rightleftharpoons PTF \\ P + \underline{TF} &\rightleftharpoons TFP \\ PTF + \underline{TF} &\rightleftharpoons TFPTF \\ TFP + \underline{TF} &\rightleftharpoons TFPTF \end{aligned}$$

and one moiety conservation relationship that relates concentrations, expressing the fact that we have a fixed finite amount of promotors,

$$p_T = p + tfp + ptf + tfptf \tag{3.24}$$

At the stationary state, all those reactions will be in thermodynamic equilibrium. We can express the concentration of the product of each reaction in terms of the substrate concentration and a dissociation constant,

$$\begin{aligned} ptf &= \frac{p \cdot tf}{K_1} \\ tfp &= \frac{p \cdot tf}{K_2} \\ tfptf &= \frac{p \cdot tf^2}{\alpha K_1 K_2} \end{aligned}$$

Here the α factor is an interaction coefficient that captures the effect of the presence of one transcription factor on the DNA when the next one binds.

So the affinity of the DNA for the second transcription factor is altered when one transcription factor is already present. This cooperativity derives from a physical interaction of the transcription factors or a modulation of the DNA conformation upon binding of the first transcription factor.

These equations yield for the free concentration of the promotor,

$$p = \frac{p_T}{1 + \frac{tf}{K_1} + \frac{tf}{K_2} + \frac{tf^2}{\alpha K_1 K_2}} \tag{3.25}$$

and for the fully occupied promotor concentration,

$$tfptf = \frac{p_T \frac{tf^2}{\alpha K_1 K_2}}{1 + \frac{tf}{K_1} + \frac{tf}{K_2} + \frac{tf^2}{\alpha K_1 K_2}} \tag{3.26}$$

The relationship between the final complex concentration and the free concentrations of the two sites and the transcription factor should depend on the path travelled through the mechanism. So whether the complex was formed through a transcription factor first binding on the first site or on the second site should have no influence on the eventual equilibrium reached. This condition is called microscopic reversibility. Show that this condition is only met when the value of the interaction coefficients, α, is fixed; it does not depend on whether the first or the second site was occupied first by a transcription factor.

Exercise

Consider the previous section.

1. How are the dissociation constants defined in terms of rate constants?
2. Why do we need to introduce the α when the regulatory sites are different or when the transcription factor can interact on the DNA?
3. Make the derivation for $tfptf$ (equation 3.26) yourself.
4. Plot the concentration of $tfptf$ as function of the transcription factor concentration. Investigate the influence of K_1, K_2, and α.
5. What does $K_1 < K_2$ indicate?
6. What does $\alpha < 1$ indicate?

3.5.3 Negative autoregulation of a gene

Some transcription factors display autoregulatory behavior: they regulate their own expression by modulating the transcription rate of their own gene. Gene autoregulation is found very often. An example of a such a gene network is shown in Figure 3.3. The synthesis of mRNA corresponds to transcription and

that of protein to translation. Both the mRNA and transcription factor are degraded. The mass balances for this system then follow,

$$\begin{aligned} \frac{dmRNA}{dt} &= \frac{k_{sm}}{1+TF^n} - k_{dm}mRNA \\ \frac{dTF}{dt} &= k_{stf}mRNA - k_{dtf}TF \end{aligned} \tag{3.27}$$

The first mass balance consider transcription and turnover of mRNA. The second mass balance concerns translation and protein degradation. Note, it is instructive to plot the transcription rate term of $mRNA$ as function of TF and investigate the effect of n and k_{sm}. This gives you some insight into the dependency of autoregulation of transcription on the transcription factor concentration.

At steady state, when $dmRNA/dt = 0$ and $dTF/dt = 0$, the mRNA concentration is a solution of,

$$0 = \frac{k_{sm}}{1+\left(\frac{k_{stf}mRNA_S}{k_{dtf}}\right)^n} - k_{dm}mRNA_S \tag{3.28}$$

(Derive this equation yourself while reading.) This equation shows that an increase in mRNA will have a inhibiting effect on its synthesis rate. The gene functions like a homeostat, it's product is actively suppressing changes in its steady-state mRNA product level!

From the previous equation, we obtain,

$$k_{sm} = k_{dm}mRNA_S + \left(\frac{k_{stf}}{k_{dtf}}\right)^n k_{dm}{mRNA_S}^{n+1} \tag{3.29}$$

If we assume that $(k_{stf}/k_{dtf})^n k_{dm}{mRNA_S}^{n+1} >> k_{dm}mRNA_S$, we find for mRNA concentration at steady state,

$$mRNA_S = \left(\frac{k_{sm}}{k_{dm}}\right)^{\frac{1}{n+1}} \left(\frac{k_{dtf}}{k_{stf}}\right)^{\frac{n}{n+1}} \tag{3.30}$$

This equation shows that any change in the transcription rate, k_{sm}, is dampened by the negative feedback. The strength of the feedback increases with n. This can be easily investigated by determining the following sensitivity coefficient,

$$\frac{\partial mRNA}{\partial k_{sm}}\frac{k_{sm}}{mRNA} = \frac{\partial \ln mRNA}{\partial \ln k_{sm}} = \frac{1}{1+n} \tag{3.31}$$

This equation indicated that a 1% change in the transcription rate, due to some other factor not modeled here, will give rise to $1/(n+1)\%$ change in the steady-state mRNA level. So, strong feedback, i.e. large n will reduce the sensitivity of the mRNA concentration to change in the transcription rate constant, k_sm; for instance, due to the effect of other regulators. The same holds for a change in

$k_d m$. Higher values of n makes the system gradually more sensitive to change in k_{stf} and k_{dtf},

$$\frac{\partial \ln mRNA}{\partial \ln k_{stf}} = \frac{\partial \ln mRNA}{\partial \ln k_{dtf}} = \frac{n}{1+n} \tag{3.32}$$

This sensitivity ranges between 0.5 and 1.

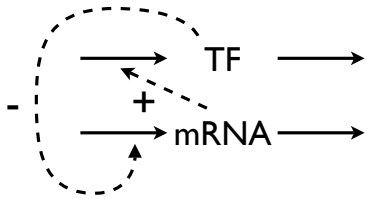

Figure 3.3: A transcription factor encoding gene that is inhibited by its own protein product.

Another aspect of negative autoregulation is that it speeds up the response of genes. This can be understood when we consider the following mass balance,

$$\frac{dmRNA}{dt} = \frac{k_{sm}}{1+\left(\frac{k_{stf} mRNA}{k_{dtf}}\right)^n} - k_{dm} mRNA \tag{3.33}$$

This mass balance may appear a bit artificial as $mRNA$ generally does not influence transcription directly. Here we assumed that the protein dynamics is so fast relative to $mRNA$ that it can be assumed in a stationary state on the time scale of $mRNA$ dynamics. Thus, $dTF/dt = 0$ is always practically zero even though mRNA levels are still changing drastically. If we want to understand the consequences of the negative autoregulation we should compare this description to the case without autoregulation. In order to do this properly we will require the steady state $mRNA$ level of the two descriptions to match. You should realize that this also forces the steady state $mRNA$ rate (J) to be equal between the two models. At steady state we find for the two descriptions,

$$\begin{aligned} \frac{k_{sm}}{1+\left(\frac{k_{stf} mRNA}{k_{dtf}}\right)^n} &= k_{dm} mRNA = J \\ k_{sm} &= k_{dm} mRNA = J \end{aligned} \tag{3.34}$$

In order for the feedback to be operative, $(k_{stf} mRNA/k_{dtf})^n > 1$, which means that the steady states of the two descriptions can only be identical if the transcription rate constant k_{sm} is larger for the network with the autoregulation. This implies that in the absence of any $mRNA$, the system with the negative feedback will response faster to a sudden increase in transcription activity!

Exercises

1. Kinase and phosphatase cycles occur very often in mammalian signaling networks (Figure 1.5). The kinase catalyzes the following reaction: $E +$

$ATP \rightleftharpoons EP + ADP$ and the phosphatase does the reverse: $EP \rightleftharpoons E + P$. Assume the kinase and phosphatase to follow simple Michaelis Menten kinetics, $v = V_{MAX}S/(S + Km)$ with S as the substrate concentration - E for the kinase and EP for the phosphatase -, V_{MAX} as the maximal enzyme rate and K_M as the Michaelis-Menten (half-saturation) affinity constant. We neglect the concentrations of ATP, ADP and P. The total concentration of enzyme substrate will remain fixed, i.e. $E+EP = E_{TOT}$, as there is no net synthesis or degradation of this enzyme.

(a) Plot the rate of the kinase and phosphatase reaction as function of the concentration of EP (which varies between 0 and E_{TOT}; take $K'_M s$ and V_{MAX} equal to 1 and $E_{tot} = 1$). You can use Excel, any other plotting package, Matlab or Mathematica. Show that the intersection of the kinase and phosphatase rate is a stable steady state.

(b) Assume that the V_{MAX} of the kinase increases linearly with a signal concentration L, i.e. $V_{MAX,kinase} = k_{cat}L$ (take k_{cat} as one and vary L from 0.1 to 10 in steps of 1. Plot the steady state concentration of EP as function of L for $E_{TOT} = 1$ and 10.

(c) Explain why the curve of EP as function of L becomes more switch-like when E_{TOT} increases. This mechanism for high-signal-sensitivity is called zero-order ultrasensitivity. We will return to this phenomenon later.

2. Consider the following reactions,

$$\begin{aligned} \underline{a} + 2x &\rightleftharpoons 3x \\ x &\rightleftharpoons \underline{b} \end{aligned}$$

The concentration of a and b are fixed.

(a) Determine the mass balance for the concentration of x, denoted by X. Take $A = 1, B = 1, k_1^+ = 10, k_1^- = 1, k_2^+ = 10$ and $k_2^- = 2$ and plot dX/dt as function of X.

(b) Count the number of intersections with the X-axis. Explain what happens to X when dX/dt is positive and negative.

(c) Show that you can identify two regions for the initial concentrations for x that each lead to different steady state concentrations of x.

(d) Conclude that depending on the initial conditions, the history of the system, the system can reach different steady states. This phenomenon is called bistability. This phenomenon will be studied later in more depth.

Chapter 4

Enzyme kinetics

4.1 Enzymes

Most reactions inside cells are catalyzed by enzymes, few reactions occur spontaneously. Metabolism, signal transduction, and gene transcription are all dependent on the actions of enzymes. By using enzymes, cells have control over what happens as they can modulate their levels and activities. Enzymes can only speed up reactions. They cannot alter the equilibrium constant of reactions. In their catalytic site, enzymes offer a favorable physicochemical environment for the reaction chemistry to occur. An enzyme may besides its catalytic site have regulatory sites, allosteric sites, that affect the kinetic properties through intramolecular signaling. Enzyme kinetics can be derived from a mass-action description of the elementary reactions involved in the enzyme mechanism, possibly under the influence of effectors. In this chapter, some of the basic concepts of enzymology will be explained. This chapter lies at the basis of the study of the dynamics and control of signaling, metabolic, and gene networks.

4.2 Irreversible Michaelis-Menten kinetics

The study of enzyme kinetics, enzymology, is a large field with many details. Many of these details and tricks can be found in this chapter. In this section, all the concepts and tricks contained in this chapter will be applied to the simplest enzyme mechanism that one can think of. Reading this section carefully will prepare you for what is to come and facilitates the reading of the next sections. Focussing only on the essential of these sections suffices.

In this section, we will consider a classical enzyme mechanism:

$$E + S \overset{v_1}{\rightleftharpoons} ES \overset{v_2}{\rightarrow} E + P \tag{4.1}$$

This description refers to a mass-action kinetics description of the two reactions

that occur,

$$\begin{aligned} v_1 &= k_1^+ e \cdot s - k_1^- es \\ v_2 &= k_2 es \end{aligned} \tag{4.2}$$

We would like to derive the rate of this reaction in terms of the familiar relationship used in biochemistry, called the Michaelis-Menten equation,

$$v = V_{MAX} \frac{s}{K_M + s} \tag{4.3}$$

This equation relates the rates of the reaction, the number of products produced per unit time, to the concentration s of the substrate and two kinetic properties of the enzyme, i.e. the maximal enzyme rate V_{MAX} and the Michaelis-Menten or affinity constant of the enzyme for S denoted by K_M. This equation prescribes a hyperbolic relationship between v and s with $v = 1/2V_{MAX}$ at $s = K_M$ and $v \to V_{MAX}$ when $s >> K_M$. When $v \approx V_{MAX}$, it is said that the enzyme is saturated.

However, we do not know yet how the kinetic constants V_{MAX} and K_M are related to the elementary rate constant k_1^+, k_1^- and k_2. This is what enzyme kinetics is all about. It can be done in two ways: by a quasi-steady state assumption and an equilibrium-binding assumption. These are explained in the next two subsections and used later for more complicated kinetics.

4.2.1 Derivation of enzyme kinetics: quasi-steady state assumption

The total amount of enzyme stays constant: $e_T = e + es$. We assume that the substrate is in excess over enzyme, $s >> e_T$. This means that we are effectively considering,

$$E + \underline{S} \overset{v_1}{\rightleftharpoons} ES \overset{v_2}{\to} E + P \tag{4.4}$$

($\underline{S}$ means S is fixed.) We then have the following two balances for the enzyme species,

$$\begin{aligned} \frac{de}{dt} &= -(k_1^+ e \cdot s - k_1^- es) + k_2 es \\ \frac{des}{dt} &= (k_1^+ e \cdot s - k_1^- es) - k_2 es \end{aligned} \tag{4.5}$$

Because the total amount of enzyme is fixed we obtain $-de/dt = des/dt$; this indicates that for every free enzyme consumed an enzyme-substrate complex is produced. The quasi-steady state assumption means that we assume that $\frac{de}{dt} = 0$ and $\frac{des}{dt} = 0$ while S is in excess and P is being produced. Then,

$$\begin{aligned} 0 &= -(k_1^+ e \cdot s - k_1^- es) + k_2 es \\ 0 &= k_1^+ e \cdot s - k_1^- es - k_2 es \end{aligned} \tag{4.6}$$

These equation allows us to solve for the (quasi-) steady state concentrations of e and es. Since, $e = e_T - es$ we can write the last equation solely in terms of es,

$$\begin{aligned} 0 &= k_1^+(e_T - es)\cdot s - k_1^- es - k_2 es \\ &= k_1^+ e_T \cdot s - k_1^+ es \cdot s - k_1^- es - k_2 es \\ &= k_1^+ e_T \cdot s - es(k_1^+ s + k_1^- + k_2) \Rightarrow \\ es &= \frac{k_1^+ e_T \cdot s}{k_1^+ s + k_1^- + k_2} \\ &= \frac{e_T \frac{k_1^+}{k_1^- + k_2}\cdot s}{\frac{k_1^+}{k_1^- + k_2} s + 1} \end{aligned} \tag{4.7}$$

The rate of the enzyme under quasi-steady state conditions equals $v = v_1 = v_2 = k_2 es$,

$$\begin{aligned} v &= k_2 es \\ &= k_2 e_T \frac{\frac{k_1^+}{k_1^- + k_2}\cdot s}{\frac{k_1^+}{k_1^- + k_2} s + 1} \\ &= V_{MAX}\frac{\frac{s}{K_M}}{\frac{s}{K_M} + 1} = V_{MAX}\frac{s}{s + K_M} \end{aligned} \tag{4.8}$$

Here the maximal rate of the enzyme, V_{MAX}, is defined as $k_2 e_T$ and the Michaelis-Menten constant as $K_M = \frac{k_1^- + k_2}{k_1^+}$.

Exercise

1. Determine the change in the substrate concentration when the enzyme rate changes from 10% to 90% of the maximal value.
2. Describe the quasi-steady state assumption in your own words.

4.2.2 Derivation of enzyme kinetics: equilibrium-binding assumption

Also for the equilibrium-binding assumption the substrate level is assumed fixed,

$$E + \underline{S} \overset{v_1}{\rightleftharpoons} ES \overset{v_2}{\rightarrow} E + P \tag{4.9}$$

Instead of assuming a steady state for the concentrations of the enzyme species while S is converted into P it is now assumed that reaction 1 is in thermodynamic equilibrium; then,

$$k_1^+ e \cdot s = k_1^- es \Rightarrow e = \frac{k_1^-}{k_1^+}\frac{es}{s} = K_S \frac{es}{s} \tag{4.10}$$

The K_S is now a dissociation constant. Using the relation for the conservation of total enzyme we can solve for the equilibrium concentration of es,

$$\begin{aligned} e_T &= e + es = es\left(\frac{K_S}{s} + 1\right) \Rightarrow \\ es &= \frac{e_T}{\frac{K_S}{s} + 1} = \frac{e_T \frac{s}{K_S}}{\frac{s}{K_S} + 1} = \frac{e_T s}{s + K_S} \end{aligned} \tag{4.11}$$

Again the rate of the enzyme equals $v = v_2 = k_2 es$ and therefore,

$$v = k_2 e_T \frac{s}{s + K_S} = V_{MAX} \frac{s}{s + K_S} \tag{4.12}$$

The V_{MAX} has the same definition as previously with the quasi-steady state approximation. The difference is in the definition of the Michaelis-Menten constant. As Michaelis and Menten defined the constant as K_M under the quasi-steady state condition, the derivation under the equilibrium-binding assumption should strictly not use the term Michaelis-Menten constant and the notation K_M. This is why we called it K_S in this section. Except for this minor difference the outcomes of the two derivations are exactly the same. Differences will appear between these two approaches when multiple substrates and products are considered.

Exercise

Which assumption is the most unrealistic the quasi-steady state or the equilibrium-binding assumption?

4.2.3 Consideration of inhibitors and activators: equilibrium-binding assumption

A pragmatic approach to the action of inhibitors and activators suggests that effectors of enzyme catalyzed reaction can influence the V_{MAX} and/or the K_M (or K_S). Hereby, the fluxes through an entire metabolic pathway can be effected because of the influence of the regulated enzyme on the pathway behavior. Thus, enzyme inhibition and activation is method to modulate pathway activity. By definition, an effector (i.e. inhibitor or activator) is not consumed by the enzyme, it only binds to the enzyme to change the enzyme properties.

In the presence of an effector, say "X", we have the following possible enzyme states,

$$e_T = e + ex + es + esx \tag{4.13}$$

Using, mass-action kinetics and equilibrium binding for X to the enzyme we can write the last equation as (you should know this by now),

$$e_T = e\left(1 + \frac{x}{K_1}\right) + es\left(1 + \frac{x}{K_2}\right) \tag{4.14}$$

We will next assume that S is in excess, X is an inhibitor, that ex and esx are dead-ends in the enzyme mechanism (conversion of ex into esx is prohibited), and that e and es are in equilibrium: then,

$$\begin{aligned}
e_T &= K_S \frac{es}{s}\left(1+\frac{x}{K_1}\right) + es\left(1+\frac{x}{K_2}\right) \\
&= es\left(\frac{K_S}{s}\left(1+\frac{x}{K_1}\right)+\left(1+\frac{x}{K_2}\right)\right) \Rightarrow \\
es &= \frac{e_T}{\frac{K_S}{s}\left(1+\frac{x}{K_1}\right)+\left(1+\frac{x}{K_2}\right)} \\
&= \frac{e_T}{\left(1+\frac{x}{K_2}\right)} \frac{1}{\frac{K_S}{s}\frac{1+\frac{x}{K_1}}{1+\frac{x}{K_2}}+1} \\
&= \frac{e_T}{\left(1+\frac{x}{K_2}\right)} \frac{\frac{s}{K_S}\frac{1+\frac{x}{K_2}}{1+\frac{x}{K_1}}}{\frac{s}{K_S}\frac{1+\frac{x}{K_2}}{1+\frac{x}{K_1}}+1} \\
&= \frac{e_T}{\left(1+\frac{x}{K_2}\right)} \frac{\frac{s}{K_S\frac{1+\frac{x}{K_1}}{1+\frac{x}{K_2}}}}{\frac{s}{K_S\frac{1+\frac{x}{K_1}}{1+\frac{x}{K_2}}}+1}
\end{aligned} \tag{4.15}$$

And the rate of the enzyme now becomes,

$$\begin{aligned}
v &= \frac{k_2 e_T}{\left(1+\frac{x}{K_2}\right)} \frac{\frac{s}{K_S\frac{1+\frac{x}{K_1}}{1+\frac{x}{K_2}}}}{\frac{s}{K_S\frac{1+\frac{x}{K_1}}{1+\frac{x}{K_2}}}+1} \\
&= V_{MAX}^{APP} \frac{\frac{s}{K_S^{APP}}}{\frac{s}{K_S^{APP}}+1} \\
&= V_{MAX}^{APP} \frac{s}{s+K_S^{APP}}
\end{aligned} \tag{4.16}$$

The last equation tells you that in the presence of inhibitor the V_{MAX} and the K_S are modulated to new values V_{MAX}^{APP} and K_S^{APP} but that the dependency of the enzyme rate on the substrate concentration remains hyperbolic.

4.2.4 Exercises

1. Derive the kinetics in the case that X cannot bind to ES but only to E.
2. Derive the kinetics in the case that X cannot bind to E but only to ES.
3. Compare the two equations that you have derived in the previous two exercises.

4. One of these mechanisms is called competitive inhibition. Which one do you think and why?

5. What should be the concentration of X if I want to inhibit the enzyme by 50%, for the two mechanisms you have derived yourself (question 1 and 2)?

4.2.5 Sensitivity of the enzyme rate to reactants and effectors

We have now assumed that the enzyme consisted of one subunit, so a single catalytic site per enzyme macromolecule. If an enzyme is composed out of multiple subunits, such that it is a protein complex, the subunits within the enzyme can affect each other's activities and sensitize and desensitize each other for their substrates. This phenomenon is known as cooperativity. Essentially, this means that the enzyme rate no longer depends on the substrate concentration in a hyperbolic fashion but that it displays a steeper dependence. This is often approximated by the Hill equation,

$$v = V_{MAX} \frac{s^n}{K_S^n + s^n} \tag{4.17}$$

This equation is completely phenomenological as we shall see later but what it does represent is an equation with greater sensitivity to the substrate concentration than a normal Michaelis-Menten type of relationship (when $n = 1$). This is easy to see when you consider the fractional change in the reaction rate upon a fractional change in the substrate concentration; i.e. the % change in the reaction rate upon a 1% change in the substrate concentration, this is much higher for enzymes with high values for n,

$$\frac{\partial \ln v}{\partial \ln s} = \frac{s}{v}\frac{\partial v}{\partial s} = n\frac{K_S^n}{K_S^n + S^n} \tag{4.18}$$

So multi-subunit enzymes can become very sensitive to their reactants and effectors, which makes them potent regulating enzymes with metabolic pathways. We will come back to this in a later section.

4.3 Reversible Michaelis-Menten kinetics

Monomeric enzymes have only one catalytic unit. We will consider enzymes with multiple subunits in section 4.7. The simplest reversible enzyme mechanism considers an enzyme, E, that converts a single substrate S into a single product P,

$$E + S \overset{v_1}{\rightleftharpoons} ES \overset{v_2}{\rightleftharpoons} EP \overset{v_3}{\rightleftharpoons} E + P \tag{4.19}$$

All three reactions are considered reversible and described by mass-action kinetics. ES and EP are often referred to as enzyme-substrate and enzyme-product

complexes, respectively. Reactions with a single substrate and single product are called uni-uni reactions; two substrates and a single product, bi-uni reactions, etc. We denote concentrations by normal letters and names of species by capitalized fonts.

The mass balances for all the species in mechanism 4.19 are given by,

$$\begin{aligned} \frac{ds}{dt} &= -v_1 \\ \frac{de}{dt} &= -v_1 + v_3 \\ \frac{des}{dt} &= v_1 - v_2 \\ \frac{dep}{dt} &= v_2 - v_3 \\ \frac{dp}{dt} &= v_3 \end{aligned} \tag{4.20}$$

To describe the entire process by a single rate equation, rather than by these 5 balances as it is now, we need additional assumptions for model reduction. This is the main achievement of enzyme kinetics besides rigorous methods for the determination of enzyme kinetic properties from experimental data. We will consider two approaches for the derivation of enzyme kinetics. They both have to do with differences in the dynamics of reactants and enzyme-reactant complexes. We will start with quasi-steady state descriptions before we consider equilibrium-binding models.

The net effect of the quasi-steady state assumption for enzyme kinetics is that the differential equations that describe the mass balance for all the enzyme species are set to zero and the concentration of the substrate and product are considered as constants. This assumption amounts to assuming that S and P have been added in such excess that any consumption or production of S and P by the enzyme, during the time it takes for the enzyme to reach a state where the concentration of the enzyme species no longer change, can be assumed not to influence the concentration of S and P. This means that on the time-scale of appreciable changes in S and P, it can be safely assumed that the concentrations of the enzyme species to be given by the equations resulting from their mass-balances set to zero. Accordingly, we are now left with the following set of equations,

$$\begin{aligned} \frac{de}{dt} &= -v_1 + v_3 = 0 \\ \frac{des}{dt} &= v_1 - v_2 = 0 \\ \frac{dep}{dt} &= v_2 - v_3 = 0 \end{aligned} \tag{4.21}$$

There is one other equation to consider that captures the conservation of enzyme species. As there is no net turnover of enzyme, we have,

$$e_{tot} = e + es + ep \tag{4.22}$$

This equation can be checked to be true from equation 4.21 as,

$$\frac{de}{dt} + \frac{des}{dt} + \frac{dep}{dt} = 0 \tag{4.23}$$

The product formation rate we are interested in is given by,

$$v = dp/dt = v_3 = k_3^+ es - k_3^- \cdot e \cdot p \tag{4.24}$$

We are considering the enzyme at steady state: $v_1 = v_2 = v_3$. We need to determine the steady-state concentrations of the enzyme species, e and es, in order to determine equation 4.24. As the equations in equation 4.21 are linearly dependent - they obey equation 4.23 - we need to use the conservation of total enzyme, equation 4.22, when solving for the enzyme species. This can be done by hand (as was done in section 4.2) or by using a matrix approach. The matrix approach is used here as this easily generalizes to more complicated enzyme mechanisms. This is done as follows, first we write the rate equations in terms of mass-action kinetics, substitute them in the mass balances, and write those in matrix format and set them to zero,[1]

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -k_1^+ s - k_3^- p & k_1^- & k_3^+ \\ k_1^+ s & -k_1 - k_2^+ & k_2^- \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} e \\ es \\ ep \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -e_{tot} \end{pmatrix} \tag{4.25}$$

Next, the concentrations of the enzyme species can be obtained through matrix inversion, which is the same as solving this system of equations by hand (section 4.2) for the three enzyme species,

$$\begin{pmatrix} e \\ es \\ ep \end{pmatrix} = \begin{pmatrix} -k_1^+ s - k_3^- p & k_1^- & k_3^+ \\ k_1^+ s & -k_1 - k_2^+ & k_2^- \\ 1 & 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \\ -e_{tot} \end{pmatrix} \tag{4.26}$$

Mathematical software packages such as Mathematica or Maple can do this matrix inversion for you. Substitution of the solutions for e and ep in equation 4.24 gives,

$$\frac{v}{e_{tot}} = \frac{\overbrace{k_1^+ k_2^+ k_3^+}^{num_1} s - \overbrace{k_1^- k_2^- k_3^-}^{num_2} p}{\underbrace{k_1^- k_2^- + k_1^- k_3^+ + k_2^+ k_3^+}_{const} + \underbrace{k_1^+ (k_2^- + k_2^+ + k_3^+)}_{coef_s} s + \underbrace{k_3^- (k_1^- + k_2^- + k_2^+)}_{coef_p} p} \tag{4.27}$$

Using the method worked out by Cleland [7], we have identified a number of terms: num_1, num_2, $const$, $coef_s$ and $coef_p$. Irregardless of the mechanism, such a term identification can always be achieved [30]. The maximal rate of the enzyme in the forward and the backward direction are defined as,

$$\begin{aligned} V_{MAX}^+ &= \frac{num_1}{coef_s} \\ V_{MAX}^- &= \frac{num_2}{coef_p} \end{aligned} \tag{4.28}$$

[1] Do this yourself once to convince yourself that can to this.

The half-saturation constants or Michaelis-Menten constants obey,

$$K_{MS} = \frac{const}{coef_s} \qquad (4.29)$$
$$K_{MP} = \frac{const}{coef_p}$$

Substitution of these equations into 4.27 gives the reversible Michaelis-Menten rate equation,

$$v = \frac{V^+_{MAX}\frac{s}{K_{MS}} - V^-_{MAX}\frac{p}{K_{MP}}}{1 + \frac{s}{K_{MS}} + \frac{p}{K_{MP}}} \qquad (4.30)$$

The product enters this equation in two ways in the denominator and numerator. The denominator term is termed kinetic inhibition and the numerator term is called thermodynamic inhibition.

In the absence of product, equation 4.30 simplifies into,

$$v = V^+_{MAX}\frac{s}{s + K_{MS}} \qquad (4.31)$$

A sketch of this curve is plotted in figure 4.1. A number of conditions clarify the meaning of the terms in this equation and give rise to a number of frequently used concepts,

1. If $s >> K_{MS}$ then $v \approx V^+_{MAX}$ and the enzyme is said to be saturated. It is no longer sensitive to the concentration of the substrate. The enzyme operates in its zero order regime,
2. If $s = K_{MS}$ then $v = V^+_{MAX}/2$. This defines the K_{MS} as a half-saturation constant,
3. If $s << K_{MS}$ the rate becomes $v \approx \frac{V^+_{MAX}}{K_{MS}}s$. The enzyme operates in its first-order regime,
4. for symmetry reasons the same definitions apply to p when $s = 0$

Exercises

1. Consider equation 4.30 and set the concentration of the product to zero. Why is the K_S often called the half-saturation constant in this equation? An enzyme that follows this rate equation is irreversible and product independent. For which concentrations of S is the rate most sensitive to the concentration of S?
2. Plot the rate of an enzyme modelled with equation 4.30 as function of S for constant values of P $(0.25, 0.75, 1.5, 7.5)$ take a V^+_{MAX} of 10 mM/min, K_S of 0.1 mM, K_P of 0.75 mM, and an equilibrium constant of 1000. Find the concentration of S where the enzyme is in thermodynamic equilibrium and check equation 4.30.

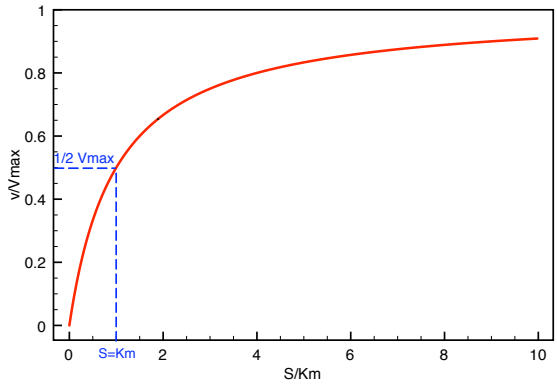

Figure 4.1: A sketch of the relative activity of an irreversible enzyme following Michaelis-Menten kinetics (equation 4.31) as function of its normalized substrate concentration.

3. Make a kinetic model of a reversible Michaelis-Menten enzyme in terms of its elementary reactions. Compare this model to its corresponding enzyme kinetics description and test whether the quasi-steady state approximation indeed works under the conditions described in the text.

4. An ordered bi-uni reaction has two substrates ('bi'), which bind in a strict order, and one product ('uni'). It has the following elementary reactions in it's catalytic mechanism,

$$\begin{aligned} e + s_1 &\rightleftharpoons es_1 \\ es_1 + s_2 &\rightleftharpoons es_1 s_2 \\ es_1 s_2 &\rightleftharpoons ep \\ ep &\rightleftharpoons e + p \end{aligned} \tag{4.32}$$

   Derive the rate equation of this reaction using the matrix method. Define the K_M's and V_{MAX}'s. Is the binding of s_1 and s_2 to the enzyme hindered by the presence of p? Show that the synthesis of p reduces at higher levels of p. Can a reduction in the rate of the enzyme, because of a decrease in the concentration of s_1, be compensated by a change in the concentration of s_2? At thermodynamic equilibrium the enzyme rate equals zero and the ratio of the product concentration over the product of the substrate concentrations equals the equilibrium constant of the reaction. This is a definition. Express the equilibrium constant in terms of kinetic parameters of the enzyme. This relationship is known as the Haldane relationship. Do you think the equilibrium constant is a property of the enzyme or of the reactants of the reaction?

5. Draw the cyclic catalytic network of an ordered bi-bi reaction without mentioning a single species twice.

4.4 Enzyme action and thermodynamics (advanced material)

Enzymes can only enhance the rate of reactions. The equilibrium constant of reactions cannot be altered by enzymes. The equilibrium constant of a reaction derives from the thermodynamic properties of its reactants. According to transition state theory (figure 4.2) enzymes enhance the rate of reactions by offering favorable conditions in their catalytic site. This lowers the activation energy of the reaction such that it occurs more rapidly in the catalytic site of an enzyme than spontaneously.[2]

Reactions taking place at constant temperature and pressure (the conditions in the living cell) occur in the direction of a reduction of the (Gibbs) free energy of a reaction. Thus, if the free energy of a certain amount of product is lower than that of substrate the reaction will produce product spontaneously. The reaction will stop when the free energy difference becomes zero. The reaction is then in thermodynamic equilibrium, a state of maximal entropy. The (partial) molar Gibbs free energy of a molecule A is given by,

$$\mu_A = \mu_A^{0'} + RT \ln a \tag{4.33}$$

The unit of molar Gibbs free energy is J/mol, the universal gas constant R has as it's unit $J/(mol \cdot K)$, and temperature T is in Kelvin.[3] The constant $\mu_A^{0'}$ is the molar Gibbs free energy (J/mol) defined under standard biochemical conditions (concentrations are 1 molar, temperature 298 K, and pH is 7). The Gibbs free energy potential of a reaction, ΔG_R, is the difference in Gibbs free energy of the products and the substrates taking into account their stoichiometric coefficients. For the reaction $2A + B \rightleftarrows A_2B$ we obtain,

$$\Delta G_R = \mu_{A_2B} - 2\mu_A - \mu_B = \Delta G_R^{0'} + RT \ln \frac{a_2 b}{a^2 \cdot b} \tag{4.34}$$

(Here: $G_R^{0'} = \mu_{A_2B}^{0'} - 2\mu_A^{0'} - \mu_B^{0'}$.) At thermodynamic equilibrium, the rate of a reaction and it's Gibbs free energy potential of the reaction are zero. At this state, we obtain,

$$\frac{a_2 b_{EQ}}{(a_{EQ})^2 \cdot b_{EQ}} = e^{\frac{-\Delta G_R^{0'}}{RT}} \equiv K_{EQ} \tag{4.35}$$

In this equation, the equilibrium concentrations appear (subscript EQ) and the equilibrium constant, K_{EQ}. The actual ratio $\frac{a_2 b}{a^2 \cdot b}$ is defined as the mass action ratio Γ. The deviation from thermodynamic equilibrium is captured by Γ / K_{EQ}. Indeed, using equation 4.34 and 4.35 we can write the Gibbs free energy potential of the reaction as,

$$\Delta G_R = RT \ln \frac{\Gamma}{K_{EQ}} \tag{4.36}$$

[2] An analogy to activation energy is the requirement of a lighter to put wood on fire. Wood will continue to burn spontaneously (wood ash has lower free energy) after it has been ignited.

[3] $R = N_A k_B$ where N_A is Avogadro's constant and k_B Boltzman's constant.

The last equation also equals $RT \ln v^-/v^+$ where the rate of the reaction is given by the difference between the forward and backward rate: $v = v^+ - v^- = k^+ a^2 b - k^- a_2 b$.

We can rewrite equation 4.30 in terms of an equilibrium constant when we realize that: i. the rate equals zero if the enzyme is at thermodynamic equilibrium and ii. under those conditions, the mass action ratio equals the equilibrium constant for the reaction; we obtain,

$$K_{EQ} = \frac{p_{EQ}}{s_{EQ}} = \frac{V^-_{MAX} K_{MS}}{V^+_{MAX} K_{MP}} \tag{4.37}$$

This relationship is known as the Haldane relationship, which allows us to rewrite equation 4.30 as,

$$v = \frac{V^+_{MAX} \frac{s}{K_{MS}} \left(1 - \frac{p}{sK_{EQ}}\right)}{1 + \frac{s}{K_{MS}} + \frac{p}{K_{MP}}} \tag{4.38}$$

As the equilibrium constant is a property of the reactants, the enzyme kinetic properties will always have to obey the Haldane relationship.

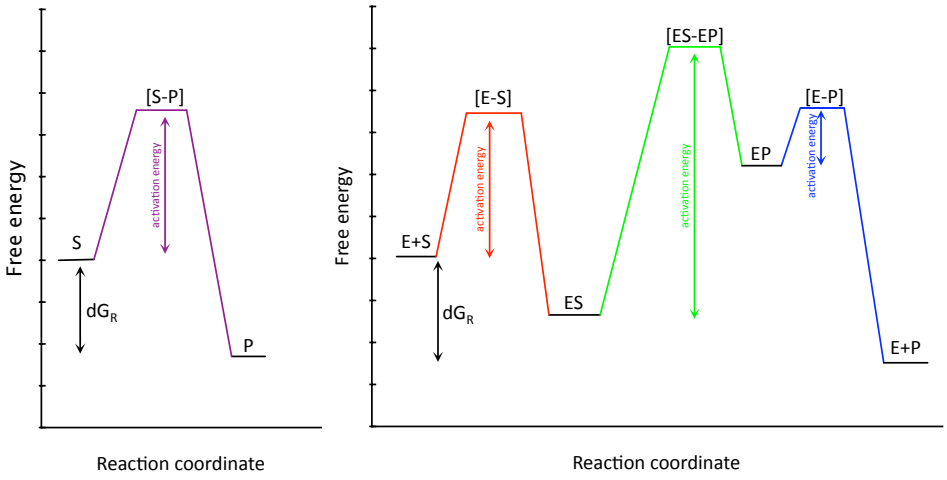

Figure 4.2: A plot of the free energy of representative states for the reaction of S to P as function of the reaction progression (reaction coordinate) when the reaction is enzyme spontaneous ($S \rightleftarrows P$, left) and enzyme-catalyzed ($E + S \rightleftarrows ES \rightleftarrows EP \rightleftarrows E + P$, right). In transition state theory, enzymes enhance the speed of reactions by lowering the activation energy for the reaction. Enzymes achieves this by offering a favorable physicochemical environment for the reaction chemistry in their catalytic site. The $dG_R = \Delta G_R = RT \ln \Gamma/K_{EQ}$ is the same for the spontaneous and enzyme-catalyzed reaction.

Exercises

1. Show that the equilibrium constant of a linear chain of enzymes equals the product of the equilibrium constants of the reactions. Show for the same system that the Gibbs free energy driving the system equals the sum of the Gibbs free energies driving the reactions.

2. Consider a kinase and phosphatase catalyzing the phosphorylation of an enzyme, $E+ATP \rightleftharpoons ADP+EP$, and its dephosphorylation $EP \rightleftharpoons E+Pi$, respectively. Show that this system is driven by the free energy potential of $ATP \rightleftharpoons ADP + Pi$ when we consider those species fixed. Show that only under those conditions a steady state can be reached. Show that at thermodynamic equilibrium the regulation of a kinase by a signal does not affect the concentration of EP and, therefore, no signal transmission can occur.

3. Adenylate kinase (AK) is a studied enzyme in the regulation of the energy balance in many organisms. It often operates at thermodynamic equilibrium. It catalyzes the following reaction: $2ADP \rightleftharpoons AMP + ATP$. It has as an equilibrium constant of 0.45. Why is this equilibrium constant dimensionless? Calculate the concentrations of ADP, AMP, and ATP at thermodynamic equilibrium when the initial conditions for ADP, AMP, and ATP are: 2 mM, 3 mM, and 5 mM. What happens to the ratio ATP/ADP when the total amount of adenosine and phosphate are independently varied from 0.2 to 10 mM. Which of them has the largest effect on this ratio?

4. Many enzymes in metabolism operate at close to thermodynamic equilibrium. Here we will study the kinetic requirements. Make a steady-state kinetic model of a linear pathway with three enzymes, each modelled with reversible Michaelis-Menten kinetics. Set the pathway substrate to 10 and the product to 1. Choose the first and the last equilibrium constant as 1000. Set all the Km's to 1 and Vmax's to 10. In the first model, you set the equilibrium constant of the second enzyme 1 and determine the Vmax of this enzyme to have it operate 10% from thermodynamic equilibrium at steady state by judging Γ/K_{eq}. In the second model, set the equilibrium constant to a 100 and determine again the value for the Vmax at which the second enzyme operates 10% away from thermodynamic equilibrium. What do you conclude? Test whether an enzyme close or far from equilibrium (10% or 90% away) has a larger or smaller effect on the steady-state flux when its Vmax is perturbed?

4.5 Enzyme inhibition: quasi-steady state approach

Enzymes are not only dependent on the concentrations of their reactants. Often, their rates are modulated by the levels of concentrations that inhibit or activate enzyme action. We can extent the mechanism underlying the Michaelis-Menten kinetic rate equation with the action of an inhibitor as shown in figure 4.3. The inhibitor can bind in principle to any of the enzyme species. At quasi-steady state conditions, the reactions between the enzyme and inhibitor will be in thermodynamic equilibrium. The conservation relationship for total enzyme then becomes,

$$\begin{aligned} e_T &= e + ei + es + esi + ep + epi \\ &= e\left(1 + \frac{i}{K_{i,4}}\right) + es\left(1 + \frac{i}{K_{i,5}}\right) + ep\left(1 + \frac{i}{K_{i,6}}\right) \end{aligned} \tag{4.39}$$

The K_i's as now defined as dissociation constants with unit concentration, e.g. mM. Using the matrix method outlined above we obtain the enzyme species from,

$$\begin{pmatrix} e \\ es \\ ep \end{pmatrix} = \begin{pmatrix} -k_1^+ s - k_3^- p & k_1^- & k_3^+ \\ k_1^+ s & -k_1 - k_2^+ & k_2^- \\ 1 + \frac{i}{K_{i,4}} & 1 + \frac{i}{K_{i,5}} & 1 + \frac{i}{K_{i,6}} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \\ -e_{tot} \end{pmatrix} \tag{4.40}$$

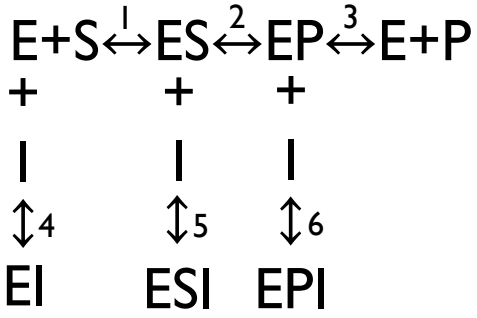

Figure 4.3: General catalytic mechanism for inhibition of an uni-uni enzyme.

The rate equation now corresponds to,

$$\begin{aligned}
\frac{v}{e_{tot}} &= \frac{num_1 s - num_2 p}{const + coef_s s + coef_p p} \\
num_1 &= k_1^+ k_2^+ k_3^+ \\
num_2 &= k_1^- k_2^- k_3^- \\
const &= \left(1 + \frac{i}{K_{i,4}}\right)(k_1^- k_2^- + k_1^- k_3^+ + k_2^+ k_3^+) \\
coef_s &= \left(1 + \frac{i}{K_{i,5}}\right) k_1^+ (k_2^- + k_2^+ + k_3^+) \\
coef_p &= \left(1 + \frac{i}{K_{i,5}}\right) k_3^- (k_1^- + k_2^- + k_2^+)
\end{aligned} \tag{4.41}$$

It is generally assumed that es and ep are indistinguishable and have the same properties such that $K_5 = K_6$. Using the same definitions as in equations 4.29 and 4.30, the enzyme kinetic properties can be expressed in terms of the those derived in the previous section,

$$\begin{aligned}
V_{MAX}^{+,APP} &= \frac{V_{MAX}^+}{1 + \frac{i}{K_{i,4}}} \\
V_{MAX}^{-,APP} &= \frac{V_{MAX}^-}{1 + \frac{i}{K_{i,4}}} \\
K_{MS}^{APP} &= \frac{1 + \frac{i}{K_{i,4}}}{1 + \frac{i}{K_{i,5}}} K_{MS} \\
K_{MP}^{APP} &= \frac{1 + \frac{i}{K_{i,4}}}{1 + \frac{i}{K_{i,5}}} K_{MP}
\end{aligned} \tag{4.42}$$

On the basis of these equations different forms of inhibition can be distinguished as shown in Table 4.1. Competitive inhibition occurs when the substrate and inhibitor can both bind in the catalytic site (K_5 absent). Mixed inhibition is when the inhibitor can compete with the substrate and bind also to the enzyme when the substrate is bound. Noncompetitive inhibition is rare, it occurs when $K_4 = K_5$. Uncompetitive inhibition is when the inhibitor can only bind to the enzyme when the substrate is bound (K_4 is absent).

Exercises

1. Is competitive or uncompetitive inhibition a more potent mechanism for inhibition?

2. Make a kinetic model of a metabolic pathway with three enzymes each catalyzing an uni-uni reaction. Make the first reaction irreversible and product-independent. Choose the other two enzymes as reversible Michaelis-Menten kinetics. Set the equilibrium constants to 100, all Km's to 1, all

Table 4.1: Different modes of inhibition.

type of inhibition	$V_{MAX}^{+,APP}$	$V_{MAX}^{+,APP}/K_{MS}^{APP}$	K_{MS}^{APP}
Competitive (K_5 absent)	V_{MAX}^{+}	$\frac{V_{MAX}^{+}/K_{MS}}{1+\frac{i}{K_{i,4}}}$	$K_{MS}\left(1+\frac{i}{K_{i,4}}\right)$
Mixed (K_4 and K_5)	$\frac{V_{MAX}^{+}}{1+\frac{i}{K_{i,5}}}$	$\frac{V_{MAX}^{+}/K_{MS}}{1+\frac{i}{K_{i,4}}}$	$K_{MS}\frac{1+\frac{i}{K_{i,4}}}{1+\frac{i}{K_{i,5}}}$
Pure noncompetitive ($K_4 = K_5$)	$\frac{V_{MAX}^{+}}{1+\frac{i}{K_{i,5}}}$	$\frac{V_{MAX}^{+}/K_{MS}}{1+\frac{i}{K_{i,4}}}$	K_{MS}
Uncompetitive (K_4 absent)	$\frac{V_{MAX}^{+}}{1+\frac{i}{K_{i,5}}}$	V_{MAX}^{+}/K_{MS}	$\frac{K_{MS}}{1+\frac{i}{K_{i,5}}}$

V_{max}^{+} to 100, the fixed pathway substrate to 10 and the fixed pathway product to 1. Show that the steady-state flux through this pathway is only sensitive to the enzyme level of the first enzyme and not to the second and the third. Test this and explain why this occurs. Introduce competitive inhibition of the first enzyme by the substrate of the third enzyme. Figure out which enzyme level can change the steady-state flux most. Explain your findings. Equip the same model with uncompetitive inhibition. Test whether this inhibition is more potent inhibition mechanism. When do you conclude one of the two mechanisms is more potent? Think carefully about a fair comparison of the two models. Do you want the models to have the same reference steady state and K_i for the inhibition?

4.6 Equilibrium binding models and convenience kinetics

An alternative and much more straightforward approach than the quasi-steady state approximation to deriving enzyme kinetics is by using equilibrium binding models. The disadvantage is that they are more approximate but often they result in rate equations that have nearly the same mathematical properties and they are also in accordance with thermodynamics, as the quasi-steady state approximation.

The simplest method to derive equilibrium binding models is to start from the conservation equation of total enzyme. Let's start with the simplest example (see also equation 4.19),

$$E + S \overset{v_1}{\rightleftharpoons} ES \overset{v_2}{\rightleftharpoons} EP \overset{v_3}{\rightleftharpoons} E + P \tag{4.43}$$

The enzyme conservation equation equals,

$$e_T = e + es + ep \tag{4.44}$$

Reactions 1 and 3 are assumed to be in thermodynamic equilibrium and the rate determining reaction is reaction 2. This assumption entails that reaction 1 and 3 are much faster than the conversion of $es \rightleftharpoons ep$. Using this assumption we can equate the enzyme-reactant complexes in terms of the reactant and free enzyme concentration through the definition of the dissociation constant,

$$e_T = e\left(1 + \frac{s}{K_S} + \frac{p}{K_P}\right) \tag{4.45}$$

K_S (is $K_1 = k_1^+/k_1^-$) and K_P (is $K_3 = k_3^+/k_3^-$) are dissociation constants and play the role of affinity constants in equilibrium binding models as we shall see shortly. The rate of the reaction is given by,

$$v = k_2^+ es - k_2^- ep \tag{4.46}$$

Using the dissociation constant definition and equation 4.45 we obtain for the rate equation,

$$v = \frac{V_{MAX}^+ \frac{s}{K_S} - V_{MAX}^- \frac{p}{K_P}}{1 + \frac{s}{K_S} + \frac{p}{K_P}} \tag{4.47}$$

The maximal rates are now defined as $V_{MAX}^+ = k_2^+ e_T$ and $V_{MAX}^- = k_2^- e_T$. This equation has the same form as the reversible Michaelis-Menten equation but the affinity constants have a different meaning!

The power of equilibrium models derives from its straightforward derivation of multi-reactant rate equations. For instance, consider the enzyme mechanisms shown in figure 4.4. For mechanism A, we have the following conservation relation,

$$e_T = e + ea + eb + eab + ep + eq + epq \tag{4.48}$$

and for mechanism B,

$$e_T = e + ae + eb + aeb + aeq + pe + eq + peb + peq \tag{4.49}$$

Using the equilbrium binding assumption we obtain for mechanism A,

$$e_T = e\left(1 + \frac{a}{K_a} + \frac{b}{K_b} + \frac{a \cdot b}{K_a K_b} + \frac{p}{K_p} + \frac{q}{K_q} + \frac{p \cdot q}{K_p K_q}\right) \tag{4.50}$$

whereas for mechanism B the expression can be simplified to,

$$e_T = e\left(1 + \frac{a}{K_a} + \frac{p}{K_p}\right)\left(1 + \frac{b}{K_b} + \frac{q}{K_q}\right) \tag{4.51}$$

The last equation has a straightforward interpretation. Each term within brackets corresponds to the saturation and competition characteristics of one binding pocket. Pocket 1 can bind either be empty, bind A or P and pocket 2 can be empty or bind B or Q. As the rate of reaction A equals $v = k^+ eab - k^- epq$ the rate equation becomes,

$$v = \frac{V_{MAX}^+ \frac{a \cdot b}{K_a K_b} - V_{MAX}^- \frac{p \cdot q}{K_p K_q}}{1 + \frac{a}{K_a} + \frac{b}{K_b} + \frac{a \cdot b}{K_a K_b} + \frac{p}{K_p} + \frac{q}{K_q} + \frac{p \cdot q}{K_p K_q}} \tag{4.52}$$

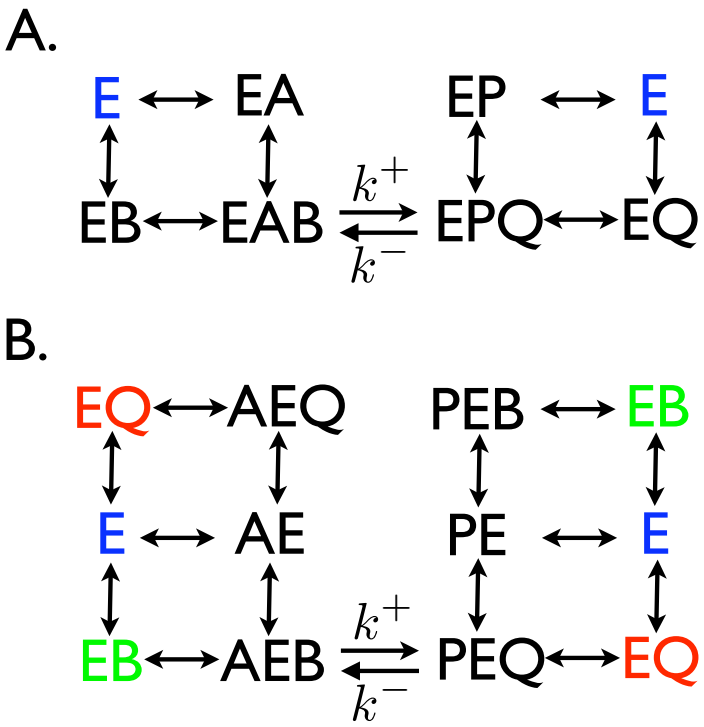

Figure 4.4: Two catalytic mechanism for a two-substrate and two-product reaction, a so-called bi-bi reaction. The equilibrium binding models will be different for the two mechanisms. Species that appear twice in the mechanisms are coloured. A. The enzyme has a catalytic site where only substrates and products can bind independently. B. The enzyme has a catalytic site where there occurs competition for a two binding pocket; A competes with P for one binding pocket and B with Q for the other. We assume here that the affinity of the reactants does not depend on the identity of the reactant that is already bound (or not) to the enzyme.

(With $V^+_{MAX} = k^+ e_T$ and $V^-_{MAX} = k^- e_T$.) The rate equation for mechanism B equals,

$$v = \frac{V^+_{MAX} \frac{a \cdot b}{K_a K_b} - V^-_{MAX} \frac{p \cdot q}{K_p K_q}}{\left(1 + \frac{a}{K_a} + \frac{p}{K_p}\right)\left(1 + \frac{b}{K_b} + \frac{q}{K_q}\right)} \tag{4.53}$$

These equations have the same Haldane relationship,

$$K_{EQ} = \frac{p_{EQ} q_{EQ}}{a_{EQ} b_{EQ}} = \frac{V^+_{MAX} K_p K_q}{V^-_{MAX} K_a K_b} \tag{4.54}$$

(The subscript EQ denotes equilibrium concentrations.) The numerator of both

rate equations can expressed in terms of the equilibrium constant in the same way as the reversible Michaelis-Menten equation as,

$$V^+_{MAX}\frac{a\cdot b}{K_aK_b}\left(1-\frac{p\cdot q}{a\cdot b\cdot K_{EQ}}\right)=V^+_{MAX}\frac{a\cdot b}{K_aK_b}\left(1-\frac{\Gamma}{K_{EQ}}\right) \tag{4.55}$$

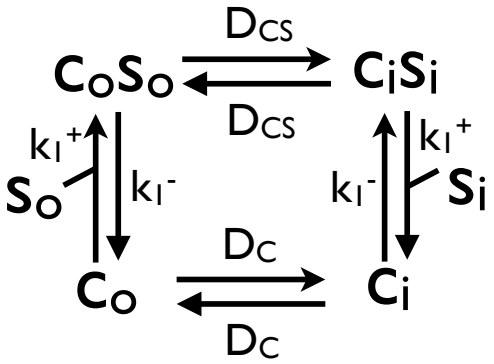

Figure 4.5: Mechanism for a membrane transporter acting as a facilitated diffusion carrier. The transport solute occurs at the external cell side with concentration, s_o, where it binds the carrier, occurring at concentration, c_o. As the carrier diffuses continuously through the membrane it sometimes occurs at the intracellular side, at concentration $c_i s_i$, where it can deposit the solute.

Exercises

In figure 4.5, an enzyme mechanism for a transporter is displayed. A molecule, S, is transported from the external to the intracellular medium, with concentrations s_o and s_i, respectively. The binding reactions of the molecule to the carrier at the extra- and intracellular side of the membrane are assumed to be at equilibrium. The rate of the reaction is determined by the diffusion of the carrier through the membrane. Given those assumptions write the rate equation for the transporter in the following form,

$$v=V_{max}\frac{\frac{s_o}{K_m}-\frac{s_i}{K_m}}{1+\frac{s_o}{K_m}+\frac{s_o}{K_m}+K_i\frac{s_is_o}{K_m^2}} \tag{4.56}$$

and determine the constants K_m, K_i and V_{max} in terms of $K_1=k_1^-/k_1^+$, D_C and D_{CS}. Study the effect of product inhibition, s_i, on the normalized uptake rate, v/V_{max}. Set s_0 to 5 mM and K_m to 1.19 mM. Those numbers are realistic for yeast, which is known to have this transporter mechanism for its glucose carrier. What is the role of K_i; when is the inhibition reduced and uptake

rate high? K_i has been shown to equal 0.91 in yeast. Does this K_i facilitate glucose uptake in yeast? Determine the equilibrium constant for this enzyme. Study the conditions for high-sensitivity for the external level of solute and a high maximal rate; equate the ratio V_{max}/K_m to do so. Take into account the Haldane relationship.

4.7 Cooperative enzymes

4.7.1 The regulatory potential of cooperative enzymes

Enzymes composed out of multiple subunits are termed multimeric enzymes. In the regulation of metabolism they play pivotal roles. Classical examples in catabolism are pyruvate kinase and phosphofructokinase. In multimeric enzymes the subunits can be identical to each other or not. The presence of multiple subunits introduces the possibility that the kinetics of one subunit depends on the binding state of other subunits within the same enzyme. Such multimeric enzymes are called cooperative enzymes or allosteric enzymes.

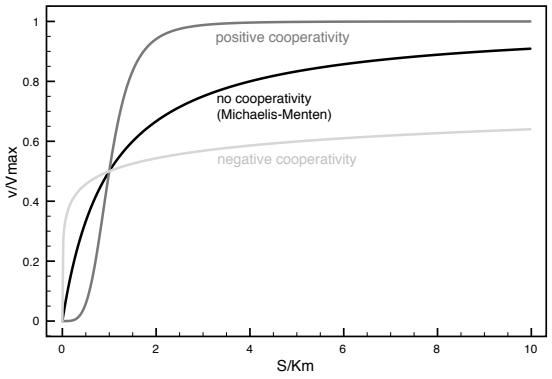

Figure 4.6: Cooperative enzymes and regulation. The saturation of three enzymes, v/V_{MAX} (their normalized rates), is plotted as function of their normalized substrate concentrations, S/K_M. They differ markedly in their sensitivity to the concentration of their substrate. The enzyme displaying positive cooperativity is clearly most sensitive. The enzyme with negative cooperativity is least sensitive to its substrate. A Michaelis-Menten enzyme, which has a hyperbolic substrate-rate dependency, has intermediate sensitivity. A convenient definition of sensitivity is reciprocal value of the concentration change required to change the rate of an enzyme from 10% to 90% of saturation, v/V_{MAX}.

Why are cooperative enzymes such good regulatory devices in metabolism? This becomes apparent when we consider the sensitivity of a Michaelis-Menten enzyme to its substrate. In figure 4.6 the rate of an enzyme following Michaelis-Menten kinetics is plotted as function of the substrate concentration (the black

line). An 81-fold change in the substrate concentration is required to change the rate from 10% to 90% of its maximal value: we obtain from the Michaelis-Menten equation,

$$s_{v/V_{MAX}} = K_M \frac{v/V_{MAX}}{1 - v/V_{MAX}} \tag{4.57}$$

Calculation of $s_{0.9}/s_{0.1}$ gives 81! A Michaelis-Menten enzyme needs an enormous change in its substrate to have a 9-fold change in its flux. So it is very hard to tune the rate of such an enzyme through regulation. In 1910, Hill studied the oxygen binding kinetics of hemoglobine and found a sigmoidal saturation curve, as the gray curve marked with positive cooperativity in figure 4.6. He fitted the following phenomenological equation to this relationship, now known as the Hill equation,

$$y = \frac{x^h}{K_{0.5}^h + x^h} \tag{4.58}$$

Here h is defined as the Hill coefficient and $K_{0.5}$ as the value for x at which y equals 0.5. If $h = 1$ this relationship is identical to a Michaelis-Menten function. One should realize however that the Michaelis-Menten equation has a physical basis and the Hill equation does not! The Hill equation however allows us to define a sensitivity index (or cooperativity index). We use,

$$x_{v/V_{MAX}} = K_M \left(\frac{y}{1 - y} \right)^{1/h}$$

to define the sensitivity index,

$$R = \frac{x_{0.9}}{x_{0.1}} = 81^{1/h} \tag{4.59}$$

This equation indicates that if the Hill coefficient is 1 a 81-fold increase in x is required increase y from 0.1 to 0.9, i.e. from 10 to 90% of the maximal output. The Hill curve becomes sigmoidal when $h > 1$. Below we will learn that this corresponds to positively cooperating subunits in a multi-subunit enzyme.

Enzymes that display positive cooperativity can have a heightened sensitivity to reactants and effectors such that small changes in their concentrations bring about large adjustments in the catalysis rate. This gives cooperative enzymes their high regulatory potential. Large change in enzyme rate can occur due to small changes in the concentrations of metabolites; the metabolites remain nearly homeostatic despite a large rate change through their pools. Negative cooperativity causes an enzyme to be very insensitive to metabolites, which is another useful property.

4.7.2 The Monod Wyman Changeux model for cooperative enzyme kinetics

In this section we will consider the model presented by Monod, Wyman, and Changeux (MWC) in 1965. The MWC model assumes the multimeric enzyme

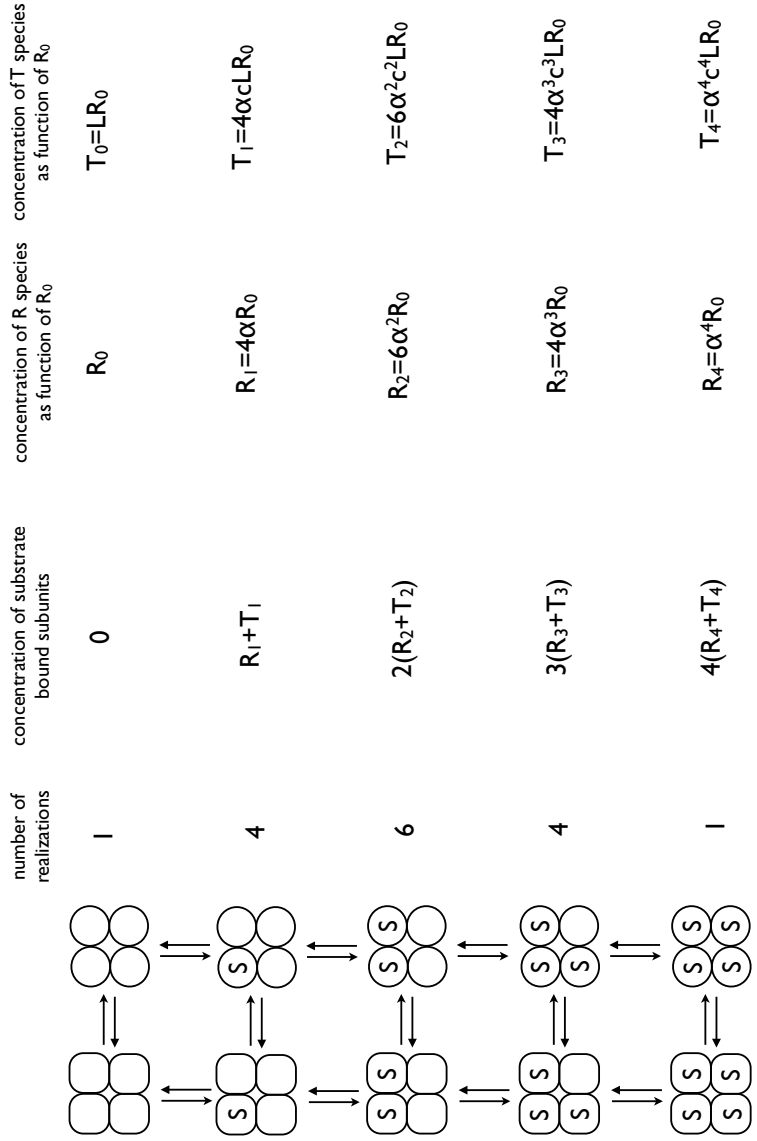

Figure 4.7: MWC scheme

to be composed out of n identical subunits that can each be in a T (taut or tight) state, which has a low affinity for the substrate S, or a R (relaxed) state, which has a high affinity for the substrate. In addition, it is assumed that the subunits are in equilibrium between their R and T state and that all subunits change from R to T or from T to R in a concerted fashion (=at the same time). The MWC model is sometimes referred to as the concerted-symmetry model.

It is an equilibrium-binding model.

The model presented by Koshland & Nemethy [24] does not make the concerted-symmetry assumption. It is therefore more general and, regretfully so, more complicated. The MWC model as it was originally presented only considered single substrate kinetics and an irreversible reaction. For reversible models, the reader is referred to Hofmeyr & Cornish-Bowden [15] (see next section) and Popova & Selkov [27].

The following reactions are considered in the MWC model,

$$\begin{array}{rcll} T_0 & \overset{L}{\rightleftharpoons} & R_0 & \\ T_0 + S & \overset{K_T}{\rightleftharpoons} & T_1 & R_0 + S \overset{K_R}{\rightleftharpoons} R_1 \\ T_1 + S & \overset{K_T}{\rightleftharpoons} & T_2 & R_1 + S \overset{K_R}{\rightleftharpoons} R_2 \\ & \vdots & & \vdots \\ T_{n-1} + S & \overset{K_T}{\rightleftharpoons} & T_n & R_{n-1} + S \overset{K_R}{\rightleftharpoons} R_n \end{array} \tag{4.60}$$

The dissociation constant is given above the reaction arrow and T_j and R_j denote the oligomers with j of their n subunits bound to the substrate S. We define the following kinetic constants,

$$\begin{aligned} \alpha &= \frac{S}{K_R} \\ L &= \frac{T_0}{R_0} \\ c &= \frac{K_R}{K_T} \end{aligned} \tag{4.61}$$

The derivation of the MWC model can be illustrated nicely with an example of a cooperative enzyme with four subunits (figure 4.7). The activity of the enzyme is given by,

$$\begin{aligned} v &= V_{MAX} \cdot \frac{\text{concentration of all substrate bound subunits}}{\text{concentration of all subunits}} \\ &= V_{MAX} \cdot \frac{4(R_1 + 3R_2 + 3R_3 + R_4 + T_1 + 3T_2 + 3T_3 + T_4)}{4(R_0 + R_1 + 3R_2 + 3R_3 + R_4 + T_0 + T_1 + 3T_2 + 3T_3 + T_4)} \\ &= V_{MAX} \cdot \frac{\alpha + 3\alpha^2 + 3\alpha^3 + \alpha^4 + \alpha cL + 3\alpha^2 c^2 L + 3\alpha^3 c^3 L + \alpha^4 c^4 L}{1 + \alpha + 3\alpha^2 + 3\alpha^3 + \alpha^4 + cL + \alpha cL + 3\alpha^2 c^2 L + 3\alpha^3 c^3 L + \alpha^4 c^4 L} \\ &= \frac{\alpha(1+\alpha)^2 + \alpha L c(1+c\alpha)^2}{(1+\alpha)^3 + L(1+c\alpha)^3} \end{aligned} \tag{4.62}$$

In the general case, we then obtain,

$$\begin{aligned}\frac{v}{V_{MAX}} &= \text{fraction of subunits bound to S} \\ &= \frac{\text{total amount of subunits bound to S}}{\text{total amount of subunits}} \\ &= \frac{\overbrace{\alpha\left(1+\alpha\right)^{n-1}}^{\text{amount of R states bound to S}} + \overbrace{L\cdot c\cdot\alpha\left(1+c\cdot\alpha\right)^{n-1}}^{\text{amount of T state bound to S}}}{\underbrace{\left(1+\alpha\right)^{n}}_{\text{total amount of R state}} + \underbrace{L\cdot\left(1+c\cdot\alpha\right)^{n}}_{\text{total amount of T state}}}\end{aligned} \tag{4.63}$$

The maximal rate of the enzyme equals $V_{MAX} = n{\cdot}k_{cat,R}{\cdot}e_T$. The difference between the R and T state disappears if $c = 1$ and $L = 1$. If $L = 0$ or $c = 1$ the equation simplifies to $v = \frac{V_{MAX}S}{S+K_R}$.

An activator and inhibitor can be defined to have an effect of the L coefficient as,

$$L^{APP} = L\frac{(1+\beta)^n}{(1+\gamma)^n} \tag{4.64}$$

With $\gamma = A/K_A$ as the activator and $\beta = I/K_I$ as the inhibitor term. Competitive inhibition can be expressed as,

$$v = \frac{v}{V_{MAX}}\frac{\alpha\left(1+\alpha+\beta\right)^{n-1} + L\cdot c\cdot\alpha\left(1+c\cdot\alpha\right)^{n-1}}{\left(1+\alpha+\beta\right)^{n} + L\cdot\left(1+c\cdot\alpha\right)^{n}} \tag{4.65}$$

So far, we have assumed that the R and T state have the same V_{MAX}. We have only taken into account differences in substrate affinity. Such systems are called V systems.

Exercises

1. Pyruvate kinase is a well-known cooperative enzyme in glycolysis of many organisms following the MWC mechanism. It catalyzes the following reaction $phosphoenolpyruvate + ADP \rightleftharpoons ATP + pyruvate$. Here we will abbreviate phospoenolpyruvate as *pep* and pyruvate as *pyr*. In *Escherichia coli*, the rate equation for this mechanism is,

$$v = V_{MAX}\frac{pep\cdot adp\left(\frac{pep}{K_{pep}}+1\right)^n}{K_{pep}\left(L\left(\frac{1+\frac{atp}{K_{atp}}}{\frac{fdp}{K_{fdp}}+\frac{amp}{K_{amp}}+1}\right)^n + \left(\frac{pep}{K_{pep}}+1\right)^n\right)\left(adp+K_{adp}\right)} \tag{4.66}$$

   The kinetic parameters are: $K_{pep} = 0.31\ mM$, $K_{adp} = 0.26\ mM$, $K_{amp} = 0.2\ mM$, $K_{fdp} = 0.19\ mM$, $K_{atp} = 22.5\ mM$, $L = 1000$, and $n = 4$. Physiological values for *pep*, *atp*, *adp*, *amp* and *fdp* are: 2.7, 4.2, 0.6, 1, and 0.27 mM. Determine whether the regulatory influences of *amp* and *fdp* are activating or inhibiting. *Fdp* is a glycolytic intermediate in the

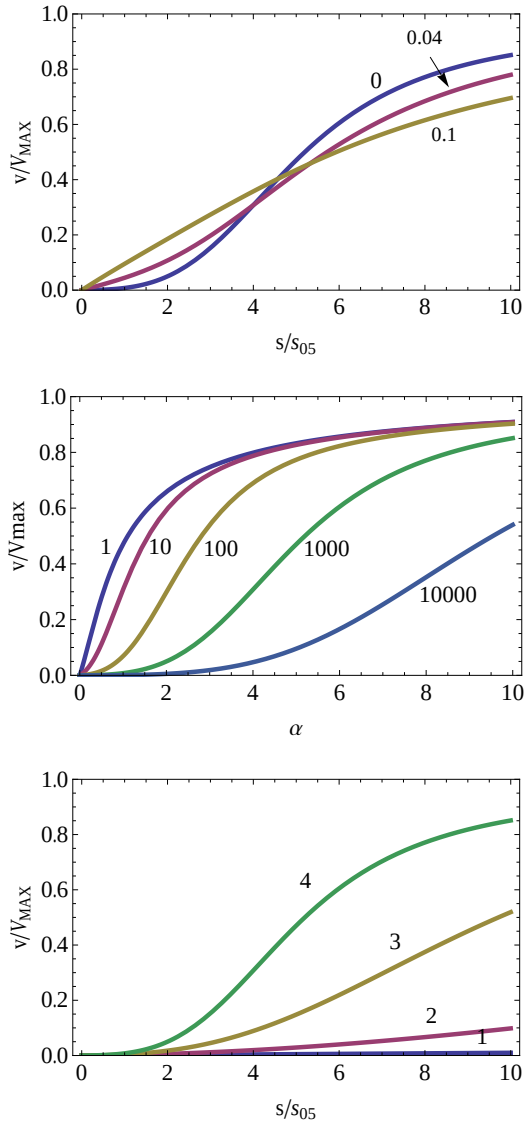

Figure 4.8: Illustrations of the consequence of cooperativity parameters on the MWC rate equation (equation 4.63). In the upper, middle and lower plot, the value of the c, L, and n parameter were changed, respectively.

upper part of glycolysis; it exert a feedforward regulation on pyruvate kinase (see figure 1.1). K_{amp} was set to an arbitrary value. Determine the effect of the chosen value on the rate equation of pyruvate kinase.

2. The intricate regulation of glycolysis and occurrence of the cooperative en-

zyme, phosphofructokinase, pyruvate decarboxylase, and puryvate kinase, inspired the analysis of kinetic models of glycolysis. A glycolysis model was developed by Goldbeter and Lefever [12]. It was a simplified model of glycolysis that illustrated the potential important role of product activation of phosphofructokinase (PFK) by ADP. PFK catalyzes the following reaction: $fructose-6-phosphate + ATP \rightleftharpoons ADP + fructose-1,6-bisphosphate$. In figure 4.9 the network diagram of this model is shown. Here γ indicates ADP and α, fructose-6-phosphate. PFK is an allosteric enzyme, modelled with a MWC mechanism,

$$v_2 = \sigma_M \frac{\alpha e(1+\alpha e)^{n-1}(1+\gamma)^n + L\theta\alpha c e'(1+\alpha c e')^{n-1}}{L(1+\alpha c e')^n + (1+\gamma)^n(1+\alpha e)^n} \tag{4.67}$$

Where $e = (1+\epsilon)^{-1}$ and $e' = (1+\epsilon')^{-1}$ with ϵ and ϵ' as relative catalytic constants of the T and R states. The first rate v_1 is fixed to 0.7 and $v_2 = k_s\gamma$, with $k_s = 0.1$. The other parameters are: $\epsilon = 0.1$, $\epsilon' = 0.1$, $L = 10^6$, $c = 10^{-5}$, $\sigma_M = 5$, and $\theta = 1$. Confirm that ADP activates PFK by studying its rate curves. Simulate this model for various value of the Hill coefficient (take reference value 2). Choose as initial conditions: $\alpha(0) = 40$ and $\gamma(0) = 0.8$. What is effect of the removal of the activation?

In figure 4.10 complicated dynamics is shown induced by regulation of cooperative enzymes. Similar complex dynamics has been observed in in vitro studies on glycolysis. The current view is that under physiological conditions, chaos and complex oscillations can be ruled out and would be hazardous for cells. Regular glycolytic oscillations have been shown for yeast, but again under particular and unphysiological conditions.

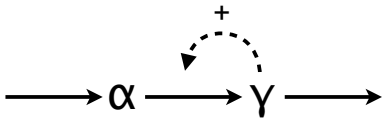

Figure 4.9: Network diagram for the simplified model developed by Goldbeter and Lefever [12] to study the role of the product activation of PFK by ADP in glycolysis. In this model, PFK was modelled according to a MWC mechanism.

4.7.3 The reversible Hill Equation

The cooperative enzymes we have treated in the previous section can be sometimes unrealistic models. They do not describe reversible enzymes and only take into account the action of the substrate. A more realistic model would describe cooperative enzymes as having reversible rates, which are sensitive to the concentrations of substrates, products, and effectors. Such a model would

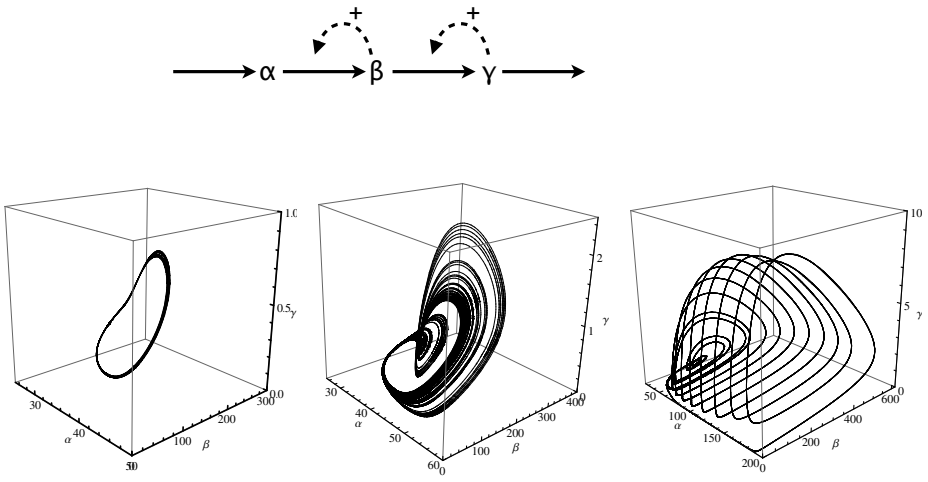

Figure 4.10: Network diagram and dynamics of an illustrative model for complex dynamics as developed by Decroly and Goldbeter [8]. The three plots with dynamics differ in the value of a first-order rate constant k_s for the degradation of γ. The plot of the left (for $k_s = 1.9$ indicates that the values of the three variables as function of time settle onto a so-called limit cycle; oscillation with a single period. At a slightly higher value for k_s, at 2, the dynamics becomes chaotic and settles onto a strange attractor (middle plot). When, k_s equals 2.032 the systems displays complex oscillations. The rate equations for this model are: $v_1 = v/K_{m1}$, $v_2 = \frac{\alpha(1+\alpha)(1+\beta)^2}{L_1+(1+\alpha)^2(1+\beta)^2}$, $v_2 = \frac{\beta(1+d\beta)(1+\gamma)^2}{L_1+(1+d\beta)^2(1+\gamma)^2}$, and $v_4 = k_s\gamma$. The differential equations are: $d\alpha/dt = v/K_{m1} - \sigma_1\theta, d\beta/dt = q_1\sigma_1\theta - \sigma_2\eta$ and $d\gamma/dt = q_2\sigma_2\eta - k_s\gamma$. The following constants were used: $v/K_{m1} = 0.45$, $\sigma_1 = \sigma_2 = 10$, $q_1 = 50$, $q_2 = 0.02$, $L_1 = 5 * 10^8$, $L_2 = 100$, and $d = 10^{-7}$. This is a sufficient description of the system to play with it yourself!

become very complicated to derive and to handle as it would depend on a large number of parameters. In addition, the experimental determination of such mechanisms would require an enormous amount of experiments. The number of experiments to determine the kinetics of yeast phosphofructokinase was about 600. Hofmeyr and Cornish-Bowden took up the challenge to derive a reversible product-sensitive cooperative enzyme kinetics that does not suffer from a great number of parameters [15]. They named it the reversible Hill equation to emphasize its two characteristics; it's reversible and phenomenological, as Hill's original equation.

The derivation of this equation is straightforward. It assumes extreme cooperativity such only the free enzyme or the fully saturated enzyme species exist; either all or none of the binding sites are occupied. We will illustrate the derivation for an enzyme with two subunits (figure 4.11). Each of the subunits catalyzes the reversible uni-uni reaction from S to P. The total enzyme

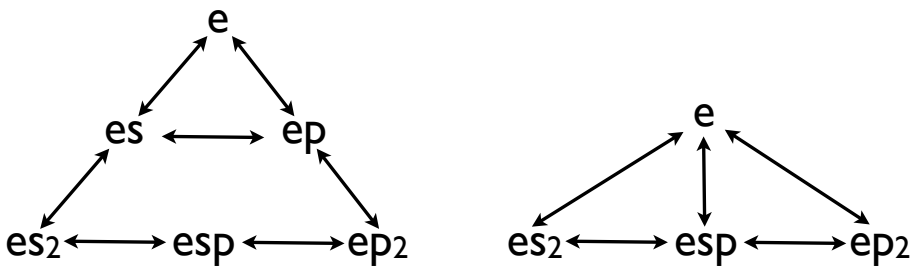

Figure 4.11: Illustration of the assumption in the derivation of the reversible Hill equation for an enzyme with two subunits each catalyzing the reversible conversion of S into P. On the left, the complete mechanism is shown and on the right the reduced mechanism as assumed for the reversible Hill equation. The enzyme is either in its free state, with none of its subunits bound to reactants, or all binding sites are occupied. This assumption means extreme cooperativity between the subunits.

concentrations obeys,

$$e_T = e + es_2 + 2esp + ep_2 \tag{4.68}$$

It was assumed that the concentrations of es and ep are negligibly small. This is the essential assumption in the derivation of the reversible Hill equation. It amounts to assuming extreme cooperativity between the subunits; if one subunit has bound a substrate the other has near infinite affinity for the substrate. The 2 in front of esp denotes the two forms of this species, e.g. with s and p once on the first and the second binding site and vice versa for the other form. The rate of the enzyme then corresponds to,

$$v = k^+(2es_2 + 2esp) - k^-(2ep_2 + 2esp) \tag{4.69}$$

The enzyme substrate species are assumed to be in thermodynamic equilibrium,

$$\begin{aligned} es_2 &= \frac{e \cdot s^2}{s_{0.5}^2} \\ esp &= \frac{e \cdot s \cdot p}{s_{0.5} p_{0.5}} \\ es_2 &= \frac{e \cdot p^2}{p_{0.5}^2} \end{aligned} \tag{4.70}$$

$p_{0.5}$ and $s_{0.5}$ are at this point defined phenomenologically. The term $s_{0.5}^2$ has to equal αK_s^2 with α as a cooperative interaction coefficient with K_s as the dissociation constant of a single binding site for s. Accordingly, $s_{0.5}$ and $p_{0.5}$ are equal to $\sqrt{\alpha} K_s$ and $\sqrt{\alpha} K_p$, respectively. To ascertain that the es and

ep are negligibly small α needs to be much smaller than 1. The free enzyme concentration can now be equated,

$$e = \frac{e_T}{1 + \frac{s^2}{s_{0.5}^2} + 2\frac{s \cdot p}{s_{0.5} p_{0.5}} + \frac{p^2}{p_{0.5}^2}} \tag{4.71}$$

and the other species as well, for instance,

$$es_2 = \frac{e_T}{1 + \frac{s^2}{s_{0.5}^2} + 2\frac{s \cdot p}{s_{0.5} p_{0.5}} + \frac{p^2}{p_{0.5}^2}} \frac{s^2}{s_{0.5}^2} \tag{4.72}$$

The rate of the enzyme now becomes,

$$\begin{aligned} v &= \frac{2k^+ e_T \left(\frac{s^2}{s_{0.5}^2} + \frac{s \cdot p}{s_{0.5} p_{0.5}}\right) - 2k^- e_T \left(\frac{p^2}{p_{0.5}^2} + \frac{s \cdot p}{s_{0.5} p_{0.5}}\right)}{1 + \frac{s^2}{s_{0.5}^2} + 2\frac{s \cdot p}{s_{0.5} p_{0.5}} + \frac{p^2}{p_{0.5}^2}} \\ &= \frac{2k^+ e_T \frac{s}{s_{0.5}} \left(\frac{s}{s_{0.5}} + \frac{p}{p_{0.5}}\right) - 2k^- e_T \frac{p}{p_{0.5}} \left(\frac{p}{p_{0.5}} + \frac{s}{s_{0.5}}\right)}{1 + \frac{s^2}{s_{0.5}^2} + 2\frac{s \cdot p}{s_{0.5} p_{0.5}} + \frac{p^2}{p_{0.5}^2}} \\ &= \frac{\left(V_{MAX}^+ \frac{s}{s_{0.5}} - V_{MAX}^- \frac{p}{p_{0.5}}\right) \left(\frac{s}{s_{0.5}} + \frac{p}{p_{0.5}}\right)}{1 + \left(\frac{s}{s_{0.5}} + \frac{p}{p_{0.5}}\right)^2} \\ &= \frac{V_{MAX}^+ \frac{s}{s_{0.5}} \left(1 - \frac{p}{sK_{EQ}}\right) \left(\frac{s}{s_{0.5}} + \frac{p}{p_{0.5}}\right)}{1 + \left(\frac{s}{s_{0.5}} + \frac{p}{p_{0.5}}\right)^2} \end{aligned} \tag{4.73}$$

Here the maximal forward and backward rate are defined as, $V_{MAX}^+ = 2k^+ e_T$ and $V_{MAX}^- = 2k^- e_T$.

Hofmeyr and Cornish-Bowden have generalized equation 4.73 to enzymes with n subunits,

$$v = \frac{V_{MAX}^+ \frac{s}{s_{0.5}} \left(1 - \frac{p}{sK_{EQ}}\right) \left(\frac{s}{s_{0.5}} + \frac{p}{p_{0.5}}\right)^{n-1}}{1 + \left(\frac{s}{s_{0.5}} + \frac{p}{p_{0.5}}\right)^n} \tag{4.74}$$

If $p = 0$ the Hill equation is obtained. Note that the exponent does have a physical meaning in the Hofmeyr & Cornish-Bowden derivation, whereas in the original equation derived by Hill it did not.

Activation and inhibition can be incorporated into this equation,

$$v = \frac{V_{MAX}^+ \frac{s}{s_{0.5}} \left(1 - \frac{p}{sK_{EQ}}\right) \left(\frac{s}{s_{0.5}} + \frac{p}{p_{0.5}}\right)^{n-1}}{\frac{1 + \left(\frac{x}{x_{0.5}}\right)^n}{1 + \beta\left(\frac{x}{x_{0.5}}\right)^n} + \left(\frac{s}{s_{0.5}} + \frac{p}{p_{0.5}}\right)^n} \tag{4.75}$$

If $\beta < 1$ the effector acts as an inhibitor and when $\beta > 1$ it becomes an activator.

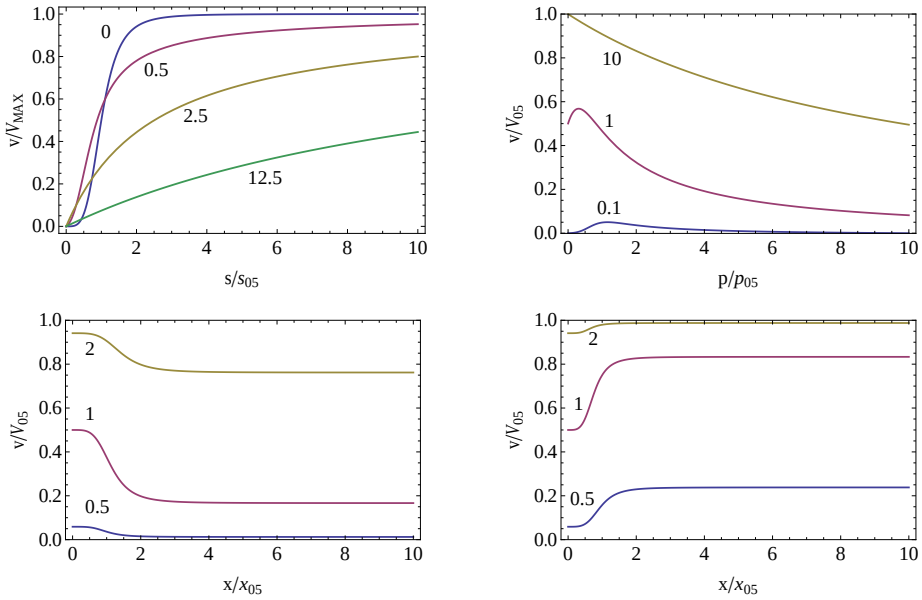

Figure 4.12: Numerical analysis of the reversible Hill rate equation (equation 4.75). In the upper left plot, product inhibition is illustrated (p was varied). The upper right plot displays product inhibition for various values of the substrate, s. The lower two plots show the effect of the β parameter ($p = 0$); on the left it equals 0.2 and on the right 5. In all plots, $s_{0.5} = p_{0.5} = 1$ and $n = 4$. $x_{0.5} = 1$ in all plots except the upper left plot where it was set to 0.1. The equilibrium constant was set to 10^6 in all plots except the upper right plot where it equals 100.

Exercises

Negative feedback regulation in metabolic pathways has profound influences on homeostasis and which enzymes influence the steady-state flux most. Make a kinetic model of three enzymes with the second and the third following reversible Michaelis-Menten kinetics with a V_{max} of 1000, K_M's of 1, and an equilibrium constant of 10. To allow for steady state the pathway substrate S and product P are fixed. We label the three variable metabolites in the pathway as x_1 to x_3. The first enzyme in the pathway is an enzyme following the reversible Hill rate equation, which is inhibited by the fixed final product of the pathway, p,

$$v = \frac{\frac{V_f s}{s_{0.5}}\left(1 - \frac{s}{x_1 K_{eq}}\right)\left(\frac{s}{s_{0.5}} + \frac{x_1}{x_{1,0.5}}\right)^{n-1}}{\left(\frac{s}{s_{0.5}} + \frac{x_1}{x_{1,0.5}}\right)^{n} + \frac{1+\left(\frac{p}{p_{0.5}}\right)^{n}}{1+\alpha\left(\frac{p}{p_{0.5}}\right)^{n}}} \tag{4.76}$$

The parameters for this enzyme are: $x_{1,0.5} = 10^4$, $p_{0.5} = 1$, $n = 4$, and $\alpha = 0.0001$, $V_f = 200$, and $K_{eq} = 400$. Set s equal to 1. The first enzyme has

been parameterized such that it is not very sensitive to its product, x_1. How was this achieved? Test your hypothesis by studying the enzyme in isolation of the pathway. Explain why the first enzyme, in the absence of the feedback, determines the steady state flux when it has little or no sensitivity towards its immediate product ? At what concentration of P will this entire pathway operate at thermodynamic equilibrium? What are the equilibrium concentration of the metabolic intermediates? Verify your hypothesis using the model and by calculating those concentrations by hand. Make a log-log plot of the steady-state flux as function of the fixed product concentration. This is called a rate characteristic. Let the fixed product concentration change from very small to its equilibrium value. Explain what you see. Change n and $p_{0.5}$ to determine how the feedback influences curve? What do you conclude? Vary the value of $s_{0.5}$. What is the immediate influence on the first enzyme? How does it influence the shape of the rate characteristic? Suppose now that the product P is consumed by a fourth enzyme, following $10p/(0.01 + p)$ as rate equation. Add this curve to the plot. Which enzyme has the largest influence on the steady-state flux - which enzyme control the flux the most - when its level is changed? How does this conclusion depend on the strength of the feedback? Homeostasis of a metabolite can be defined as little changes in its concentration over a range of steady states while the flux through this metabolite changes very much. When is P more homeostatic with weak or strong feedback? Write in a single sentence your conclusion about the interplay between homeostasis, flux control, and negative feedback. You can read more about these issues in references [16, 17].

Chapter 5

Stoichiometric network analysis

5.1 Introduction

Kinetic models require information about the initial concentrations of the variable intermediates, environment, reaction kinetics, thermodynamics, and stoichiometry. They allow for the calculation of the dynamics of the concentrations of molecules in networks and their dependencies on parameters. Very useful information for many biological studies. But in many cases not all of that information is available. What to do in such cases? For metabolic networks a number of system properties can be found with stoichiometric network analysis (SNA). Signaling and gene networks can analyzed with graph theoretical methods, which will not be discussed here.

Stoichiometric models solely consider the stoichiometry of a metabolic network. Often, SNA requires in addition a specification of the nutrient availability in the environment. Stoichiometry can be a potent constraint for the behavior of metabolic networks as we shall see. Collectively, all the methods developed for the study of stoichiometric models have been termed stoichiometric network analysis. SNA has grown into quite a large and active field with many different methods, often with a biotechnological application in mind.

In this chapter, we will introduce some of the basics of SNA: conservation relationships, independent fluxes, flux modes, flux space, flux balance analysis and flux variability analysis.

5.2 The stoichiometric matrix

The mathematical description of a kinetic model with r reactions and m intermediates is given by (an example is introduced below),

$$\frac{d}{dt}\mathbf{x}(t, \mathbf{p}, \mathbf{x_0}) = \mathbf{N}\mathbf{v}(\mathbf{x}(t, \mathbf{p}, \mathbf{x_0}), \mathbf{p}) \tag{5.1}$$

with: $\mathbf{x}$ as the $m \times 1$ concentration (or state) vector, $\mathbf{N}$ as the $m \times r$ stoichiometric matrix, $\mathbf{x}_0$ as the $m \times 1$ vector of initial conditions ($\mathbf{x_0} = \mathbf{x}(0, \mathbf{p})$), t as time, and $\mathbf{p}$ as the parameter vector. Often equation 5.1 is written in shorthand notation as $d\mathbf{x}/dt = \mathbf{N}\mathbf{v}(\mathbf{x}, \mathbf{p})$ or $d\mathbf{x}/dt = \mathbf{N}\mathbf{v}$.

For example consider the network show in Figure 5.1. The stoichiometric matrix is given by (we have indicated the names of the columns and rows here, this is in the definition of the N matrix of course),

$$\mathbf{N} = \begin{pmatrix} & v_1 & v_2 & v_3 & v_4 & v_5 \\ x_1 & 1 & -1 & -1 & 0 & 0 \\ x_2 & 0 & 0 & 1 & -1 & 0 \\ a & -1 & 0 & 0 & 1 & 1 \\ b & 1 & 0 & 0 & -1 & -1 \\ x_3 & 0 & 1 & 0 & 0 & -1 \end{pmatrix} \tag{5.2}$$

The rate vector is composed out of the rate equations, e.g. reversible Michaelis-Menten, convenience kinetics, or Monod Wyman Changeux models.

5.3 Conservation relations and the link matrix L

Within molecular networks molecules are being cut and pasted by enzymes. This means that a molecule within a network is composed out of parts of other molecules, e.g. its first carbon atom may derive from 2-oxoglutarate, its second nitrogen from alanine, etc. Often such parts of molecules are being recycled, they are not taken up or excreted by a cell. Examples of such recycled molecules within metabolism are the adenoside moiety of ATP, NAD, FAD, and COA.

Consider the metabolic pathway depicted in figure 5.2. In the first reaction, some part of A is glued to S to give rise to the remainer of A, B, and the newly synthesized molecule X. The second reaction isomerizes X into Y. Finally, the part of A glued to S in the first reaction is returned to B to form A and P. If you follow this logic then you will realize that P is an isomer of S. To make this more clear we rename the species in the metabolic pathway, see figure 5.3.

Figure 5.3 shows that species A is composed out of two parts, C and D, and that C is glued to S in the first reaction and CD is formed again in the last reaction. It also indicates that C and CD do not leave the system, only S and S^* do. In other words, the total amounts of C and D denoted by C_T

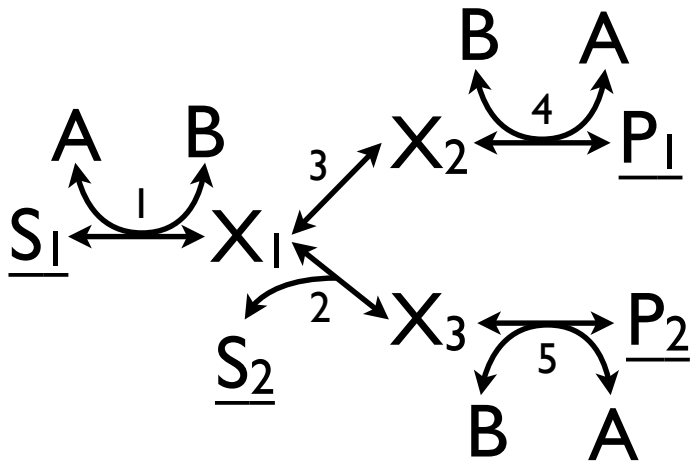

Figure 5.1: A branched metabolic pathway with five enzymes, five intermediates, and four fixed external concentrations. The intermediates A and B are being recycled.

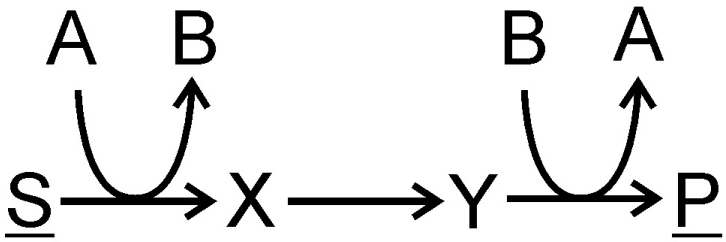

Figure 5.2: A metabolic pathway with three enzymes, four intermediates, and two fixed external concentrations (S,P). The intermediates A and B are being recycled.

and D_T, remain fixed as function of time and are related to initial metabolite concentrations,

$$\begin{aligned} C_T &= CD(t) + SC(t) + S^*C(t) = CD_0 + SC_0 + S^*C_0 \\ D_T &= D(t) + CD(t) = D_0 + CD_0 \end{aligned} \tag{5.3}$$

These equations tell you if you start with 5 mM of D and 3 mM of CD at

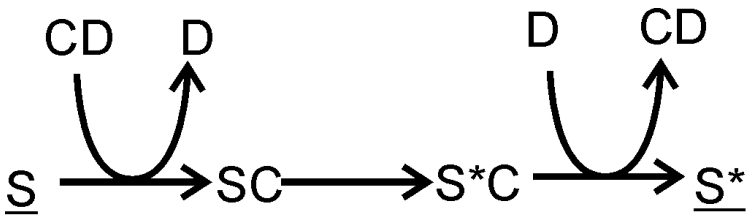

Figure 5.3: The metabolic pathway of figure 5.2 with new names for the metabolites.

time zero that $D(t) + CD(t)$ at any time point will sum to 7.5 mM. Thus D and C are conserved! Equations 5.3 are called moiety-conservation equations. Equation 5.3 also specifies a relationship for the rates of change,

$$\begin{aligned} 0 &= \dot{CD}(t) + \dot{SC}(t) + \dot{S^*C}(t) \\ 0 &= \dot{D}(t) + \dot{CD}(t) \end{aligned} \tag{5.4}$$

These equations indicate their exist linear combinations between the rates of change, which defines $\mathbf{L}$ as a transformation matrix,

$$\begin{pmatrix} \dot{CD}(t) \\ \dot{SC}(t) \\ \dot{D}(t) \\ \dot{S^*C}(t) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} \dot{CD}(t) \\ \dot{SC}(t) \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{I} \\ \mathbf{L_0} \end{pmatrix}}_{\text{link matrix } \mathbf{L}} \begin{pmatrix} \dot{CD}(t) \\ \dot{SC}(t) \end{pmatrix} \tag{5.5}$$

The metabolites CD and SC are termed the independent intermediates and D and S^*C the dependent intermediates. The last equations indicate that the dynamics of all the species can be obtained from the dynamics of only the independent intermediates. The mass balances for all the metabolites (equation 5.1) are given by,

$$\begin{pmatrix} \dot{CD}(t) \\ \dot{SC}(t) \\ \dot{D}(t) \\ \dot{S^*C}(t) \end{pmatrix} = \underbrace{\begin{pmatrix} -1 & 0 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}}_{\mathbf{N}} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \tag{5.6}$$

Using the moiety conservation relationship (equation 5.3), we can substitute for the concentrations of $D(t)$ and $S^*C(t)$ in all the rate equations,

$$\begin{aligned} D(t) &= D_T - CD(t) \\ S^*C(t) &= C_T - CD(t) - SC(t) \end{aligned}$$

The dynamics of the independent intermediates are now given by,

$$\begin{pmatrix} \dot{CD}(t) \\ \dot{SC}(t) \end{pmatrix} = \underbrace{\begin{pmatrix} -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}}_{\mathbf{N_R}} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \tag{5.7}$$

Note that from this set of equation all reference to the concentrations of the dependent intermediates have been eliminated! They are redundant for determining the dynamics. The dynamics of all intermediates can be obtained from equation 5.5. $\mathbf{N}$ has been decomposed as,

$$\mathbf{N} = \begin{pmatrix} \mathbf{N_R} \\ \mathbf{N_0} \end{pmatrix} \tag{5.8}$$

with $\mathbf{N_0}$ as,

$$\mathbf{N_0} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} \tag{5.9}$$

and $\mathbf{N_R}$ as defined above.

Using the above described method, involving renaming of the metabolites, will allows help you to derive the moiety conservation relationships by hand. In systems of the size of glycolysis can easily be handled. In the general case, we then obtain for equation 5.1,

$$\begin{pmatrix} \dot{\mathbf{x}}^{\mathbf{I}} \\ \dot{\mathbf{x}}^{\mathbf{D}} \end{pmatrix} = \begin{pmatrix} \mathbf{N_R} \\ \mathbf{N_0} \end{pmatrix} \mathbf{v} = \mathbf{L}\mathbf{N_R}\mathbf{v} \tag{5.10}$$

$$\dot{\mathbf{x}}^{\mathbf{I}} = \mathbf{N_R}\mathbf{v} \tag{5.11}$$

$$\dot{\mathbf{x}}^{\mathbf{D}} = \mathbf{L_0}\dot{\mathbf{x}}^{\mathbf{I}} \tag{5.12}$$

The general derivation of these equation will be given below using linear algebra. Identification of the moiety conservation relationships can be done by,

$$\mathbf{x}^{\mathbf{D}} - \mathbf{L_0}\mathbf{x}^{\mathbf{I}} = \mathbf{T} \tag{5.13}$$

When we again consider the metabolic network in figure 5.1 the $\mathbf{L}$ matrix is given by,

$$\mathbf{L} = \begin{pmatrix} & x_1 & x_2 & a \\ x_1 & 1 & 0 & 0 \\ x_2 & 0 & 1 & 0 \\ a & 0 & 0 & 1 \\ b & 0 & 0 & -1 \\ x_3 & -1 & -1 & -1 \end{pmatrix} \tag{5.14}$$

This means that the independent variables are x_1, x_2, and a. The concentrations of the remaining dependent metabolites are linearly to the independent species

through,

$$\mathbf{L_0} = \begin{pmatrix} & x_1 & x_2 & a \\ b & 0 & 0 & -1 \\ x_3 & -1 & -1 & -1 \end{pmatrix} \tag{5.15}$$

The reduced stoichiometric matrix is now corresponds to the first three rows of the $\mathbf{N}$ matrix,

$$\mathbf{N_R} = \begin{pmatrix} & v_1 & v_2 & v_3 & v_4 & v_5 \\ x_1 & 1 & -1 & -1 & 0 & 0 \\ x_2 & 0 & 0 & 1 & -1 & 0 \\ a & -1 & 0 & 0 & 1 & 1 \end{pmatrix} \tag{5.16}$$

The moiety conservation equations are (using equation 5.13),

$$\begin{aligned} constant_1 &= a(t) + b(t) = a(0) + b(0) \\ constant_2 &= a(t) + x_1(t) + x_2(t) + x_3(t) \\ &= a(0) + x_1(0) + x_2(0) + x_3(0) \end{aligned} \tag{5.17}$$

The last relationship will surprise many of you.

Exercises

The metabolic network displayed in Figure 5.4 is a simplified representation of the glycolysis as it occurs in *Trypanosomes*. Determine the following matrices $\mathbf{N}$, $\mathbf{N_R}$, $\mathbf{L}$ and the moiety conservation relationships. You should realize that any linear combination of those relationships is again a set of valid conservation relationships. Try to write the relationships you find in a form which is most insightful. For instance by only having sums of concentrations.

5.4 Independent fluxes and the kernel matrix K

The previous section determined a stoichiometric relationship between the concentrations of metabolites that is also valid for transient states of the network. We will now consider a stoichiometric relationship between fluxes through metabolic pathway that hold only at steady state.

At steady state, all the fluxes through a linear pathway without moieties are equal.[1] Thus to determine all the fluxes for a steady-state linear pathway we only have to know one them, the so-called independent flux. All the other fluxes are then equal to this value and are termed the dependent fluxes. For complicated pathways with branches the identification of the independent and dependent fluxes becomes a bit more involved. Again linear algebra will proof useful.

[1] Remember the convention that enzyme conversion rates are called fluxes at steady state conditions.

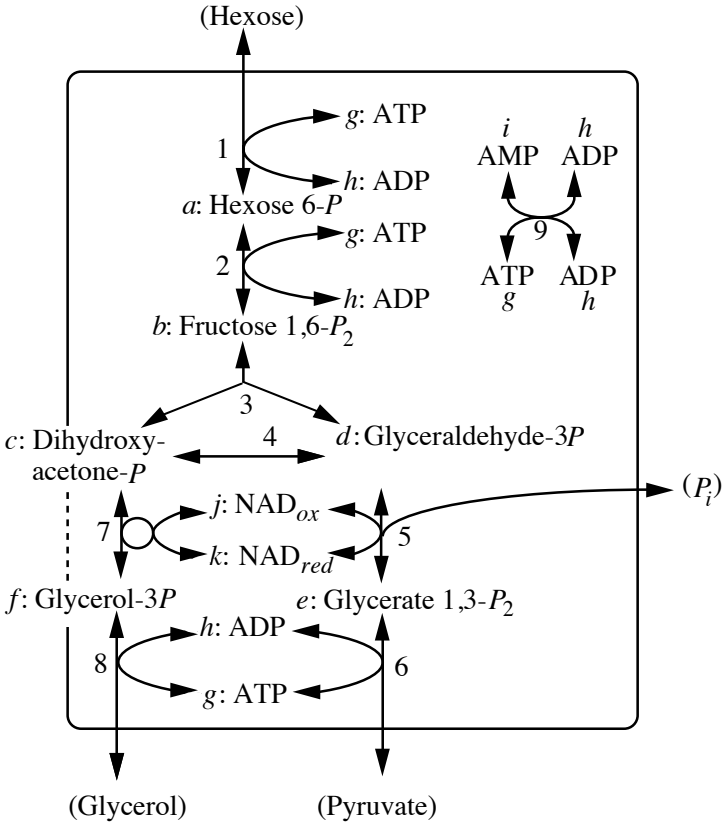

Figure 5.4: Simplified representation of the glycolysis as it occurs in *Trypanosomes*.

We will consider in this section, networks without moiety conservation relationships before we outline generic linear algebra methods for their determination in the next method. In this case, the number of dependent fluxes is given by the number of variable metabolite concentrations, i.e. mass balances. This can be easily understood. If there are r fluxes and m variable metabolites, then m (steady-state mass balance) relationships exist between all fluxes and $r - m$ fluxes need to be supplied - those are the independent fluxes - to determine them all.

Let's consider an example network, displayed in figure 5.5. This network has no conserved moieties, seven fluxes and five variable metabolites. So we need to know two flux values to determine all values. Clearly, some combinations of two fluxes will not work. Knowing for instance J_1 and J_2 will not help to determine the fluxes after the branch. When we would know J_1 and J_3, all fluxes can be determined using the mass balances for the variable metabolites at steady state

(in the order of the metabolite numbering),

$$\begin{aligned} J_1 &= J_2 \\ J_2 &= J_3 + J_5 \\ J_3 &= J_4 \\ J_5 &= J_6 \\ J_6 &= J_7 \end{aligned} \tag{5.18}$$

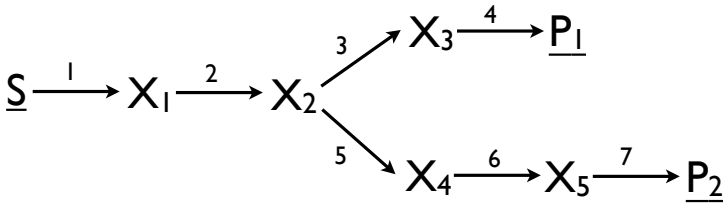

Figure 5.5: A branched metabolic pathway with five variable metabolites and seven fluxes.

These equations can be captured in matrix form,

$$\begin{pmatrix} J_1 \\ J_3 \\ J_2 \\ J_4 \\ J_5 \\ J_6 \\ J_7 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} J_1 \\ J_3 \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{I} \\ \mathbf{K_0} \end{pmatrix}}_{\mathbf{K}} \begin{pmatrix} J_1 \\ J_3 \end{pmatrix} \tag{5.19}$$

This equation indicates that the **K**-matrix relates all flux values to the values of the independent fluxes. Each column of the **K**-matrix is a segment of the metabolic network that can attain a steady-state state by itself. They are called flux modes. So any flux distributions of the entire network is a linear combinations of its flux modes.

For systems without moieties the method outlined in this section always works but becomes cumbersome for large systems. Linear algebra can help us make this task computable by Matlab or Mathematica as will be shown in the next section.

5.4.1 Exercise

Identify the flux models of the network displayed in figure 5.5. Construct a new valid **K** matrix that has only positive entries by taking a linear combinations of the columns of the **K** matrix given in equation 5.19.

5.5 Derivation of the K and L matrix using linear algebra for any N matrix

In order to determine the **K** and **L** for any **N** matrix we have to follow the scheme displayed in figure 5.6. From above, we conclude from equation 5.12,

$$\mathbf{L_0 N_R} = \mathbf{N_0} \tag{5.20}$$

This equation can be rearranged to yield,

$$\begin{pmatrix} -\mathbf{L_0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{N_R} \\ \mathbf{N_0} \end{pmatrix} = \mathbf{0} \tag{5.21}$$

and therefore,

$$\begin{pmatrix} \mathbf{N_R} \\ \mathbf{N_0} \end{pmatrix}^T \begin{pmatrix} -\mathbf{L_0} & \mathbf{I} \end{pmatrix}^T = \mathbf{0} \tag{5.22}$$

So the right nullspace of the transpose of **N** (so **N**'s left nullspace) equals $\begin{pmatrix} -\mathbf{L_0} & \mathbf{I} \end{pmatrix}^T$. The rank of **N**, m_0 given the number of independent rows, which equals the number of independent metabolites. The dimensions of **L** are therefore $m \times m_0$.

At this point, we have determined the reduced stoichiometry matrix $\mathbf{N_R}$. The number of dependent fluxes in steady state is given by the number of row of N_R, m_0. The number of independent fluxes is $r - m_0$. The **K** matrix can be obtained from the right nullspace of the reduced stoichiometry matrix $\mathbf{N_R}$,

$$\mathbf{N_R K} = \mathbf{0} \tag{5.23}$$

Note that the K matrix is not unique, any linear combination its combinations of its columns or their multiplication with a constant will give rise to a valid nullspace matrix of **N**. Elementary flux modes and extreme pathways are unique definitions.

There is one more useful property of kernel matrices. We have already concluded that their columns are called flux modes and have the property that they can attain steady-state by themselves; that is, they are either cycles in the network or link source to sink metabolites. They also have an overall reaction stoichiometry. For instance, the overall stoichiometries of the flux modes of the stoichiometric network displayed in figure 5.5 are $S \rightarrow P_1$ and $P_2 \rightarrow P_1$, for the first and second flux mode, respectively. These can be determined in the following manner,

$$\mathbf{K}^T \cdot \mathbf{r} \tag{5.24}$$

where **r** describes the stoichiometry of the reactions as,

$$\mathbf{r} = \mathbf{X}^T \mathbf{N} \tag{5.25}$$

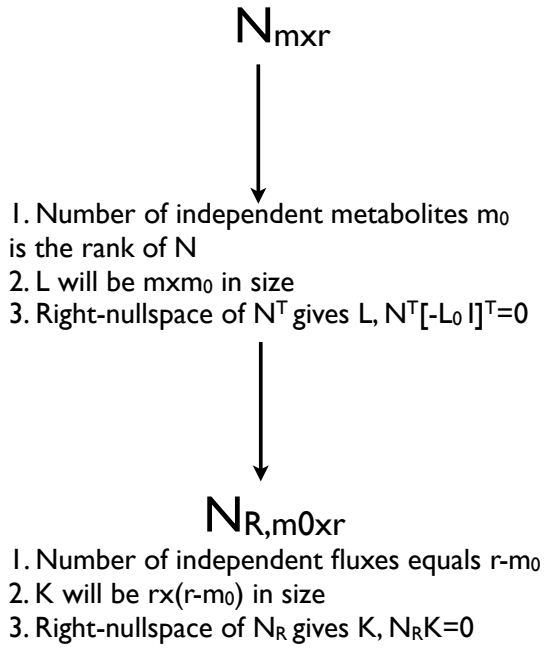

Figure 5.6: Workflow for **L** and **K** determination.

Where **X** is the vector of species names. This gives for the network displayed in figure 5.5,

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} -S + X_1 \\ -X_1 + X_2 \\ -X_2 + X_3 \\ -X_3 + P_1 \\ -X_2 + X_4 \\ -X_4 + X_5 \\ -X_5 + P_2 \end{pmatrix} = \begin{pmatrix} -S + P_1 \\ -P_2 + P_1 \end{pmatrix} \quad (5.26)$$

Exercises

1. Use Mathematica to derive the **K** for the network displayed in figure 5.4.
2. How can the overall stoichiometry of a flux distribution **J** be determined? If the network in figure 5.5 has the following distribution $(10, 10, 6, 6, 4, 4, 4)^T$ what is the overall stoichiometry of the pathway?
3. Determine the **N**, **K** and **L** matrix for the four metabolic networks displayed in figure 5.7.

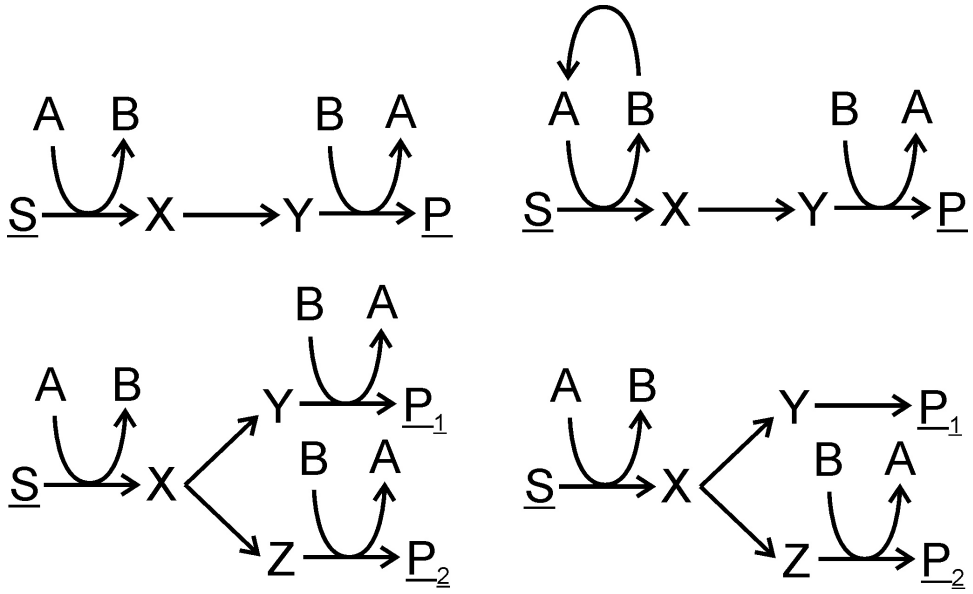

Figure 5.7: Four metabolic networks with that differ in independent fluxes and metabolic intermediates.

5.6 Steady-state flux space and constraint-based modeling

In the previous section we have seen how to analyse a stoichiometric matrix to determine the moiety conservation and flux relationships of a metabolic network. In this section we will explore the steady-state solution space further, particularly with respect to constraints that can be imposed on flux distributions. This analysis and the collection of techniques involved are often called constrained-based modeling. We will discuss the uses and limitations of such approaches, provide examples of simple systems to get the basic idea, and then discuss its use in much larger systems: genome-scale metabolic models.

In Figure 5.8 a simple branched pathway is depicted, with $\mathbf{N} = [1, -1, -1]$. The rank of $\mathbf{N}$ is 1, and since there are 3 rates, the dimensions of the null space of $\mathbf{N}$ is 2, which is equivalent to 2 independent fluxes. If we take $\mathbf{v_1}$ and $\mathbf{v_2}$ as independent fluxes, the Kernel $\mathbf{K}$ equals:

$$\mathbf{K} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{pmatrix} \tag{5.27}$$

The columns of the Kernel can be interpreted as flux modes, as indicated in the figure 5.8. These are pathways running from S_1 to P_3 (b_1), and P_3 to P_2 (b_3),

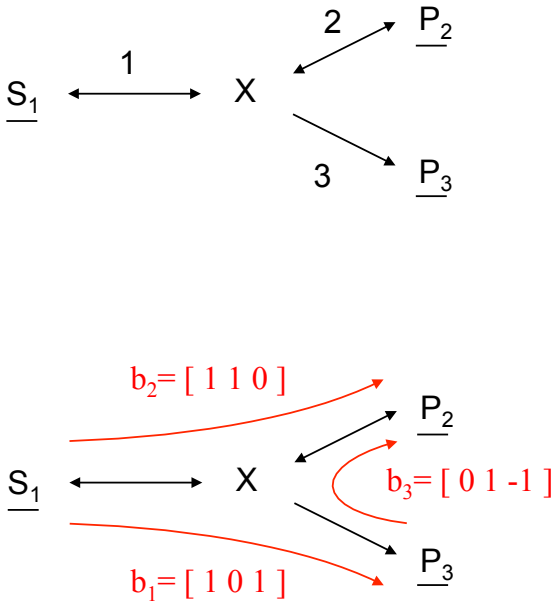

Figure 5.8: Simple three-enzyme pathway with a branch.

respectively. They are base vectors that span the steady-state solution space, which in this case is a plane in 3D-space (see Figure 5.9. Note that we have no constraints at all on the values that reactions can take, i.e., the plane stretches out to infinity in both directions. However, suppose that reaction 3 is irreversible under physiological conditions. What we mean by this is that the mass action ratio of X_1 and P_3 cannot compensate, within reasonable bounds of $[X_1]$ and $[P_3]$, for a very large equilibrium constant of reaction 3. This is in fact a first constraint on the reaction network: $v_3 > 0$, which is rooted in thermodynamics. The base vector (flux mode) defined in terms of negative values for v_3 is therefore physiologically unattainable, even though it is mathematically a perfect base for the null space of $\mathbf{N}$. Adding column 1 to column 2 in equation 5.27 results in a more useful set of base vectors under the constraint $v_3 > 0$, the original flux mode S_1 to P_3, and the new flux mode S_1 to P_2 (b_2). Note that this Kernel equally well describes the unconstrained null space of $\mathbf{N}$ as did the one of equation 5.27.

The constraint $v_3 > 0$ can graphically be seen as a plane that cuts to solution space into two halves, only one of which fulfills the constraint. We see that these constraints reduces the number of feasible base vectors to describe the null space. In the case that all reactions are (required to be) positive, the solution space is constraint to the positive quadrant of the flux space, which is then called a convex space. This is most often not a biologically relevant condition as many reactions are reversible, but nevertheless it has become the standard way in

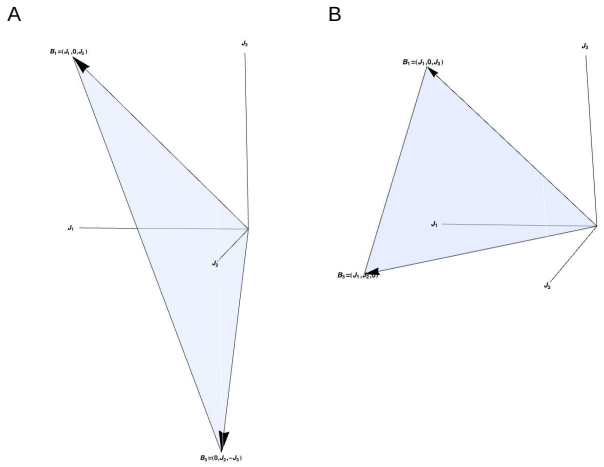

Figure 5.9: Null space of the simple branched pathway of of figure 5.8 spanned by two different sets of basis vectors. Note that they span the same plane, which stretches out in all directions if there are no additional constraints: any linear combination of the basis vectors is a point in the null space.

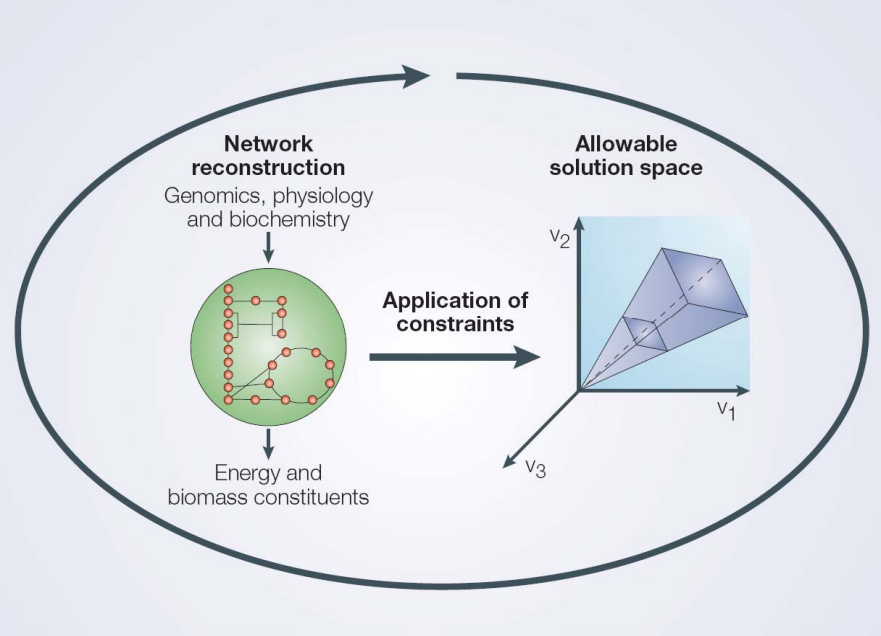

Figure 5.10: Popular represenation of the flux cone in 3D. Picture from Price (2004) Nat Rev Micro 2, 886.

constraint-based modeling reviews to visualise the flux space, which is under non-negativity a flux cone, some sort of ice cream cone spanned by base vectors of the Kernel (see Figure 5.10. Under such strict constraints, the basis vectors for spanning the convex solution space are unique and they are known in the systems biology literature as extreme pathways. For the pathway of figure 5.8 this flux cone is the triangle spanned by $[J_1, 0, J_3]$ and $[0, J_2, J_3]$, i.e. the triangle in figure 5.9.

Apart from constraints on the direction of reactions, there are also constraints possible on the maximal rates of certain enzymes: these can also be represented as planes that reduce the solution space of possible steady-state flux distributions. This is illustrated in Figure 5.11, with $0 < J_1 < J_{1,max}$ and $0 < J_2 < J_{2,max}$. Within constraint-based modeling, constraints on the direction of fluxes and on their rates are collectively called capacity constraints. The maximal rates of enzymes are needed to bound the solution space into a closed polytope: without proper capacity constraints the solution space is unbounded.

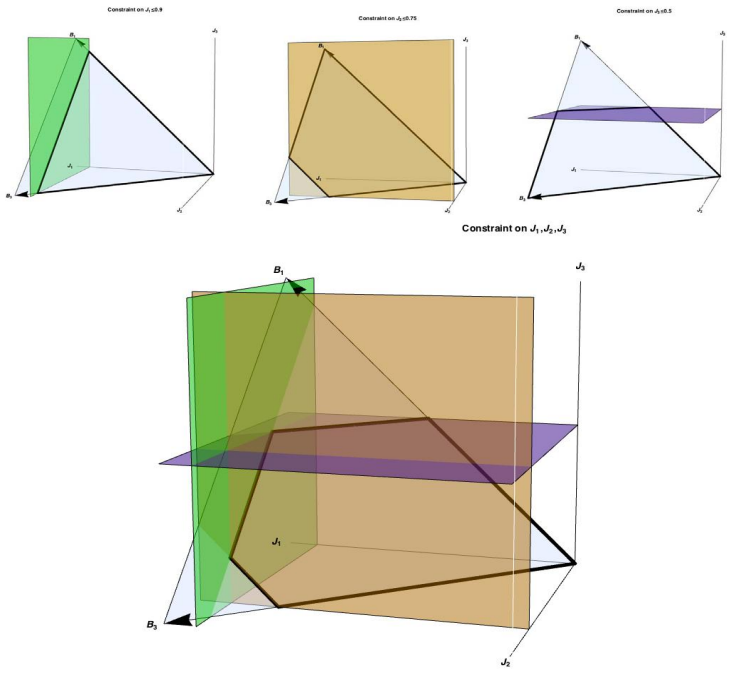

Figure 5.11: Flux cone of simple branched pathway of figure 5.8 as the result of capacity constraints applied to the original null space.

This is obviously not biologically possible, and bounded solution spaces are required to ask quesions about optimality, as we will see shortly.

The steady state solution space is important for a number of applications in systems biology, including dynamic modeling. Stoichiometric analysis is therefore the first step to perform when constructing a kinetic model, not only to find moiety conservation and hence reduce the number of variables that determine the dynamic behaviour of the model, as shown above. The steady state flux space gives the set of states to which *any* kinetic model with this reaction stoichiometry will eventually evolve to in time, i.e. this state space is not dependent on the kinetic parameters. [2] The kinetic parameters, as we will see, determine the specific steady state that will be reached in the solution space, it will determine, together with the initial conditions, the trajectory towards the steady state, and it will determine the control structure in the steady state (see

[2]given fixed source and sink concentrations, and assuming for now a unique and stable solution to the set of ODE's of the model.

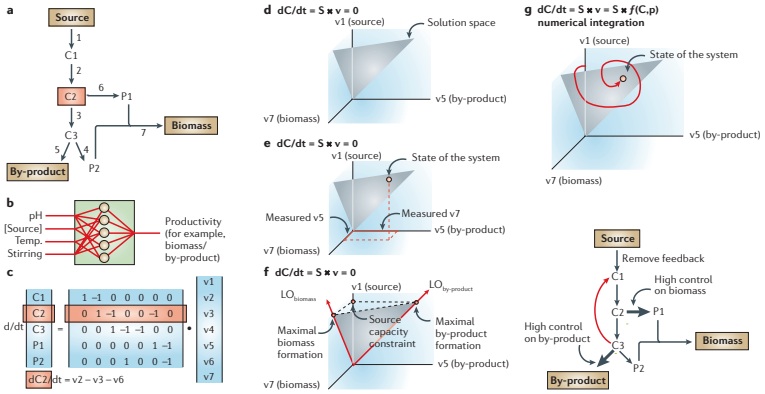

Figure 2 | **The basics of common modelling strategies. a** | A simplified metabolic network is shown in which a source can be converted into biomass and by-product through several chemical conversions, and which involves 3 intracellular metabolites, C1, C2 and C3, and two biomass precursors, P1 and P2. The concentrations of source, by-product and biomass are set as being fixed; for example, they are continuously fed into the system and removed, or the changes in these concentrations are slower than the changes in the intracellular metabolites, so at each point in time these variables can be treated as constant (pseudo steady-state assumption through timescale separation). A desired outcome might be to increase by-product formation (for example, lactate) or biomass formation. These products compete for common precursors, in this case C2 and C3. **b** | Black box approach. Here, several inputs, such as process parameters and variables, are linked to relevant output variables. Such a model might be used for process optimization and control. The connections inside the black box do not need to be biologically meaningful. **c** | Matrix formulation of the explicit metabolic network model. One can derive for each compound in the network a differential equation describing its dynamics as a function of the reaction rates. For compound C2 in part **a**, this reads: $dC2/dt = v_2 - v_3 - v_6 - D$ (eq. 1), in which D is a term that describes the dilution of compound C2 into new cells; that is, $D = \mu \times C2$, μ being the growth rate. This dilution effect, however, is usually negligible and therefore left out, but for some compounds, such as some vitamins, no degradation pathway is known, and dilution into new cell is the only sink and should be defined. The entire system is described by a set of coupled, ordinary differential equations, which is most conveniently notated in matrix form: $dC/dt = S \times v$ (eq. 2), in which dC/dt is a vector with all time derivatives of the metabolites, v is the vector with all reaction rates, and S is the stoichiometry matrix linking all the metabolites to the reactions. The differential equation for C2 (eq. 1) corresponds to the second row in the stoichiometry matrix S. **d** | Steady-state approach. In metabolic flux analysis and in constraint-based modelling, steady states are assumed, which postulates that all metabolites are balanced and therefore $dC/dt = S \times v = 0$ (eq. 3). This results in a set of homogeneous equations, which can be solved with basic linear-algebra techniques. The set of all possible solutions to this set of equations is called the solution space, which mathematically corresponds to the null space of the stoichiometric matrix S[51]. In general, stoichiometric matrices tend to have more reactions than metabolites, which is caused by branching in metabolic networks. Moreover, some combinations of metabolites might be conserved (for example, the adenosine moieties in the adenine nucleotides), introducing linear dependencies in the set of equations. Therefore, the set of equations in equation 3 is usually underdetermined, resulting in a large solution space. Also in this example, there are 5 metabolites and 7 reactions (that is, 7 unknowns and 5 equations). The solution space is therefore two-dimensional. In part **d**, the relationship between source, biomass formation and by-product formation is shown for this system. From the stoichiometry matrix S it can be derived that in steady state $v1 = v5 + 2v7$ (a relationship that can be found by inspection for this system, but for any realistic system requires linear-algebra software such as Matlab or Mathematica). **e** | Metabolic flux analysis. Equation 3 is solved by measuring enough unknown fluxes (reaction rates in steady state) so that the system becomes determined (or preferably overdetermined). Alternatively, ^{13}C-labelling studies can be used to estimate fluxes. Therefore, the state the system is in, a point in the solution space, is estimated based on experimental data. For example, if the biomass and by-product formation rates were measured, the consumption of source would be known. Agreement with measured rates of source consumption would validate the underlying network structure. Metabolic flux analysis therefore describes a metabolic system. There are two approaches to actually predict the steady state of a system: **f** | Flux balance analysis. A predicted solution to equation 2 is found by assuming that the system fulfils an optimality condition, and therefore the state of the system could be predicted by finding the optimal flux distribution for that objective function. This technique is often used in genome-scale models, in which the number of unknown fluxes is too large to measure. The optimal solution should lie on one of the red vectors (sometimes referred to as the 'line of optimality', indicated as LO[105]), corresponding to optimal biomass formation or by-product formation. To find a particular maximum, one or more capacity constraints are required, otherwise the solution to the optimization will be unbounded (the vector goes to infinity). At a given source capacity constraint (indicated in the figure), the corresponding optimal solutions for growth and by-product formation are indicated as red dots. **g** | Kinetic modelling. Alternatively, one might make a full kinetic model. In such an approach, all kinetic parameters are collected and equation 3 is solved numerically with $v = f(C,p)$; that is, each reaction rate is a function of some kinetic parameters p and the concentration of substrates and products, C. Such an approach results in a prediction of not only all steady-state fluxes but also all metabolite concentrations. Additionally, one obtains the dynamics of the system as well as the control structure of the system, that is, the sensitivity of the fluxes and metabolites to parameters[80].

Figure 5.12: From [34].

also Figure 5.12).

But also without any kinetics, stoichiometric analyses have been very popular, and also successful, in analyzing metabolic networks. In metabolic engineering, people use stoichiometric network analysis to deduce flux distributions from measured data. The idea is simple: if the dimension of the solution space is, let say, 5, meaning there are 5 independent fluxes that together determine all dependent fluxes, one needs to just measure those 5 fluxes. The issue is of course that the only fluxes that are easily accessible are the external ones, the product formation rates and the nutrient consumption rates. The trick is therefore to find a $\mathbf{K}$ matrix in which all independent fluxes can be measured

experimentally. This activity is often referred to as metabolic flux analysis.

In the genomic era, especially the stoichiometric analysis of very large genome-scale metabolic networks have been very useful to turn sequenced genomes into mathematical models representing the complete metabolic network of the sequenced organism. The approaches, pitfalls and bioinformatic tools for making such metabolic reconstructions fall outside the scope of this book, but have been described in more detail elsewhere. These genome-scale metabolic networks have sizes in the order of 500 to 1500 (!) reactions, and similar number of metabolites. The number of independent reactions is often over 100! The size of these networks makes it close to impossible to model it comprehensively with kinetic models. The complexity would be devastating, and the number of parameters involved are simply so large that we do not have the capacity to measure them all. Therefore people have resorted to stoichiometric analyses of such networks. Clearly, even the steady state solution space of these networks is hugely complex, but there are a number of analyses that have proven to be useful. The most popular and influential one is called Flux Balance Analysis (FBA). With FBA we use linear optimization to find flux distributions that maximize or minimize a certain *objective function*. This objective function is itself a flux or a linear combination of fluxes. The optimisation problem reads:

$max\ Z$

subject to

$\mathbf{Nv} = \mathbf{0}$ (steady state constraint, often called "mass-balance" constraints)

$a_i < v_i < b_i$ for all v_i elements of $\mathbf{v}$ (capacity constraints) (5.28)

Note that if a_i is 0, the corresponding reaction is put to be irreversible. For most internal, reversible fluxes, a_i and b_i are set to (-)infinity, which is fine as long as sufficient input fluxes are constraint by some maximum to bound the solution space. To understand what the optimisation does, we will make use of the same example as in the previous part, the network of Figure 5.8. In Figure 5.13 the solution space is drawn as a projection on the (J_2, J_3) plane. Important is the line $J_2 + J_3 = J_1$ (mass-balance). By constraining J_1 to J_1, max, the solution space is already bounded; additional capacity constraints for J_2 and J_3 introduce the typical polytope shape, shaped by a number of edges that represent constraints. In this graph we can immediately see what flux distribution would maximize a certain objective function, Z, that we can express as $Z = w_2 J_2 + w_3 J_3$, where w_2 and w_3 are some scalars weighting the contribution of each flux in the objective function. In this particular example there are several possibilities: (i) unique solution: if w_2 is not w_3 and w_2 and w_3 are nonzero. If $w_2 > w_3$ point B in Figure 5.14 is reached, if $w_2 < w_3$, point A is reached (ii) an infinite number of solutions: if $w_2 = w_3$, all points between A and B maximizes Z (see Figure 5.15) (iii) an infinite number of solutions: if $w_2 = 0$ and w_3 is nonzero, any point between the y-axis and point A would be a solution (there is obviously an equivalent situation if w_3 is zero). (iv) unbounded solutions are also possible, where the value of Z would reach infinity, but this is not possible in this example where the flux cone is fully bounded.

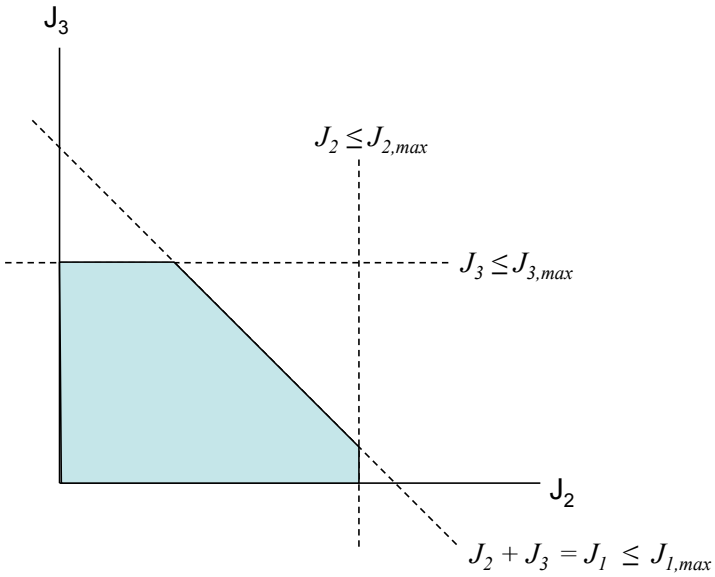

Figure 5.13: 2D projection of the solution space of the network displayed in figure 5.8. Indicated are a mass balance constraint ($J_2 + J_3 = J_1$), and two capacity constraints, on J_2 and on J_3

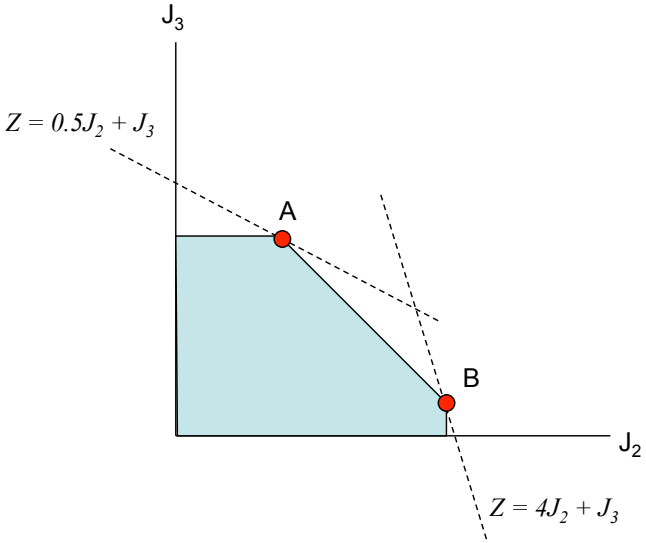

Figure 5.14: Solution of FBA of the network displayed in figure 5.8 with an objective function with different weights on J_2 and J_3.

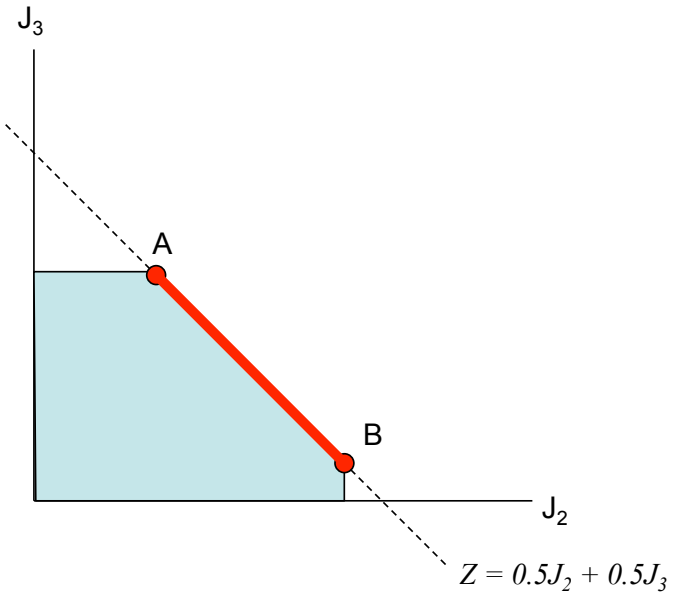

Figure 5.15: Solution of FBA of the network displayed in figure 5.8 with an objective function with equals weights on J_2 and J_3.

In the case of 500 to 1000 reactions, we obviously have to resort to computers rather than inspection to find the optima. In engineering very efficient algorithms for these type of linear optimisation problems have been developed, collectively called linear programming. An FBA problem with $>$ 1000 reactions is solved within seconds on a desktop using standard solvers built-in in Mathematica or Matlab. Specifying constraints, the objective function, and interpreting the result are the most difficult parts of genome-scale constraint-based modelling. The constraints come from pysiology and experimental data, i.e. what nutrients are in the medium that the organisms can consume, and at what rate are these nutrients consumed; what products can be made, what is the composition of the cell in terms of proteins, lipids, RNA etc. So, in general, FBA yields unique maximal values for Z, but not necessarily unique flux distributions to reach this value of Z. This is extremely important in real-life use of the technique, as one cannot rely on a single optimisation to conclude what the optimal flux through a step should be to reach Z (for example as a candidate reaction to delete or augment by metabolic engineering). The technique to check for uniqueness or degeneracy of flux values of individual reactions of an FBA solution, is called flux variability analysis. It should follow any FBA solution of interest. FVA is formulated as:

$max/min\ v_i$ for all v_i element in v
subject to
$\mathbf{Nv} = \mathbf{0}$ (steady state constraint, often called "mass-balance" constraints)
$a_i < v_i < b_i$ for all v_i elements of $\mathbf{v}$ (capacity constraints)
$Z = Z_{max}$ FBA result (5.29)

The last constraint ensures that the maximal and minimal fluxes through each individual flux is evaluated at the optimum objective function value of the FBA problem. FVA involves many rounds of FBA, 2 times the number of reactions in the network. Subsequently the span of a reaction is defined as the maximal minus the minimal value: if the flux through a reaction is completely fixed by the maximal value of Z, its span is 0. In the example of Figure 5.13, if $Z = J_2 + J_3$, the span for J_2 is the difference of its values in A and B (B-A). A large difference means that this flux is not constraint by the objective function and that there is a high degree of flexibility in that part of the network. FVA can also be used for other purposes than testing alternative FBA solutions, such as a genome-scale equivalent of metabolic flux analysis. In this case *measured* flux data can be used as constraints in the FVA formulation, rather than $Z_{max,FBA}$, and FVA will test which parts of the network are predictable by the measurements (small span) and which parts are not (large span). Remember that a genome-scale model has in the order of 100 degrees of freedom (independent fluxes) so one would need quite a lot of data to completely predict all fluxes in such a model!

5.7 Applications and pitfalls

One key issue of course is the definition of the objective function: if we want to predict flux distributions, using FBA, that make biological sense, we need objective functions that make biological sense. Maximizing ethanol production for biofuel production would be a good objective function from the perspective of man (not necessarily from the perspective of yeast!). In the literature, maximization of growth rate has most often been used as an objective function, with quite some success, but also failure. Maximization of growth rate makes sense for microorganisms; it is in fact used in population dynamic studies as the proxy for fitness. But do we really optimise growth rate with FBA? To answer this question, we first need to define the objective function, growth rate, within the FBA format. Within stoichiometric model, growth is modeled as a sink of biomass components (protein, DNA, RNA, lipid, carbohydrates) that reflect the biomass composition of the cell. Here is an example for the lactic

acid bacterium *Lactobacillus plantarum*:

$$\begin{aligned} &2.45\ protein + 0.279\ RNA + 0.062\ DNA + vitamins + 0.081\ lipids \\ &+ 0.129\ polysaccharides + 0.146\ peptidoglycans + 0.014\ wall\ teichoicacids \\ &+ 0.013\ lipoteichoicacids + 27.4\ ATP + 27.4\ H2O \\ &\rightarrow biomass + 27.4ADP + 27.4H^{+} \end{aligned}$$

The stoichiometric coefficients have unit mmol gDW^{-1}. Note the use of 27.4 ATP to form biomass: this number reflect lumped ATP costs for putting all the biomass components into actual living cells. Many of the processes involved are still unknown and unaccounted for in the models. Hence this is an empirical number fitted by varying the growth rate (e.g. in a chemostat) and estimating the amount of ATP that is being formed by catabolism. Assuming that this ATP is used for growth, the growth-associated ATP requirement can be estimated. Note also that because of the unit of the stoichiometric coefficients (mmol gDW^{-1}), the unit of this reaction is different from the other reactions in the network. Convention is to express fluxes as mmol h^{-1} gDW^{-1}; consequently, the unit of the biomass formation rate is h^{-1}, the specific growth rate! So it appears indeed that we optimise growth rate! This is wrong. Remember we are dealing with stoichiometric network models, and that we need some capacity constraints to bound the solution space: the maximal growth rate is therefore always bounded by some limiting input flux: therefore we ask what the maximal growth rate is *relative* to the input flux. This is in fact a yield (i.e. a ratio of fluxes). Figure 5.16 makes this point very clearly.

Optimising yield or optimising rate is biologically a completely different thing: one can be very fast at the expense of efficiency (yield), and this strategy may win the batlle of the fittest. Under other conditions (e.g. of poor energy resources), maximizing yield may be the best strategy. Importantly, stoichiometric analyses can only predict optimal yields, not a priori rates. To turn concentrations into rates, one needs kinetics.

5.8 Sensitivity analysis

Once an FBA solution has been found, there are two types of sensitivity coefficients that are useful for interpretation and understanding of the metabolic network. These are called shadow prices and reduced costs. Reduced costs are relatively easy to understand: they quantify how much the objective function would change if a capacity constraint is changed:

$$r_i = \frac{\partial Z}{\partial b_i} \tag{5.30}$$

Or by differentiating with respect to a_i. This is illustrated in Figure 5.17.

Obviously, reduced costs are always zero if the FBA solution does not hit the minimal or maximal constraint, and so reduced costs indicate reactions that are somehow constraining the objective function. In many applications, these

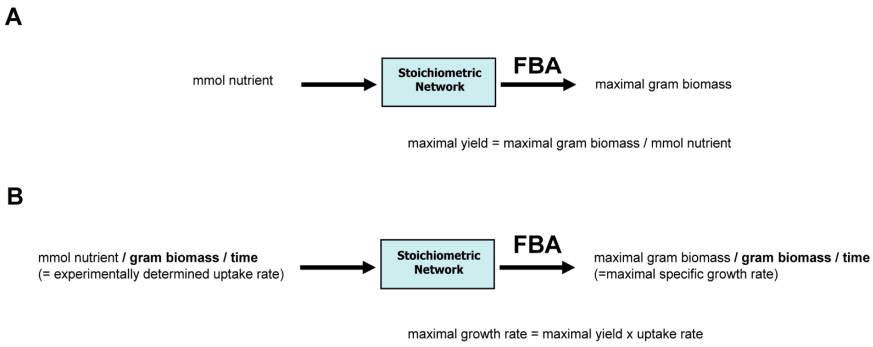

Figure 5.16: 1A A stoichiometric network can be used, with FBA, for optimization of maximal yield of biomass on a certain nutrient. 1B by providing an experimentally measured input rate (capacity constraint in constraint-based modeling terms), FBA predicts a specific growth rate. The two situations are, however, exactly the same except for some scaling factor (indicated in bold). In both cases, a flux distribution through the stoichiometric network will be found that maximized the yield of biomass on the nutrient.

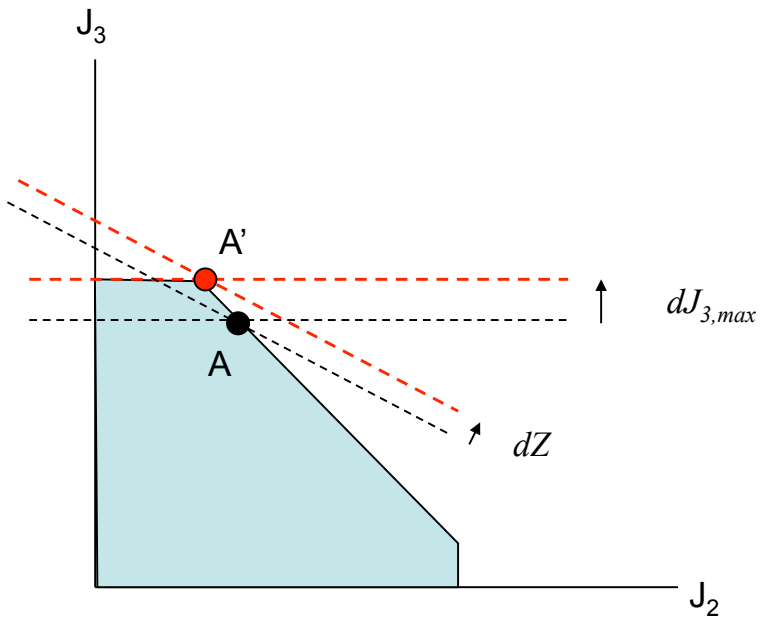

Figure 5.17: Illustration of reduced costs

are input or output fluxes, e.g. the glucose input flux will often constrain the maximal growth rate. These numbers are therefore interesting if one is interested in medium optimisation. The other sensitivity coefficient is the shadow price: it is the change in objective value Z if the mass-balance equation is altered. It therefore deals with metabolites, not rates! The easiest way to think of shadow prices is as the "value" of a metabolite in terms of the objective value, i.e. if one would introduce a metabolite in the medium together with a transporter (and so tap in the metabolite for free), would it affect Z? It therefore introduces an extra column in the stoichiometric matrix with only a "1" at the metabolite evaluated. This is illustrated in figure 5.18.

Exercises

1. Draw examples of different network topologies where at least 5 fluxes can be estimated from only 2 measurements. Explore with what structure external flux measurements do not suffice to solve the set of fluxes. 2. In Figure 5.12 a metabolic network is shown. Assume all reactions are irreversible. - Construct a K matrix by inspection and by linear algebra (section 5.5) and draw the flux modes in the network. - Draw the solution space in 3D in Mathematica. - try to maximize byproduct formation using FBA. Do you get a solution? - constrain the network such that it becomes bounded. - calculate maximal byproduct and biomass production under such constraints - do FVA for each optimization -

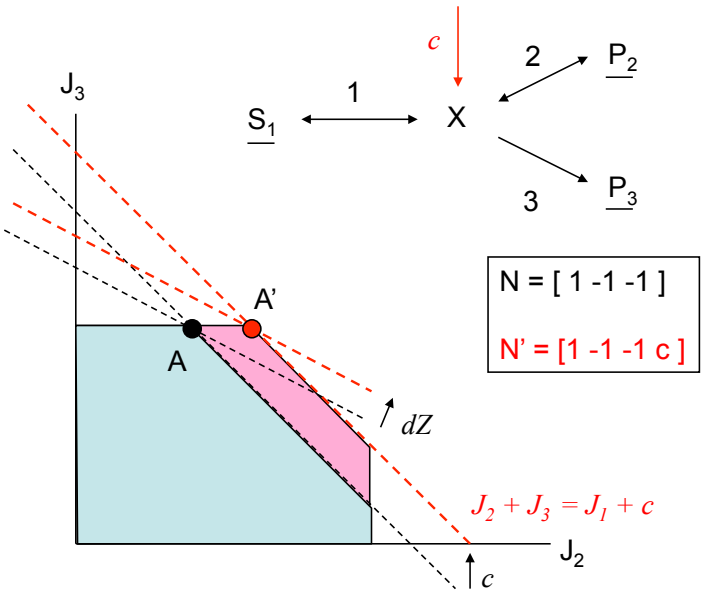

Figure 5.18: Illustration of shadow price

a flux of v_1 of 10 +/- 2 was measured, and a biomass production of 4 +/- 1. Calculate the predicted span of the byproduct formation flux. 3. In Figure 5.19 a metabolic network is shown with parallel pathways. Assume all reactions are irreversible. - The input flux was measured to be 10. Give the span of each reaction in the network. - Now assume all reactions in the network are reversible: does this affect the span? Show by computation. Explain the result and discuss if this is realistic in real life.

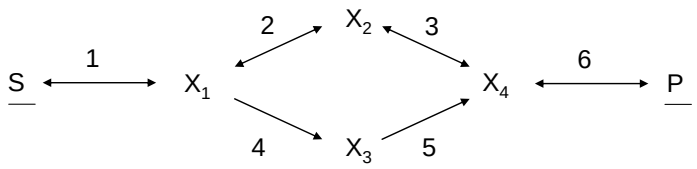

Figure 5.19: Metabolic network with parallel pathways

Chapter 6

Biological control theory

6.1 Introduction

Cells adapt their physiological strategy upon environmental changes. This may involve network rewiring, changes in metabolic rates, alterations in covalent-modification levels of signaling proteins, and new sets of membrane receptors. All these changes are induced upon perception of the environmental change and processing by signaling, metabolic and gene networks. To some environmental influences cells do not respond at all even though some of their processes are sensitive to those disturbances. How can cells achieve perturbation-specific sensitivity and robustness? This question has to do with the extent by which the initial effect of a perturbation on a process rate, say a change in the level of glucose on the rate of a glucose transporter or growth factor on the autophosphorylation capacity of a growth factor receptor, is propagated through the entire molecular network to bring about a global change. Metabolic control analysis can give insight into such network responses.

Physiological adjustments have to be carried out within strict constraints. For instance, many reactions rely on ATP and NADH and large changes in their levels would perturb many processes simultaneously, causing havoc. So some concentrations have to kept within small bounds (homeostasis) despite the fact that the flux through those pools may change by orders of magnitude. On top of that, cells have limited internal space and energy budgets, which means that the benefit of a physiological adaptation should at least compensate for the cost. Another complication are trade-offs, a change in one part of the network may enhance physiological performance while at the same time causing a reduction in performance by another part of the network. So somehow cells are tinkered by natural selection to achieve sophisticated constraint multi-objective control and optimization tasks.

In this chapter, we will introduce a number of concepts and tools in the framework of metabolic control analysis (MCA) to tackle how system properties of networks are being controlled and regulated by cells. Examples of such

systems properties are fluxes, concentrations, response times, sensitivity and robustness. The following aspects will be addressed:

- In general, no rate-limiting reactions or master regulator exist in molecular networks,
- Any system property is under the influence - controlled - by the activity of all the reactions in the network,
- The extent of control that a particular reaction exerts on a system property depend on the state of the network, all the kinetic parameters and characterization of the environment - it is a network property itself,
- The sensitivity of reactions to their reactants and effectors determine largely the control distribution of system properties,
- Feedback circuitry are potent mechanisms to make network (ultra-)sensitivity and robust to changes in their environment

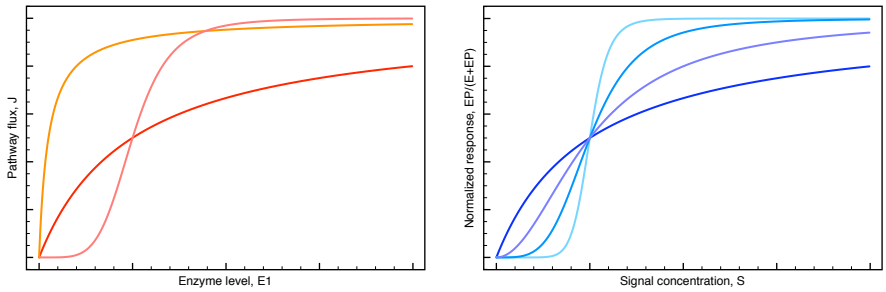

Figure 6.1: Different responses of steady-state system properties of a molecular network, i.e. a metabolic flux (J) and the covalent modification fraction of a protein $EP/(E + EP)$, as function of a physiological parameter, such as an enzyme or signal level. Networks can generate different input/output characteristics depending on their structure and parameterization. Control theories analyze how the sensitivity of system properties to parameters can be explained in terms of network design and process parameterizations. In this way more insight can be gained into the molecular mechanisms underlying robustness (parameter insensitivity) and fragility (parameter sensitivity). The examples in this plot all show stimulatory responses but the same variability can be observed in repressive responses.

6.2 System properties

Cells change their physiological states upon external stimuli. Changes in pathway fluxes, covalent modification levels of signaling proteins, rates of gene ex-

pression are all responses involved in adaptive behavior. Each of these system properties responds in a specific manner to a specific stimulus. Figure 6.1 shows some examples of system responses.

Having made it this far into this book you will realize at this point that not a single molecular property will explain how the dependency of network functions on parameters but rather that many processes will be typically responsible for it to varying extents. This indicates that a metabolic pathway flux, a gene transcription rate, or the covalent modification state of a protein are each dependent in some unintuitive nonlinear manner on all kinetic properties of the enzymes, which specifies the network structures with all its branches, pathways, and feedback circuitry. This function is not tractable in most cases.

Consider for instance the simplest pathway imaginable to illustrate some of the concepts behind MCA (network A in figure 6.2),

$$\underline{S} \overset{v_1}{\rightleftharpoons} X \overset{v_2}{\rightleftharpoons} \underline{P} \tag{6.1}$$

The linear pathway is composed out of two reversible reactions where S and P are held fixed. For illustrative purposes, we will assume those reactions to follow mass-action kinetics. In the next section, we consider enzyme kinetics. The mass balance for X is now given by,

$$\frac{d}{dt}x = v_1 - v_2 = k_1^+ s - k_1^- x - k_2^+ x + k_2^- p \tag{6.2}$$

At steady state, $dx/dt = 0$, the concentration of X, x_s equals,

$$x_s = \frac{k_1^+ s + k_2^- p}{k_1^- + k_2^+} \tag{6.3}$$

This concentration depends on all the parameters of the network, i.e.

$$x_s = x_s(k_1^+, k_1^-, s, k_2^+, k_2^-, p), \tag{6.4}$$

which is defined at steady state, but this functional dependency of the concentration of X on all parameters of the network extends to time-dependent properties. This dependency can be understood by realizing that the steady-state rate of both reactions depends on x_s such that the pathway flux and steady-state concentration of X depend on all kinetic and environmental parameters. In other words, the steady-state concentration and rates are defined by the entire network, which makes them system properties!

Often the functional dependency of a system property with respect to all the parameters of the system - the kinetic model - is not known in closed form, as in the case of equation 6.3. Yet we can gain insight into how network circuitry and enzyme properties shape responses of system properties to changes in parameters, such as enzyme levels, external signals and nutrients. Control theory is particularly useful for this purpose. There exist two sorts of control theory: engineering and biological control theory. The latter is often referred to

as metabolic control analysis.[1] Do not let the name "metabolic" control analysis fool you; it's application is by now far more broad than it's initial application. It now covers signaling networks, gene networks, and hierarchical networks besides metabolic networks. Engineering and biological control theory has much in common. Here we will mostly outline biological control theory.

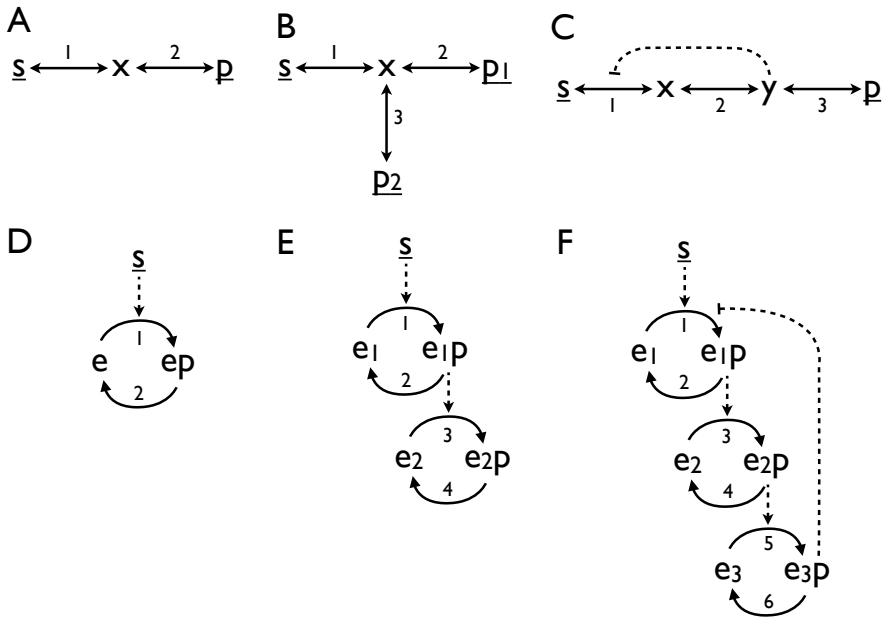

Figure 6.2: Recurrent network structures in molecular networks that will be studied with metabolic control analysis in this chapter. A. A 2-enzyme linear pathway with reversible enzymes. B. A branched pathway with three enzymes. C. A 3-enzyme linear pathway with feedback. D. A signal transduction cycle composed out of a kinase and a phosphatase. E. Two signaling cycles in series. F. Three signaling cycles in series with a negative feedback. All the reactions in the metabolic pathways are reversible and sensitive to their reactants and effectors. All the reactions in the signaling systems are irreversible and only sensitive to their substrate and effectors. The sensitivities of rates to their reactants and their effectors will be quantified with elasticity coefficients that express the fractional change in the rate of a reaction upon a fractional change in the concentration of a reactant or effector at a given state.

[1] A related theory is biochemical systems theory developed by Michael Savageau.

6.3 The machinery of MCA: response, control and elasticity coefficients

The coefficients of metabolic control analysis will be introduced using network A in figure 6.2 as an example. Representative steady-state responses of such a system to a change in the fixed pathway substrate concentration, s, are shown in figure 6.3. The steady-state concentration of x and the steady-state flux J increase with the concentration of s. This is not a surprise, the first enzyme can run faster with more s. As we analyze the system in steady state, the second enzyme needs to operate as fast as the first one. Since, this enzyme is not directly sensitive to s, the rate of enzyme 2 can only increase if the concentration of its substrate, x, increases. The problem becomes a bit more complicated due to the fact that x inhibits the first enzyme through product inhibition. So the exact levels of x depends on properties of both enzymes. It is again a system property and so will be the flux. Without knowing the exact functional dependence of those system properties on the concentration of s can we then still understand the shapes of the plots in figure 6.3? Yes, we can. We can understand the slope of these curves at every value of s in terms of enzyme properties. This is the approach of metabolic control analysis.

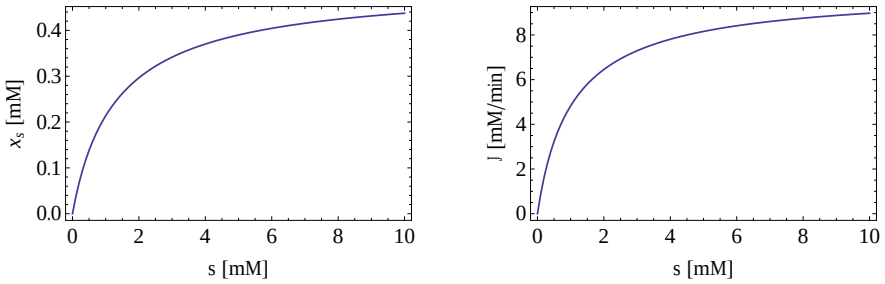

Figure 6.3: The steady state concentration of X and the steady-state flux J for network A in figure 6.2 as function of the fixed concentration of the pathway substrate, s. Both enzyme follow irreversible product-sensitive Michaelis-Menten kinetics with parameters: $V_{max,1} = 10\ mM\ min^{-1}$, $K_{M,1,s} = 1\ mM$, $K_{M,1,x} = 3\ mM$, $V_{max,2} = 50\ mM\ min^{-1}$, and $K_{M,2,x} = 2\ mM$. The concentration of p was set to zero. Note that the subscript s of x_s does not refer to the substrate concentration of the pathway but to the fact that this is the steady-state concentration of x rather than the time-dependent concentration of x.

The slope in the left plot of figure 6.3 is defined for infinitesimally-small changes in s as, $\frac{\partial x_s}{\partial s}$. This sensitivity coefficient is called an unscaled response coefficient in metabolic control analysis. "Unscaled" because in MCA we gen-

erally consider a response coefficient as,

$$R_s^{x_s} = \frac{\partial \ln x_s}{\partial \ln s} = \frac{s}{x_s}\frac{\partial x_s}{\partial s} \tag{6.5}$$

MCA related this systemic coefficient to enzyme properties. For notational convenience we will write the steady state concentration of x, x_s, simply as x. Figure 6.3 shows that this response coefficient is not constant and depends on s as the slope varies with s.

Since, we are interested in the sensitivity of the steady-state concentration of x to s, we rewrite the mass balance for x at steady state with its explicit dependence on s,

$$0 = v_1(s, x(s)) - v_2(x(s)) \tag{6.6}$$

This equation tells you that x depends on s and that: (i) the rate of the first enzyme depends directly in s and indirectly through x and (ii) the rate of the second enzyme depends only indirectly on s through its dependence on x. To study the slope in figure 6.3, we differentiate the previous equation with respect to s,

$$0 = \left(\frac{\partial v_1}{\partial s} + \frac{\partial v_1}{\partial x}\frac{\partial x}{\partial s} - \frac{\partial v_2}{\partial x}\frac{\partial x}{\partial s}\right) ds \tag{6.7}$$

In MCA, all derivatives are scaled using the relationship, $\frac{1}{x}dx = d\ln x$. Scaling the previous equation then gives,

$$\begin{aligned} 0 &= \left(\frac{\partial v_1}{\partial s} + \frac{\partial v_1}{\partial x}\frac{\partial x}{\partial s} - \frac{\partial v_2}{\partial x}\frac{\partial x}{\partial s}\right) ds \\ &= \left(\frac{s}{v_1}\frac{\partial v_1}{\partial s} + \frac{x}{v_1}\frac{\partial v_1}{\partial x}\frac{s}{x_1}\frac{\partial x}{\partial s} - \frac{x}{v_2}\frac{\partial v_2}{\partial x}\frac{s}{x}\frac{\partial x}{\partial s}\right)\frac{ds}{s} \\ &= \left(\frac{\partial \ln v_1}{\partial \ln s} + \frac{\partial \ln v_1}{\partial \ln x}\frac{\partial \ln x}{\partial \ln s} - \frac{\partial \ln v_2}{\partial \ln x}\frac{\partial \ln x}{\partial \ln s}\right) d\ln s \end{aligned} \tag{6.8}$$

In this equation, we have two sorts of coefficients: the response coefficient we have seen before $R_s^x = \frac{\partial \ln x}{\partial \ln s}$ and the normalized sensitivity of a reaction rate to either a variable reactant, $\frac{\partial \ln v_1}{\partial \ln x}$, or a fixed external reactant, $\frac{\partial \ln v_1}{\partial \ln s}$. The latter two coefficients are termed elasticity coefficients in MCA and capture the sensitivity of reactions to reactants and effectors. They are denoted by an epsilon, e.g. $\epsilon_x^{v_1}$, for $\frac{\partial \ln v_1}{\partial \ln x}$.[2] An elasticity coefficient $\epsilon_x^{v_i}$ quantifies the fractional change in the rate of the i-th reaction upon a fractional change in the concentration of reactant or effector, x, when all other intermediates are held fixed at their concentrations of some reference state. So an elasticity coefficient is a property of an enzyme while it operates at some state of reference, often a steady state. For this reason it is referred to as a local property. Response coefficients capture the fractional change in a system property upon a fractional change in a parameter

[2] In some literature, an elasticity coefficient to an external metabolite or fixed concentrations has been termed a 'π' elasticity, we not do this here.

when the entire network, all concentrations, are allowed to attain a new steady state. A response coefficient is therefore a global property, a network property.

Using the introduced terminology, equation 6.8 can be written in terms of MCA notation,

$$0 = \epsilon_s^{v_1} + \epsilon_x^{v_1} R_s^x - \epsilon_x^{v_2} R_s^x \tag{6.9}$$

and solved for the response coefficient,

$$R_s^x = \frac{-1}{\epsilon_x^{v_1} - \epsilon_x^{v_2}} \epsilon_s^{v_1} \tag{6.10}$$

The elasticity coefficient, $\epsilon_s^{v_1}$, is positive as s stimulates the first reaction. The term $\frac{-1}{\epsilon_x^{v_1} - \epsilon_x^{v_2}}$ is positive as well; $\epsilon_x^{v_1} < 0$ (product inhibition) and $\epsilon_x^{v_2} > 0$ (substrate activation). We will come back to the values of those elasticity coefficients in next section.

The response coefficient relationship in equation 6.10 can be decomposed in to the multiplication of two terms, the elasticity coefficient $\epsilon_s^{v_1}$ and a so-far not introduced coefficient, a so-called concentration control coefficient. A moment of reflection on this equation will tell you that (following the rule of partial differentiation),

$$\frac{\partial \ln x}{\partial \ln v_1} = \frac{-1}{\epsilon_x^{v_1} - \epsilon_x^{v_2}} \tag{6.11}$$

This coefficient is called the concentration control coefficient of the first reaction on the concentration of X, denoted by C_1^X. It corresponds to the fractional change in the steady-state concentration of X upon a fractional change in the activity of the first reaction. Changing the activity of a reaction corresponds to changing the forward and backward rate to the same extent as otherwise one would alter the equilibrium constant of the process, which is not defined by kinetics but by the thermodynamic properties of the reactants.

Thus a control coefficient is defined as a system response (of a concentration or flux) to a perturbation of the rate of a reaction. This can be envisioned as a perturbation of a reaction by some multiplication factor,

$$v_1(\lambda) = \lambda v_1 \tag{6.12}$$

at a reference value of $\lambda = 1$. At the moment λ is defined in a very general manner; as a linear parameter that perturbs an entire reaction rate - forward and backward rate simultaneously. In molecular networks the λ parameter can often be thought of as the enzyme concentration; for those reactions that are not catalyzed by enzyme complexes. This holds because for all such enzyme kinetics the rate depends linearly on the total enzyme concentration, i.e. the $V_{MAX}^+ = k_{cat}^+ e$ and $V_{MAX}^- = k_{cat}^- e$. So the λ parameter is nothing mysterious. We can now determine the effect of the rate of a reaction on system properties through control coefficents. The control coefficient on the steady-state concentration of X can be obtained by differentiating with respect to λ,

$$0 = v_1(\lambda, x_s(\lambda)) - v_2(x_s(\lambda)) \tag{6.13}$$

and a scaling this equation,

$$0 = \left(\frac{\partial \ln v_1}{\partial \ln \lambda} + \frac{\partial \ln v_1}{\partial \ln x} \frac{\partial \ln x}{\partial \ln \lambda} - \frac{\partial \ln v_2}{\partial \ln x} \frac{\partial \ln x}{\partial \ln \lambda} \right) d \ln \lambda \tag{6.14}$$

As $\frac{\partial \ln v_1}{\partial \ln \lambda} = 1$ the control coefficient $\frac{\partial \ln x}{\partial \ln \lambda} = \frac{\partial \ln x}{\partial \ln v_1} = C_1^x$ can be solved from the previous equation and equals,

$$C_1^X = \frac{-1}{\epsilon_x^{v_1} - \epsilon_x^{v_2}} \tag{6.15}$$

This is exactly the relationship we were searching for (equation 6.11).

In this section we have defined the main definitions of the coefficients in MCA; elasticity, control and response coefficients. Response and control coefficients measure the relative change in steady-state system properties while elasticity coefficients capture the sensitivity of reactions to reactants and effectors. What we lack is intuition and their application to recurrent network designs of of cellular regulation. This we will train in the next sections by studying metabolism, signaling and gene expression examples.

Exercises

1. Explain why an elasticity coefficient for a product is often negative. What do you expect for an elasticity coefficient to a substrate, competitive inhibitor, and allosteric activator?

2. Make the rate characteristic for the 2-enzyme pathway described in the legend to Figure 6.3, set s to 2 mM. A rate characteristic is obtained by plotting the rate of the first enzyme and the second enzyme as function of x. Verify that you predict from the rate characteristic the same steady state as shown in the left plot of figure 6.3. Determine the elasticities coefficients, $\epsilon_x^{v_1}$, $\epsilon_s^{v_1}$, and $\epsilon_x^{v_2}$. Which enzyme is more sensitive to x? Determine the concentration control coefficient, C_1^x. What does its value tell you?

3. Determine the concentration control coefficient of the second reaction on x. How does this control coefficient relate to C_1^X? Explain what you have found.

6.4 Control coefficients for a linear pathway

In the previous section we have determined the concentration control coefficient of the first reaction (equation 6.15). The concentration control coefficient for the second reaction can be found in the same manner,

$$C_2^X = \frac{1}{\epsilon_x^{v_1} - \epsilon_x^{v_2}} \tag{6.16}$$

It equals $-C_1^X$! It is negative for most enzyme kinetics as typically $\epsilon_x^{v_1} < 0$ and $\epsilon_x^{v_2} > 0$. It has to be negative as it is reasonable to assume that the flux through the pathway will increase upon the addition of more enzyme 2. This can only be achieved at steady state if the rate of the first enzyme also increases, which can only occur if x goes down; hence, $C_2^X < 0$.

We have established that,

$$C_1^X + C_2^X = 0 \tag{6.17}$$

The interpretation of equation 6.17 is that the two enzymes are simultaneously increased in activity to the same extend and that the resultant change in steady state x is zero. This we can easily understand from the mass balance of x at steady state, $v_1(x) - v_2(x) = 0$. As a multiplication of the two rates by the same factor, i.e. $\alpha v_1(x) - \alpha v_2(x) = \alpha 0$, indeed gives the same steady state concentration of x. So there is nothing puzzling about this relationship. In fact it extends to all steady-state concentrations irregardless of the complexity of the molecular network,

$$\sum_i^r C_i^{X_j} = 0 \tag{6.18}$$

The index i runs over all reactions of which there are r in total in the network and holds for all steady state concentrations, X_j. This equation is known as the summation theorem for concentration control coefficients.

A more useful control coefficient for metabolic pathways is a flux control coefficient denoted by $C_{v_i}^{J_k}$ for the control coefficient of the rate of the i-th reaction on the k-th flux. They are defined as,

$$d\ln J_k = \frac{\partial \ln J_k}{\partial \ln v_i} d\ln v_i = C_{v_i}^{J_k} d\ln v_i \tag{6.19}$$

We have already concluded that,

$$J_1 = v_1(e_1, x(e_1)) \tag{6.20}$$

As $J_1 = J_2 = v_1 = v_2$ at steady state we denote the flux by J. We can determine the flux control coefficient from the differentiation of the previous equation to e_1 and scaling the resulting equation,

$$d\ln J = \frac{\partial \ln v_1}{\partial \ln e_1} + \frac{\partial \ln v_1}{\partial \ln x}\frac{\partial \ln x}{\partial \ln e_1} d\ln e_1 \tag{6.21}$$

So we have,

$$C_1^J = \epsilon_{e_1}^{v_1} + \epsilon_x^{v_1} C_1^X = 1 + \epsilon_x^{v_1} C_1^X \tag{6.22}$$

You can easily verify that $\epsilon_{e_1}^{v_1} = 1$ because v_1 is a linear function of e_1.[3] We have already determined the concentration control coefficient, the flux control coefficient then becomes,

$$C_1^J = \frac{-\epsilon_x^{v_2}}{\epsilon_x^{v_1} - \epsilon_x^{v_2}} = \frac{1}{1 - \frac{\epsilon_x^{v_1}}{\epsilon_x^{v_2}}} \tag{6.23}$$

[3] Check this.

So the first enzyme typically has a flux control coefficient larger than 0 and below 1; as $-\frac{\epsilon_x^{v_1}}{\epsilon_x^{v_2}}$ is positive for regular kinetics.[4] For the flux control coefficient of the second enzyme we obtain,

$$C_2^J = \epsilon_{e_2}^{v_2} + \epsilon_x^{v_2} C_2^X = 1 + \epsilon_x^{v_2} C_2^X = \frac{\epsilon_x^{v_1}}{\epsilon_x^{v_1} - \epsilon_x^{v_2}} \tag{6.24}$$

Also this control coefficient is for most enzyme kinetics positive. The observant reader has noticed that,

$$C_1^J + C_2^J = 1 \tag{6.25}$$

Enzyme 1 has a larger effect on the flux when its level is changed - a larger flux control coefficient - than the second enzyme when $C_1^J > 0.5$ as $C_1^J + C_2^J = 1$ and the C^J's are positive. This occurs when $-\frac{\epsilon_x^{v_1}}{\epsilon_x^{v_2}} < 1$.[5] The ratio $-\frac{\epsilon_x^{v_1}}{\epsilon_x^{v_2}}$ indicates the ratio of the sensitivity of the first reaction for x over the sensitivity of the second reaction for x. So the least sensitive enzyme has most flux control! This often translated to more complicated networks.

Why do the flux control coefficients sum to 1? A similar proof as for the sum of the concentration coefficients applies. Consider again $\alpha v_1(x) - \alpha v_2(x) = \alpha 0$, it indicates that the steady state flux J increases to the same extend as the rates where increased, that is by a factor of α. So a simultaneous change of the rates by a factor α causes the flux to change with a factor of α too. Hence, the summation theorem for flux control coefficients (equation 6.25). The summation theorem for flux control coefficients can be generalized to networks of any complexity,

$$\sum_i^r C_i^{J_k} = 1 \tag{6.26}$$

Exercises

1. Determine the elasticity coefficient for the reversible Michaelis-Menten mechanism (equation 4.30) with respect to its substate and product. Write it as a difference between a term that contains the mass-action ratio, Γ/K_{eq} and one that contains the S/K_S and P/K_P terms. Show that close to thermodynamic equilibrium the enzyme properties do not matter for the value of these elasticity coefficients. In fact this is a general property for all reversible enzyme kinetics.

2. Show that the first enzyme in a two-enzyme pathway that is not sensitive to its product has a flux control coefficient of 1. Use the control coefficient expression and a rate characteristic. Do you think this result is limited to pathways of length 2 or does it apply also to larger systems?

[4] Explain why.

[5] Verify this by plotting C_1^J as function of $-\frac{\epsilon_x^{v_1}}{\epsilon_x^{v_2}}$

3. Use the concentration and flux control coefficients derived above to validate that,

$$\begin{aligned} C_1^J \epsilon_x^{v_1} + C_2^J \epsilon_x^{v_2} &= 0 \\ C_1^X \epsilon_x^{v_1} + C_2^X \epsilon_x^{v_2} &= -1 \end{aligned} \tag{6.27}$$

   Those are called connectivity theorems of flux and concentration control coefficients, respectively.

4. Use the connectivity theorems (equations 6.27) to show that an enzyme that is only sensitive to a single metabolite only controls the concentration of that metabolite and not the flux. Such an enzyme is called a slave enzyme. Even though you have now derived it for a 2-enzyme pathway, this is a general result.

5. Use the summation and connectivity theorems to determine the control coefficients! This is one of easiest ways of deriving control coefficients.

6.4.1 Summation and connectivity theorems for linear pathways

The connectivity theorems discovered for the 2-enzyme pathway can be extended to any linear pathway with any number of feedback and feedforward loops,

$$\sum_k^r C_k^{X_l} \epsilon_{X_j}^{v_k} = \delta_j^l,\ \delta_j^l = 1 \text{ if } j = l \text{ else } 0 \tag{6.28}$$

$$\sum_k^r C_k^{J_l} \epsilon_{X_j}^{v_k} = 0 \tag{6.29}$$

The summation theorems hold for molecular networks, including signaling and gene networks, of any complexity,

$$\sum_k^r C_k^{X_l} = 0 \tag{6.30}$$

$$\sum_k^r C_k^{J_l} = 1 \tag{6.31}$$

For linear pathways, all summation and connectivity theorems are sufficient to express all the control coefficients in terms of elasticity coefficients.

Exercises

1. Determine the flux control coefficients of a 3-enzyme linear pathway without feedback. Do the same for the pathway displayed in 6.2C. What is the effect of the feedback on the flux control coefficients?

2. Which enzymes have most flux control in a 3-enzyme linear pathway if the second enzyme is not sensitive to its product?

3. Use the connectivity theorems to show that an enzyme in any linear pathway that is only sensitive to one metabolite only controls the concentration of that metabolite and no other systemic property.

4. A transient time, τ_i, is defined as the concentration of a metabolite divided by the flux through that metabolite; all at steady state. Derive the summation and connectivity theorems for transient times for linear pathways.

6.5 A branched pathway: one robust branch and the other highly sensitive

6.5.1 Control coefficients

Consider figure 6.2B, it displays a three-enzyme metabolic pathway with a branch. Using the by-now-familiar approach of differentiation we can obtain the concentration control coefficient of the first enzyme on x and then subsequently the flux control coefficient. Let's start with the concentration control coefficient. We have the following functional relationship at steady state as a result of the mass balance,

$$0 = v_1(e_1, x(e_1)) - v_2(x(e_1)) - v_3(x(e_1)) \tag{6.32}$$

We can take the derivative of this equation to x,

$$0 = \frac{\partial v_1}{\partial e_1} + \frac{\partial v_1}{\partial x}\frac{\partial x}{\partial e_1} - \frac{\partial v_2}{\partial x}\frac{\partial x}{\partial e_1} - \frac{\partial v_3}{\partial x}\frac{\partial x}{\partial e_1} \tag{6.33}$$

and scale it,

$$0 = v_1\frac{\partial \ln v_1}{\partial \ln e_1} + v_1\frac{\partial \ln v_1}{\partial \ln x}\frac{\partial \ln x}{\partial \ln e_1} - v_2\frac{\partial \ln v_2}{\partial \ln x}\frac{\partial \ln x}{\partial \ln e_1} - v_3\frac{\partial \ln v_3}{\partial \ln x}\frac{\partial \ln x}{\partial \ln e_1} \tag{6.34}$$

As these rates are all steady-state rates, we will denote them as fluxes, J's. After having recognized the elasticity coefficients and the concentration control coefficient, we can solve for the concentration control coefficient,

$$C_1^X = \frac{-1}{\epsilon_x^{v_1} - \frac{J_2}{J_1}\epsilon_x^{v_2} - \frac{J_3}{J_1}\epsilon_x^{v_3}} \tag{6.35}$$

Here we have used $\epsilon_{e_1}^{v_1} = 1$. One should realize that at steady-state, $J_1 = J_2 + J_3$. This equation is the same as equation 6.15 if we consider,

$$\frac{J_2}{J_1}\epsilon_x^{v_2} + \frac{J_3}{J_1}\epsilon_x^{v_3} \tag{6.36}$$

as an overall elasticity coefficient for the degrading reactions of x. So from the perspective of the control of the first reaction on the concentration of x it does not matter how many consuming reactions of x occur.

The flux control coefficient can be obtained again through equation 6.22. Rewriting the result gives,

$$C_1^J = \frac{-\frac{J_2}{J_1}\epsilon_x^{v_2} + \frac{J_3}{J_1}\epsilon_x^{v_3}}{\epsilon_x^{v_1} - \left(\frac{J_2}{J_1}\epsilon_x^{v_2} + \frac{J_3}{J_1}\epsilon_x^{v_3}\right)} \tag{6.37}$$

This equation shows that the control coefficient no longer only depends on the elasticity coefficients but in addition on the flux ratio's. This has an interesting consequences as we shall see in the next section.

6.5.2 Low-flux branches and branch point control

The flux control coefficient of the second reaction on the third is given by,[6]

$$C_2^{J_3} = \frac{-1}{\frac{J_3}{J_2} + \frac{\epsilon_X^{v_2}}{\epsilon_X^{v_3}} - \frac{J_1\epsilon_X^{v_1}}{J_2\epsilon_X^{v_3}}} \tag{6.38}$$

This equation has a few interesting properties. If enzyme 2 is saturated with x such that $\epsilon_X^{v_2} \approx 0$ and the third enzyme has an elasticity coefficient of ≈ 1 for x, we obtain,

$$C_2^{J_3} \approx \frac{-J_2}{J_3 + J_1|\epsilon_X^{v_1}|} \tag{6.39}$$

You may wonder whether these conditions are exotic. They are for many branches actually quite realistic [36, 25]. Consider a branch point where the branching is regulated by the K_M's of the enzyme after the branch point; enzyme 2 and 3. The conditions we are considering corresponds to the case where the high-affinity branch enzyme is saturated at a level of the branch metabolite that equals the K_M of the low affinity enzyme. This equation illustrates that only the branch fluxes and the elasticity coefficient of the first enzyme now determine the control coefficient.

If, in addition, $\epsilon_X^{v_1}$ is small, which can easily happen, the control coefficient becomes

$$C_2^{J_3} \approx -J_2/J_3 \tag{6.40}$$

This control coefficient can become much larger than 1 if the flux through the branch with enzyme 3 is small (then: $J_1 \approx J_2$). So the major pathway has a large control on the flux through the minor pathway. (Interesting information for a biotechnologist!) Otherwise, i.e. $\epsilon_X^{v_1} \approx -1$ and $J_1 \approx J_2$, the equation reduces to $C_2^{J_3} \approx -1$. Alternatively, if none of the elasticity coefficient conditions hold but J_3 is much smaller than J_2, the flux control coefficient becomes,

$$C_2^{J_3} \approx \frac{\epsilon_X^{v_3}}{\epsilon_X^{v_2} + |\epsilon_X^{v_1}|} \tag{6.41}$$

[6] Equate this yourself if you wish

Under this condition, enzyme 3 does not control the concentration of X and therefore also not the flux through enzyme 1 and 2.[7] The branch with the large flux has become ignorant of the small branch. The small branch has become a slave of the large flux pathway, it becomes very sensitive to it when it has the highest sensitivity to the concentration of the branch metabolite X. This is relevant for the production of the production of relevant metabolites for biotechnology such as flavour compounds that typically branch from pathways in central metabolism carrying major fluxes.

Exercise

Show that if $J_1 \approx J_2$ the small branch does not control the concentration of X and therefore it does not control the flux through J_1 and J_2. Plot the control coefficient $C_2^{J_3}$ as function of the ratio $\epsilon_x^{v_2}/\epsilon_x^{v_3}$ for various values of the ratio $\epsilon_x^{v_2}/\epsilon_x^{v_3}$. Consider cases such $J_3 << J_2$. Under what conditions becomes $C_2^{J_3} < -1$? This phenomenon is called branch-point ultrasensitivity.

6.6 A metabolic pathway with negative feedback: homeostasis

We will study in this section network C in figure 6.2. Using the summation and connectivity theorems for flux control,

$$\begin{aligned} C_1^J + C_2^J + C_3^J &= 1 \\ C_1^J \epsilon_x^{v_1} + C_2^J \epsilon_x^{v_2} &= 0 \\ C_1^J \epsilon_y^{v_1} + C_2^J \epsilon_y^{v_2} + C_3^J \epsilon_y^{v_3} &= 0 \end{aligned}$$

we can solve for all the flux control coefficients in terms of the elasticity coefficients. The expression of the flux control coefficient of the third enzyme then becomes,

$$C_3^J = \frac{\epsilon_{x_1}^{v_1}\epsilon_{x_2}^{v_2} - \epsilon_{x_2}^{v_1}\epsilon_{x_1}^{v_2}}{\epsilon_{x_2}^{v_1}\epsilon_{x_1}^{v_2} - \epsilon_{x_1}^{v_1}\epsilon_{x_2}^{v_2} + \epsilon_{x_1}^{v_1}\epsilon_{x_2}^{v_3} - \epsilon_{x_1}^{v_2}\epsilon_{x_2}^{v_3}} \tag{6.42}$$

We will analyze the consequences of the feedback strength ($-\epsilon_{x_2}^{v_1}$) in this section. This we do in a more transparent fashion when we consider the product inhibition of the x_1 on enzyme 1 to be negligible, $\epsilon_{x_1}^{v_1} \approx 0$; the flux control coefficient then equals,

$$C_3^J = \frac{\epsilon_{x_2}^{v_1}}{\epsilon_{x_1}^{v_2} - \epsilon_{x_2}^{v_3}} \tag{6.43}$$

Given the signs of all the elasticity coefficient this flux control coefficient will be positive, indicating that an increase in the rate of enzyme 3 will lead to a higher flux in the new steady state. Assuming that $\epsilon_{x_1}^{v_1} \approx 0$ is not unrealistic. For many pathways equiped with negative feedback on the first reaction it is known that the first reaction is often irreversible and has a low affinity for the

[7] Do you understand why?

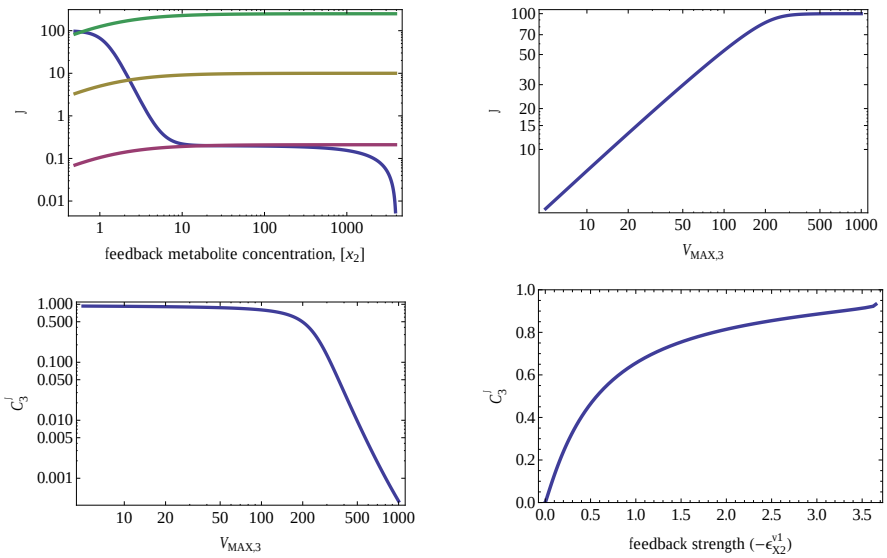

Figure 6.4: Analysis of the consequence of negative feedback for the flux control of metabolic pathways (network C in figure 6.2). In the upper left figure the log-log rate characteristic is plotted of the metabolic segment (supply) composed out of enzyme 1 and 2 (blue line) and three rate curves for the third enzyme, which differ in the maximal rate of the third enzyme (values 0.2, 10, and 250). When this maximal rate is between 0.2 and 250 the steady state lies in the steep region of the rate curve for the supply block. The right upper figure indicates that in this parameter region the dependency between the steady-state flux and maximal activity of the third enzyme is plotted on doubly logarithmic axes. The slope in this curve is the flux control coefficient of the third enzyme, which is for a large range of maximal rate values of the third enzyme constant. This is also shown in the left lower figure. The lower figure on the right indicates that strong feedback inhibition shifts the control to the last enzyme in the pathway.

product, a high K_M [17, 16]. This last equation illustrates that if the negative feedback is strong, i.e. $-\epsilon_{x_2}^{v_1}$ is high, the control shifts to the last enzyme. But what about the control of the second enzyme - since, all the flux control coefficients sum to 1? The metabolite x_1 only influences the second enzyme with a strength of $\epsilon_{x_1}^{v_2}$. Given the connectivity theorems for x_1, that since $\epsilon_{x_1}^{v_1} \approx 0$ simplify to: $C_{v_2}^{J}\epsilon_{x_1}^{v_2} = 0$, $C_{v_2}^{x_1}\epsilon_{x_1}^{v_2} = -1$ and $C_{v_2}^{x_2}\epsilon_{x_1}^{v_2} = 0$, indicate that $C_{v_2}^{J} = 0$, $C_{v_2}^{x_1} = -1/\epsilon_{x_1}^{v_2}$, and $C_{v_2}^{x_2} = 0$. A metabolite that is only sensed by a single enzyme is called a slave metabolite and only control the concentration of that metabolite and nothing else [35]. Since, $C_{v_2}^{J} = 0$ the flux control is distributed between enzyme 1 and 2. If $\epsilon_{x_1}^{v_1} \neq 0$ then the flux control is distributed between all three enzymes.

The previous equations have already suggested that the flux control of the

third enzyme increasing with the feedback strength. This is illustrated in figure 6.4 where the metabolic pathway displayed in figure 6.2C was modelled with reversible Hill kinetics for the first enzyme and reversible Michaelis-Menten kinetics for the other enzymes. As long as the steady state of the metabolic pathway occurs at a concentration of x_2 to which the metabolic segment composed out of enzyme 1 and 2 is really sensitive the flux control lies predominantly in the third reaction. We drew a related conclusion at the end of Chapter 4.

6.7 Ultrasensitivity of signaling networks

6.7.1 A single covalent modification cycle

Figure 6.2D shows an enzyme E that is covalently modified, e.g. phosphorylated, by a dedicated enzyme, e.g. a kinase, into EP and the reverse reaction is catalyzed by another dedicated enzyme, e.g. a phosphatase. Alternatively, the covalent modification may involve ubiquitination, methylation, acetylation, or adenylylation. Note that kinase reaction involves the hydrolysis of ATP into ADP and the transfer of the phosphate to E. The phosphatase reaction liberates the phosphate in the form of inorganic phosphate. In Chapter 4 we already briefly studied this system and concluded that this network can display versatile steady-state input-output relationship between the activity of the kinase and the steady-state level of EP. The biological function of this network is to transduce information. Typically, information about the presence and concentration of a signaling molecule, S, that either acts on the kinetics of the kinase, phosphatase or both. The left plot in figure 6.1 and figure shows a number of input and output characteristics ranging from hyperbolic to sigmoidal and switch-like. Similar plots and the sensitivity of the steady-state of the cycle to the activity of the kinase is shown in figure 6.5. The slopes in those curves can be studied with metabolic control analysis.

To derive the sensitivity of steady-state EP as function of the activity of the kinase, which we consider modulated by a signal concentration, we have to start with the steady state mass balance (we consider product independent kinetics of the kinase and phosphatase),

$$0 = v_k(V_k, e(V_k)) - v_p(ep(V_k)) \tag{6.44}$$

In this equation $v_k(V_k, e(V_k))$ and $v_p(ep(V_k)$ denote the rate equation for the kinase and phosphatase and their dependencies on the maximal rate of the kinase V_k and the concentrations of the unphosphorylated and phosphorylated enzyme. We can take the derivative of this equation to V_k,

$$0 = \frac{\partial v_k}{\partial V_k} + \frac{\partial v_k}{\partial e}\frac{\partial e}{\partial V_k} - \frac{\partial v_p}{\partial ep}\frac{\partial ep}{\partial V_k} \tag{6.45}$$

Since $e + ep$ is conserved, we have $\partial e/\partial V_k = -\partial ep/\partial V_k$,

$$0 = \frac{\partial v_k}{\partial V_k} - \frac{\partial v_k}{\partial e}\frac{\partial ep}{\partial V_k} - \frac{\partial v_p}{\partial ep}\frac{\partial ep}{\partial V_k} \tag{6.46}$$

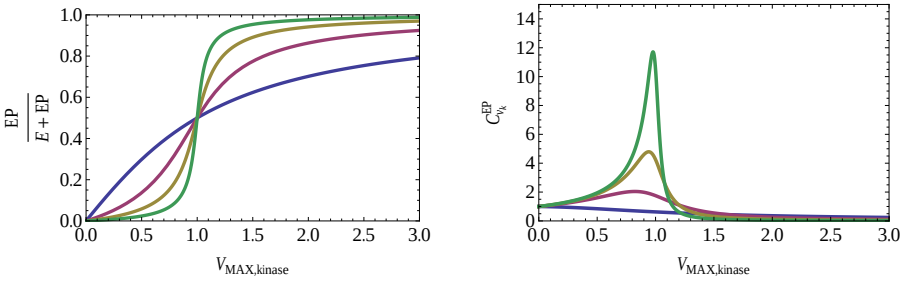

Figure 6.5: Input-output relationship and the sensitivity of a signaling cycle composed out of a kinase and a phosphatase and another enzyme as substrate (network D in figure 6.2). The kinetics was modelled with irreversible Michaelis-Menten kinetics ($K_{M,kinase} = 1$,$K_{M,phosphatase} = 1$, and $V_{MAX,phosphatase} = 1$). In the left plot the input-output relationship is shown as the steady-state phosphorylated fraction as function of the kinase maximal activity. The different lines in the two plots correspond to different total enzyme level (E+EP; 0.5, 5, 15, and 40). Higher concentrations give more sigmoidal input-output relationships. In the right plot, the concentration control coefficient of the kinase on the phosphorylated enzyme concentration is shown, which correspond to the normalized slope in the left plot, $\partial \ln EP / \partial \ln V_{MAX,kinase}$. The maximal rate of the kinase was varied to simulate the action of a signal on this activity. The input-output relationship of the phosphorylation fraction with respect to the signal concentration can then even be steeper than what is shown in the left plot. This occurs when $\epsilon_{signal}^{v_{kinase}} > 1$; for instance, when the kinase is a cooperative enzyme.

MCA requires the normalization of those derivatives,

$$0 = \frac{\partial \ln v_k}{\partial \ln V_k} - \frac{\partial \ln v_k}{\partial \ln e}\frac{ep}{e}\frac{\partial \ln ep}{\partial \ln V_k} - \frac{\partial \ln v_p}{\partial \ln ep}\frac{\partial \ln ep}{\partial \ln V_k} \tag{6.47}$$

In terms of MCA those normalized derivatives become,

$$0 = \epsilon_{V_k}^{v_k} - \epsilon_e^{v_k}\frac{ep}{e}C_{v_k}^{ep} - \epsilon_{ep}^{v_p}\frac{ep}{e}C_{v_k}^{ep} \tag{6.48}$$

As the rate of the kinase depends linearly on its maximal rate, $\epsilon_{V_k}^{v_k} = 1$ and we obtain for the concentration control coefficient of the kinase on EP,

$$C_{v_k}^{ep} = \frac{1}{\epsilon_e^{v_k}\frac{ep}{e} + \epsilon_{ep}^{v_p}} \tag{6.49}$$

This coefficient is the scaled slope $\partial ep / \partial V_k \cdot V_k / ep$ of the left plot of figure 6.5. It becomes larger than 1 when the two elasticity coefficients are small. They are small when the kinase and phosphatase are saturated with their substrate. This typically occurs when the total enzyme concentration, $e + ep$, exceeds the

sum of the K_M's of the kinase and the phosphatase. This agrees with the results in figure 6.5 as the sigmoidality and sensitivity increase with the total enzyme concentration. The response coefficient of the phosphorylated enzyme concentration, ep, to a signal concentration acting on the kinase can be equated in terms of a control coefficient and an elasticity coefficient as,

$$R_s^{ep} = C_{v_k}^{ep} \epsilon_s^{v_k} \tag{6.50}$$

This equation illustrates that the dependency of steady-state ep and s can be steeper or shallower than the dependency of steady-state ep on V_k depending on the elasticity coefficient of the kinase for the signal. When the value of this response coefficients exceeds a signaling cycle displays ultra-sensitivity; it amplifies a change in its input to a large change in its output [23, 11, 3].

6.7.2 A signaling cascade: sensitivity amplification

Equation 6.50 has an interesting consequence. In many cases, signaling cycles as treated in the previous section occur in linear cascades, a well-known example is the MAPK pathway (figure 1.5; and 6.2E). In a cascade the kinase of a signaling cycle is the output, the phoshorylated enzyme, of the previous cycle. So for a cascade of length three we obtain,

$$\begin{aligned} R_s^{e_1p} &= C_{v_{k,1}}^{e_1p} \epsilon_s^{v_{k,1}} \\ R_{ep_1}^{e_2p} &= C_{v_{k,2}}^{e_2p} \epsilon_{e_1p}^{v_{k,2}} \\ R_{ep_2}^{e_3p} &= C_{v_{k,3}}^{e_3p} \epsilon_{e_2p}^{v_{k,3}} \end{aligned} \tag{6.51}$$

As the kinase of cycle 2 and 3 are the previous phosphorylated enzyme species, i.e. e_1p and e_2p, and the kinase activity depends linearly on the concentration of those species, $\epsilon_{e_1p}^{v_{k,2}} = \epsilon_{e_2p}^{v_{k,3}} = 1$. So the sensitivity of the last phosphorylated enzyme, e_3p, to S is given by,

$$R_s^{e_3p} = R_{e_2p}^{e_3p} R_{e_1p}^{e_2p} R_s^{e_1p} \tag{6.52}$$

This equation shows the phenomenon of sensitivity amplification. If each cycle is ultrasensitive, the sensitivity of the output of the entire cascade to its input is higher than any of its components' sensitivity [4, 5, 22].

6.7.3 Signaling cascades with feedback

In the previous section, we illustrated the occurrence of sensitivity amplification along a signaling cascade. Often, feedbacks occur in signaling cascades. As a consequence the sensitivity of a signaling cycle to its input in isolation of the network can be different than with is embedded in the network because of its output eventually modulates its input. This means we have to distinguish the sensitivity of a signaling cycle in isolation from its sensitivity in the network.

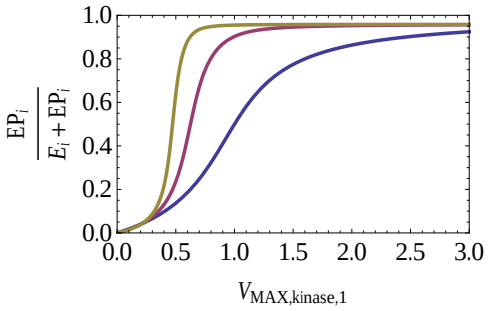

Figure 6.6: Sensitivity amplification of a signaling cascade composed out of a linear chain of signaling cycles. Each of the signaling was modeled in the same way, each has the blue input/output characteristic. The output of the second cycle, $EP_2/(EP_2 + E_2)$ (red line), is more sensitive to a change in the activity of first kinase (at the top of the signaling cascade) and the first cycle and less sensitive than the third (brown line).

In the last section, this was not necessary as we considered a cascade without feedback. We will now consider the network shown in figure 6.2F; a signaling cascade composed out of three levels and a feedback from the output to the input. At steady state the functional dependencies of the concentrations on each others is given by,

$$\begin{aligned} e_1p &= e_1p(e_3p, s) \\ e_2p &= e_2p(e_1p) \\ e_3p &= e_3p(e_2p) \end{aligned} \tag{6.53}$$

These equations may appear a bit strange to you they show that at steady state the concentration of e_1p, as a solution of its steady state mass balance, depends on e_3p and s, e.g.

$$0 = V_{max,1} \cdot \frac{K_D}{K_D + e_3p} \cdot \frac{s}{K_s + s} \cdot \frac{e_{1,total} - e_1p}{e_{1,total} - e_1p + K_{k,1}} - V_{MAX,ph,1} \cdot \frac{e_1p}{K_{ph} + e_1p} \tag{6.54}$$

Here the first term gives the rate equation for the kinase of e_1, with maximal rate $V_{max,1}$! The activity of this kinase depends on s and e_1 ($e_{1,total} - e_1p$). But in addition to e_3p as an inhibitor; here we modelled it such that e_3p and the kinase of e_1 can form a complex with dissociation constant K_D and the complexed kinase has no activity. If we solve e_1p from this equation we conclude that the dependency, $e_1p = e_1p(e_3p, s)$, is indeed correct. The other dependencies can be understood in similar ways. A change in the level of the signal, s, will bring about a change in the steady state level of e_1p, which brings about a change in

e_2p, that effects e_3p, that feedback to e_1p; see we have,

$$\begin{aligned} de_1p &= \frac{\partial e_1p}{\partial e_3p} de_3p + \frac{\partial e_1p}{\partial s} ds \\ de_2p &= \frac{\partial e_2p}{\partial e_1p} de_1p \\ de_3p &= \frac{\partial e_3p}{\partial e_2p} de_2p \end{aligned} \tag{6.55}$$

We can normalize those derivatives again and divide by $d\ln s$,

$$\begin{aligned} \frac{d\ln e_1p}{d\ln s} &= \frac{\partial \ln e_1p}{\partial \ln e_3p}\frac{d\ln e_3p}{d\ln s} + \frac{\partial \ln e_1p}{\partial \ln s} \\ \frac{d\ln e_2p}{d\ln s} &= \frac{\partial \ln e_2p}{\partial \ln e_1p}\frac{d\ln e_1p}{d\ln s} \\ \frac{d\ln e_3p}{d\ln s} &= \frac{\partial \ln e_3p}{\partial \ln e_2p}\frac{d\ln e_2p}{d\ln s} \end{aligned} \tag{6.56}$$

All the $d\ln e_ip/d\ln s$ factors are (global) response coefficients, $R_s^{e_ip}$. The partial derivatives we have not encountered before, those we will terms local response coefficients, denoted by $r_Y^X = \partial \ln X/\partial \ln Y$. The term "local" refers to the fact that those response coefficients only denote the response at the level of a single signaling unit, not at the level of entire network circuit that is captured by the response coefficients, $R_s^{e_ip}$. With those definitions the previous equation becomes,

$$\begin{aligned} R_s^{e_1p} &= r_{e_3p}^{e_1p} R_s^{e_3p} + r_s^{e_1p} \\ R_s^{e_2p} &= r_{e_1p}^{e_2p} R_s^{e_1p} \\ R_s^{e_3p} &= r_{e_2p}^{e_3p} R_s^{e_2p} \end{aligned}$$

With those equations we can express the response coefficients at the level of the entire network, the R's, in terms of the response properties of its signaling components, the r's. We obtain then for the sensitivity of the output of the signaling network, e_3p to s,

$$R_s^{e_3p} = \frac{r_{e_2p}^{e_3p} r_{e_1p}^{e_2p} r_s^{e_1p}}{1 - r_{e_2p}^{e_3p} r_{e_1p}^{e_2p} r_{e_3p}^{e_1p}} \tag{6.57}$$

This equation illustrates that the response of the cascade without the feedback loop, i.e. $r_{e_2p}^{e_3p} r_{e_1p}^{e_2p} r_s^{e_1p}$, is reduced by the feedback strength of the entire loop $r_{e_2p}^{e_3p} r_{e_1p}^{e_2p} r_{e_3p}^{e_1p}$. Note that $r_{e_3p}^{e_1p} < 0$ for a negative feedback loop, the phosphorylation level of e_1p is reduced upon a increase in e_3p.

Chapter 7

Dynamics of molecular systems

7.1 Stability of steady states

So far, we have considered how the changes in the concentrations of molecules inside cells can be expressed in terms of reactions rates and how those rates depend in turn on concentrations of molecules and kinetic rate constants. We have considered enzyme and massaction kinetics. We have concluded that steady states are often stable states and limited ourselves to the study of these systems. We came to this conclusion by considering a single variable system as an example. We considered a single molecule X and plotted in the same figure it's rate of synthesis and degradation, resp. v_s and v_d, as function of the concentration of X, denoted by x. The intersection of the rate curves then indicated a steady state. We denote the steady state concentration of X at this intersection by x_s, thus we have $v_s(x_s) = v_d(x_s)$. We concluded that this steady state was stable if two conditions were met: i. for $x < x_s$, we required $v_s > v_d$ and ii. for $x > x_s$, we required $v_s < v_d$. Both conditions make sure that the systems evolves in the direction of x_s.

Exercise

Sketch the plot that was described in the previous paragraph and make sure you understand it.

We will now write those conditions for stability in more mathematical terms to be able to work with more complicated models later and to have a proper measure for (in-)stability. The dynamics of X is described by,

$$\frac{dx}{dt} = v_s(x) - v_d(x) \tag{7.1}$$

We are interested in figuring out what happens to the concentration of X when it is perturbed slightly from its steady state value, x_s, by a value δx. If after some time it returns to x_s we call the steady state stable and otherwise unstable. So we want to know how

$$\frac{d(x_s + \delta x)}{dt} \tag{7.2}$$

evolves, i.e. whether it converges to zero (stability) or not (instability).

Here the definition of a derivative comes to the rescue, for small enough values of δx, this definition states,

$$\frac{f(x + \delta x) - f(x)}{\delta x} \approx \frac{\partial f}{\partial x}$$

and therefore,

$$f(x + \delta x) \approx f(x) + \frac{\partial f}{\partial x}\delta x \tag{7.3}$$

By following this logic, and taking dx_s/dt as our $f(x)$, we obtain for equation 7.2,

$$\frac{d(x_s + \delta x)}{dt} = \frac{dx_s}{dt} + \frac{\partial}{\partial x}\frac{dx_s}{dt} = 0 + \left(\frac{\partial v_s}{\partial x} - \frac{\partial v_d}{\partial x}\right)\delta x \tag{7.4}$$

Since, $\frac{d(x_s+\delta x)}{dt} = \frac{dx_s}{dt} + \frac{d\delta x}{dt} = 0 + \frac{\delta x}{dt}$ we get,

$$\frac{d\delta x}{dt} = \left(\frac{\partial v_s}{\partial x} - \frac{\partial v_d}{\partial x}\right)\delta x \tag{7.5}$$

This is a linear differential equation and can be solved for $\delta x(t)$ as function of time. We denote $\frac{\partial v_s}{\partial x} - \frac{\partial v_d}{\partial x}$ by λ. To understand the concept of λ as a stability measure we solve the previous differential equation,

$$\begin{aligned}
\frac{d\delta x}{\delta x} &= \lambda dt \\
\int_{\delta x(0)}^{\delta x(t)} \frac{d\delta x}{\delta x} &= \int_0^t \lambda dt \\
\ln \delta x(t) - \ln \delta x(0) &= \lambda t \\
\delta x(t) &= \delta x(0) e^{\lambda t}
\end{aligned} \tag{7.6}$$

From the last equation you can see that the perturbation, i.e. the applied $\delta x(0)$, dies out if and only if $\lambda = \frac{\partial v_s}{\partial x} - \frac{\partial v_d}{\partial x} < 0$; because $e^{\lambda t}$ converges to 0 if $\lambda < 0$ for large enough times! And this will be very often the case when you consider the signs of $\frac{\partial v_s}{\partial x}$ and $\frac{\partial v_d}{\partial x}$, which are typically negative and positive, respectively, for realistic rate equations. In mathematics, the λ parameter is called an eigenvalue.

Steady states can become unstable upon a change in a parameter. A location in parameter space where a steady state becomes unstable is an example of a bifurcation point. When a steady state becomes unstable several phenomena can occur. We will consider two such phenomena: i. the system jumps to

a distant steady state (saddle-node bifurcation; associated with bistability) or ii. the system starts to display regular oscillations (Hopf bifurcation). We will start with studying bistable systems. Bistability is often associated with systems that have a positive feedback loop whereas oscillations typically involve negative feedback.

Exercise

1. Explain in terms of the sketch of the previous exercise that the condition of $\lambda < 0$ makes sense for stability. And explain why $\frac{\partial v_s}{\partial x}$ is often negative and why $\frac{\partial v_d}{\partial x}$ is often positive for realistic rate equations (mass action and enzyme kinetics).

2. Determine whether the following systems have a stable steady state. Determine also the steady state concentration of X. You can use Mathematica or do it by hand.

   - $\frac{dx}{dt} = k_1^+ s - k_1^- x - (k_2^+ x - k_2^- p)$ with the k^+'s as 10 and k^-'s as 1, s equals 10 and p equals 1.
   - $\frac{dx}{dt} = \frac{1}{1+x} - \frac{x}{1+x}$
   - $\frac{dx}{dt} = \frac{100(1-x/10)}{1+3+x} - \frac{x}{1+x+2}$ (By the way, what kind of kinetics do these rate equations suggest?)

3. Consider the following system $\frac{dx}{dt} = \underbrace{5 + \frac{20x^5}{1+x^5}}_{\text{synthesis rate}} - \underbrace{15x}_{\text{degradation rate}}$. This one is a bit more complicated. Plot in Mathematica the synthesis and the degradation rates as function of the x. How many intersections do you count? Those are steady states. Which of those are stable and which are unstable?

4. Consider again $\frac{dx}{dt} = v_s(x) - v_d(x)$:

   - Does strong product inhibition make the system more stable or less stable?
   - Can a system with product inhibition and substrate activation become unstable?
   - Can a system with product activation and substrate activation become unstable?
   - Can a system with product inhibition and substrate inhibition become unstable?

5. Consider equations 7.6, does a system with a more negative λ return to steady state faster or slower than one with a larger (but also negative) λ value?

6. What happens to $\delta x(t)$ as a function of time for a system with a positive λ and what happens in the case of a negative λ?

7.2 Bistable dynamics of single-variable systems

Bistability is a phenomenon which occurs very often in cell biology, it has been found in various signaling and gene networks. For instance in the MAPK pathway in oocytes, the galactose regulon in yeast, and in the lac operon in *E. coli*. It is often associated with systems that have a positive feedback. Bistability appears a bit counterintuitive at first sight but it is not very hard to understand it in terms of a mathematical model. We will explain it in this section.

Figure 7.1 show a simple model with autocatalytic synthesis and linear degradation of some molecular species. This model could for instance model a transcription factor, which activates the transcription of its own mRNA. The net transcription rate is modeled as the sum of a basal rate and the influence of the transcription factor on transcription, i.e. $5 + \frac{20x^5}{1+x^5}$, and the degradation is first order, i.e. $15x$ (indicating that per unit time 15 mRNAs are degraded),

$$\frac{dx}{dt} = \underbrace{5 + \frac{20x^5}{1+x^5}}_{\text{synthesis rate}} - \underbrace{15x}_{\text{degradation rate}} \tag{7.7}$$

This system is capable of generating three steady states: two are stable and one is unstable.

Let's first determine the steady states of this system (equation 7.7). This means we have to solve,

$$0 = 5 + \frac{20x^5}{1+x^5} - 15x$$

for x. This is a frustrating exercise by hand, so we use the Mathematica function Solve and select only the positive solutions (i.e. those values of x that correspond to the intersections with the $dx/dt = 0$ axis in Figure 7.1). You will then find 0.34, 1, and 1.52 (Hint: do this yourself).

The question now is which of those steady states are stable? A graphical explanation is given in the legend to Figure 7.1: simply from the sign of dx/dt right and left from a steady state you can determine whether the state is attracting (stable) or expelling (unstable). But we can also calculate the eigenvalue and determine its sign. This means we have to determine,

$$\lambda = \frac{\partial}{\partial x}\left(5 + \frac{20x^5}{1+x^5}\right) - \frac{\partial}{\partial x}15x \tag{7.8}$$

This you can with the derivative function (D) in Mathematica; this gives (test this!),

$$\lambda = -15 - \frac{100x^9}{(1+x^5)^2} + \frac{100x^4}{1+x^5} \tag{7.9}$$

To determine whether the steady state $x_s = 0.34$ is stable we computed the λ for this value of x, which gives: -13.7. So this state is stable, λ is negative. For $x_s = 1$, we find $\lambda = 10$. This is an unstable steady state! The last steady state is stable because it gives $\lambda = -8.6$.

$$\frac{dx}{dt} = \underbrace{5 + \frac{20x^5}{1+x^5}}_{\text{synthesis rate}} - \underbrace{15x}_{\text{degradation rate}}$$

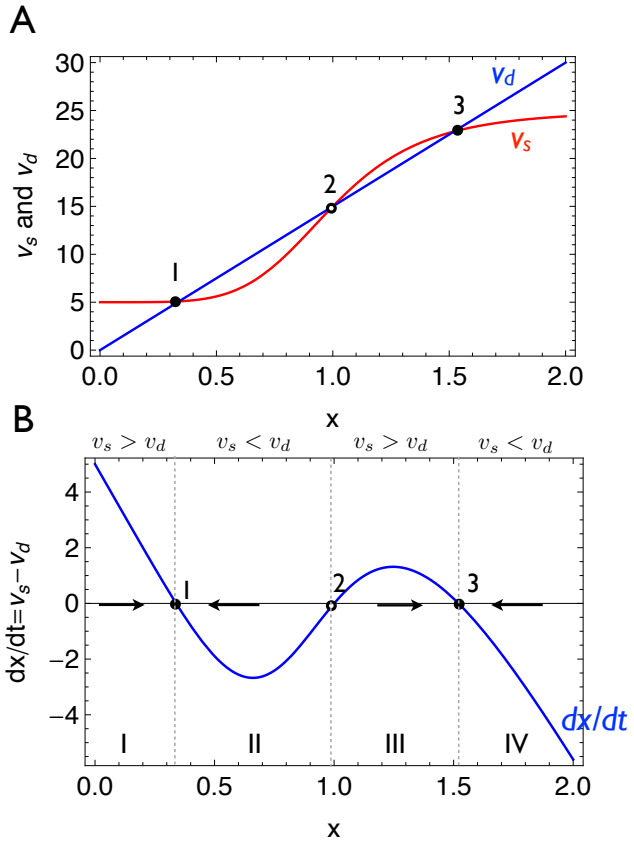

Figure 7.1: **Explanation of bistability.** We consider a system that has autocatalytic synthesis and linear degradation. In A the rates of synthesis (red line) and degradation (blue line) are drawn as function of x. Three intersections occur labelled 1, 2, and 3 and those are the steady states of system. The closed circles denote stable steady states and the open circle denote an unstable steady state. In B the explanation of the stability properties of the steady states is given. On the basis of the three steady states we distinguish four regions labelled I to IV. In region I: $v_s > v_d$ (which can be seen from plot A) and therefore the concentration of x will rise until $v_s = v_d$ (steady state 1); in region II: the concentration of x drops because $v_s < v_d$ until $v_s = v_d$ (steady state 1); in region III: the concentration of x rises again until $v_s = v_d$ (at steady state 3) and finally in region IV: the concentration drops until steady state 3 is reached. This means that steady state 2 is propelling and is never reached, so it is unstable.

Exercises

1. Suppose a synthesis rate of x is inhibited by x as $\frac{1}{1+x}$. What should the shape of the degradation function be to make the system bistable? How would you call such kinetics? Is it realistic? Which steady state would be stable and which ones would be unstable? Use the graphical method to decide this.

2. Take again the model from Figure 7.1 and take it as an example of transcription model. Suppose now that a second transcription regulator, y, influences the maximal influence of x on its own transcription rate, i.e. the factor 20 in the synthesis rate of x. Suppose the dependency is is like this, $v_s = 5 + y\frac{x^5}{1+x^5}$. Investigate in Mathematica the influence of y on the number of steady states. Describe what happens do you find steady state with high values of x, low values of x, or both? Or does this depend on the exact concentration of the second transcription regulator y?

7.2.1 Emergence and disappearance of bistability as function of a parameter

Not all systems with a positive feedback will be bistable. The feedback only suggests the possibility for bistability. This means we can control the emergence of bistability with kinetic parameters! This is what this subsection is all about. The parameter we will consider is the affinity of the synthesis process for the x, which we denote with K,

$$\frac{dx}{dt} = 5 + \frac{20x^5}{K^5 + x^5} - 15x \tag{7.10}$$

In the previous sections, K was chosen as 1. A higher K value means a lower affinity and a lower K a higher affinity. In figure 7.2, we decrease and increase the affinity constant and find that the bistability disappears; only a single steady state is now possible. The synthesis rate curve is either shifted to the left or the right, which in both cases forces a single intersection with the degradation rate curve. This indicated that the number of steady states changes as function of K! This we show in Figure 7.3 in a so-called bifurcation plot. These kinds of plots are very informative and can be experimentally measured (Van Oudenaarden Lac operon, Gal regulon; and Ferrell/Oocytes). Depending on the parameters S or Ƨ shaped curves can be found. Note that for complicated systems with a lot more variables bistability remains qualitatively the same behavior as it was discussed in this section. Calculation of the bifurcation diagrams is then often a bit more involved and then researchers often resort to dedicated software such as Auto and XPaut. With a little effort a continuation algorithm can be programmed in Mathematica that does parameter scans and make S/Ƨ-shaped bifurcation curves.

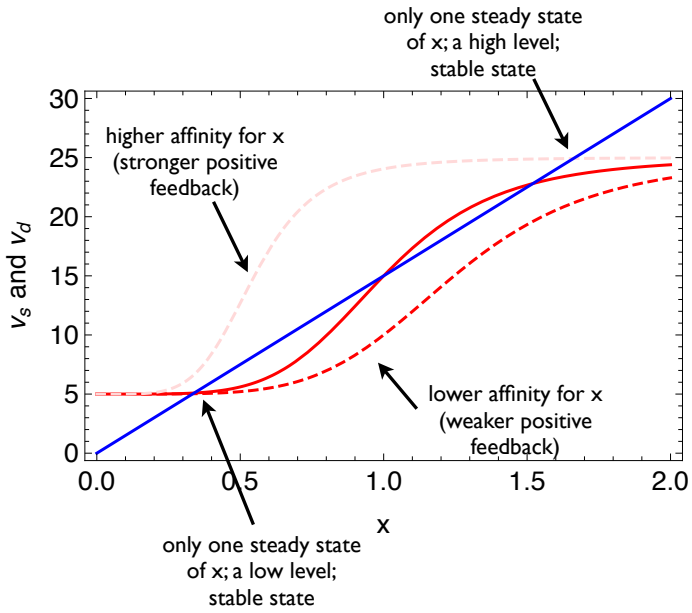

Figure 7.2: **Emergence and disappearance of bistability as function of a parameter.** This plot refers to the same model as analyzed in figure 7.1. Here we consider three different positive feedback strengths of x on its own synthesis rate - the two dashed lines. We modulate the feedback strength by changing the affinity of the synthesis process for x, i.e. the 1 in equation 7.7. The observant reader will note that the 1 in fact corresponds to the affinity raised to the 5-th power as we are considering Hill kinetics. But this does not change our argumentation. The light red dashed curve, has an affinity parameter 0.05 whereas the other dashed line has value 3; a stronger and weaker feedback, respectively. As a result of those parameter changes the number of possible steady state has shifted from 3 to 1. Note that in the two cases different steady state of x will be reached.

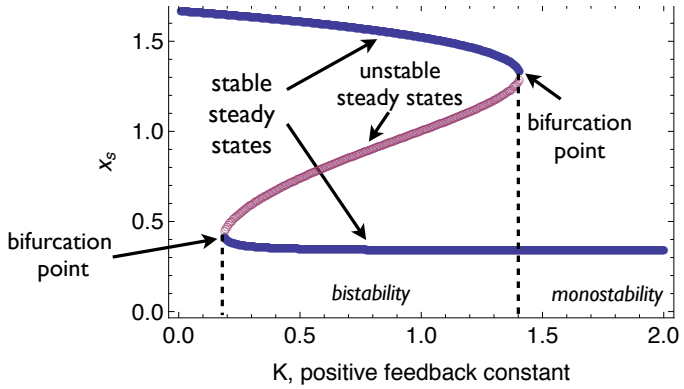

Figure 7.3: **A bifurcation diagram indicating bistability.** For different values of the feedback parameter K the steady state were determined. Each steady state was checked for stability: stable steady state are denoted by the blue line and unstable steady state are denoted by the purple line. As function of K, the system starts in a monostable region, then enters a bistable region through a bifurcation (a so-called saddle-node bifurcation) at a critical value for K, and a monostable region again follows after a saddle-node bifurcation at a second critical parameter value of K.

7.2.2 Bistability as a mechanism for memory

What are the benefits that bistability offers for cells? It has at least two advantages for living cells. Firstly, it is a mechanism for a population of cells to generate two subpopulations: one in the high state and another in the low steady state. But you can then ask yourself, how can this happen? Shouldn't all cells have the same kinetic parameters and therefore be in the same state? If cells would behave deterministically you would be right but cells often do not. We will not discuss this in great depth now but cells have the tendency to carry out signaling and gene-regulation processes at low concentrations of the participating molecules such that these processes have a strong stochastic component. This stochasticity causes cells with the same genome and growth history to function differently and end up in a different steady state when the system is bistable. For experimental examples, see the Lac Operon work by Van Oudenaarden or the sporulation switch in Bacillus subtilis. Secondly, bistability gives some memory of previous events. This is a deterministic property and can therefore be illustrated with differential equations. This we will be explained next.

So memory, where does that come from? Consider again Figure 7.3 and imagine that we slowly increase the value of the parameter K. This means we start in a high-steady state value for x_s and that it slowly decreases. At the bifurcation point, the system jumps to the lower steady state branch and enters the monostable region on the right. Now we decrease the parameter K and we

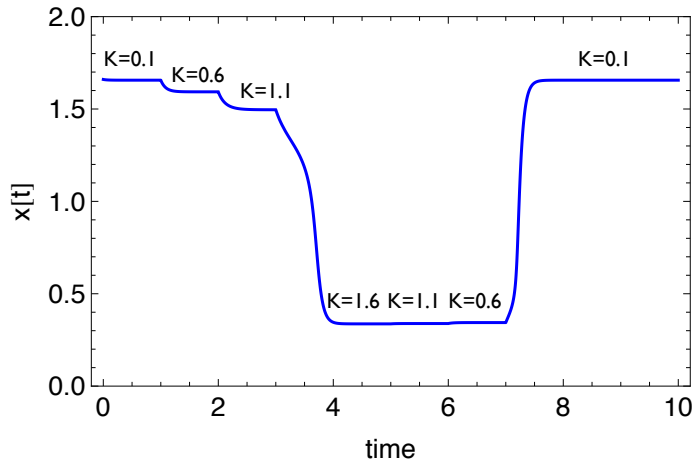

Figure 7.4: **Illustration of hysteresis (memory).** Bistable systems can have different states for the set of parameters depending on their history. This is illustrated in this figure where the dynamics of X is simulated upon a stepwise decrease of K followed by a stepwise decrease in this parameter. This figure can be understood if figure 7.3 is taking into account because the systems "walks" over this $\mathcal{S}$-curve from left to right in this simulation. As you can see depending on the history the system reaches different steady states at K values 0.6 and 1.1. It depends whether the system came from a high or low steady state value for x.

remain on the lower branch of steady states until we hit the other bifurcation point; see Figure 7.4. This means that depending on the history, i.e. starting a high or a low value of K, the state of the system is different! This is memory! (Also sometimes referred to as hysteresis.) This is intriguing isn't it? We have created a simple molecular network with memory. In synthetic biology such devices have been constructed as well, Gardner Nature.

Exercises

Joost, enkele op basis van de snijers? Laat ze dan eerst de G functie zelf even afleiden, die kennen ze nog niet.

7.3 Stability of two variable systems

So far, we have only considered systems with a single variable concentration. Some of you may think that those are not the most realistic systems but this is not always true. Especially in gene networks those systems can be relevant; a bunch of genes can be under the control of a single transcription factor with auto-regulatory behavior leading to bistability (an example is the *lac* operon in

Escherichia coli).

Another dynamic behavior often found in biological systems is oscillations and to understand this dynamics we have to consider minimally two variables. To understand the stability of steady state of systems with two variables we can again consider the signs of eigenvalues but this we will postpone for later. For we will start with a simpler approach called phase plane analysis.

7.4 Phase plane analysis for two variable dynamic systems

Phase-plane analysis is a useful graphical method to determine the number of steady states, their stability properties, and whether oscillations or bistability can occur in the system.

To introduce phase-plane analysis, we start with a general description of the dynamics of a two-variable dynamic system,

$$\begin{aligned} \frac{dx}{dt} &= f(x,y) \\ \frac{dy}{dt} &= g(x,y) \end{aligned} \tag{7.11}$$

Here $f(x,y)$ contains rate equations, i.e. the net rate of synthesis of x as function of x and y minus the net degradation rate of x as function of x and y. The same applies for $g(x,y)$. So nothing new.

A specific example is the dynamic description of the following chemical reaction system. The reactions are[1],

$$\begin{aligned} \underline{A} &\overset{v_1}{\rightleftharpoons} X \\ \underline{B} &\overset{v_2}{\rightarrow} Y \\ 2X + Y &\overset{v_3}{\rightarrow} 3X \end{aligned} \tag{7.12}$$

We consider the concentration of A and B fixed and the mass balances for X and Y are now given by,

$$\begin{aligned} \frac{dx}{dt} &= k_1^+ a - k_1^- x + k_3 x^2 y = f(x,y) \\ \frac{dy}{dt} &= k_2 b - k_3 x^2 y = g(x,y) \end{aligned} \tag{7.13}$$

We choose this example as it has a minimal mathematical complexity, which allows us to focus more on the essence of the phase plane approach.

At steady state, when $x = x_s$ and $y = y_s$ both of these equations are zero by definition,

$$\begin{aligned} 0 &= f(x_s, y_s) \\ 0 &= g(x_s, y_s) \end{aligned} \tag{7.14}$$

[1]This section follows closely a section in *Mathematical Biology* by J D Murray

This means that in the (x, y)-plane the two lines defined by $f(x, y) = 0$ and $g(x, y) = 0$ intersect at steady states. These two lines are called nullclines. Multiple intersections between the nullclines can occur, which indicates the occurrence of bistability for example. This (x, y)-plane is called the phase plane.

For our explicit example setting the equations 7.13 to zero allows us to solve for the steady state concentration of X and Y, those are (I used Mathematica to find those),

$$\begin{aligned} x_s &= \frac{ak_1^+ + bk_2}{k_1^-} \\ y_s &= \frac{k_2 b (k_1^-)^2}{(ak_1^+ + bk_2)^2 k_3} \end{aligned} \tag{7.15}$$

This means only a single intersection in the phase plane occur for the chemical reaction system.

For every (x, y) point in the phase plane $(dx/dt, dy/dt)$ will have a value and "point" in a direction: x and y can go up or down and remain fixed at the steady state. This means that if we start in state $(3, 4)$ (i.e. where $x = 3$, $y = 4$) then after some time δt we are in state $(3 + \frac{dx}{dt}\delta t, 4 + \frac{dy}{dt}\delta t)$, which is a new point in the plane at which new values for $(dx/dt, dy/dt)$ hold, which define a new direction, etc. You get the picture: the dynamics of the system - sometimes called "flow" - can be visualized onto the (x, y)-plane, the so-called phase plane. After long enough times, the system may end up in a point where $(dx/dt, dy/dt)$ equals $(0, 0)$ and then the system is stuck in a stable steady state. If may also happen that the system is initially attracted to a state and then expelled, etc. Or else, the system never settles on a steady state, but ends up circling around a state. Then the system oscillates as function of time and it is said to have settled on a "limit cycle". A limit cycle means that after a time limit (i.e. long enough times) the system cycles (=oscillates). A stable steady state is sometimes called a fixed point.[2]

We will now return to our example system and study its phase plane characteristics and the associated dynamics of x and y. First we need to determine the nullclines for this specific example. This means we have to set equations

[2]Clearly oscillations are periodic (with one or more frequencies). Dynamics can also occur that is not periodic and therefore never returns to the same state. If it would return to a state the system has visited before the system would again display the dynamics of the intervening period because the system is deterministic (for every (x, y, z) state exists only one $(dx/dt, dy/dt, dz/dt)$). Systems that do not settle on a fixed point or a limit cycle and which are therefore not periodic are called chaotic systems. The minimal dimension for chaotic systems is three variables; essentially because in a cube you can draw a line (a "strange" attractor), which represents the flow of the dynamic system, that has infinite length and never intersects with itself. Chaos is an intriguing dynamics but not very important for molecular systems biology, sorry. ☺

7.13 to zero and solve for y. We then obtain for the nullclines,

$$\begin{aligned} y = f^{-1}(x) &= \frac{k_1^- x - a k_1^+}{k_3 x^2} \\ y = g^{-1}(x) &= \frac{b k_2}{k_3 x^2} \end{aligned} \tag{7.16}$$

As a sanity check we can determine the intersection point, where the system is at steady state,

$$\begin{aligned} \frac{k_1^- x_s - a k_1^+}{k_3 x_s^2} &= \frac{b k_2}{k_3 x_s^2} \Rightarrow \\ k_1^- x_s - a k_1^+ &= b k_2 \Rightarrow \\ x_s &= \frac{b k_2 + a k_1^+}{k_1^-} \end{aligned} \tag{7.17}$$

And indeed this agrees with our earlier findings.

In figure 7.5 we plot the dynamics of the system for a set of parameters that gives rise to oscillations. The phase plane is also displayed with the nullclines (red and blue) and the dynamics of figure A (in black). The gray lines with arrows indicate the direction of the dynamics of the 2-variable system. These arrow indicate around the steady state that it is unstable; as the arrow moves away from it in an oscillatory motion towards the limit cycle. In other words, a phase plane allows you to study the stability of steady states when in every point in the plane the direction of flow in calculated, i.e. dx/dt and dy/dt are determined in every point. This makes it a very useful method. However, it does not immediately indicate the effect of parameter changes on qualitative changes in dynamics (called bifurcations), e.g. the appearance or disappearance of bistability and oscillations. To achieve this we to construct a bifurcation diagram (like we did for in the previous section for bistability, i.e. the S-curve).

Figure 7.6 indicates that the kinetic parameters k_1^+ is an interesting parameter to study bifurcations. In figure 7.7 the bifurcation diagram for x as function of the bifurcation parameter ak_1^+ is displayed. It is customary to indicate the amplitude of the oscillations in such a figure as well. At a value of ak_1^+ around 0.18 the oscillations suddenly disappear. For ak_1^+ values just below 0.18, the oscillation were already occurring with low amplitude. After this value, the steady states are stable and oscillations are absent. To be able to reproduce we have to learn a trick to determine the stability of steady states of two variable systems.

7.5 Stability of steady states of dynamic systems with two variables

At the beginning of this chapter the eigenvalue was introduced as stability measure. When it was positive the associated steady states was unstable and other-

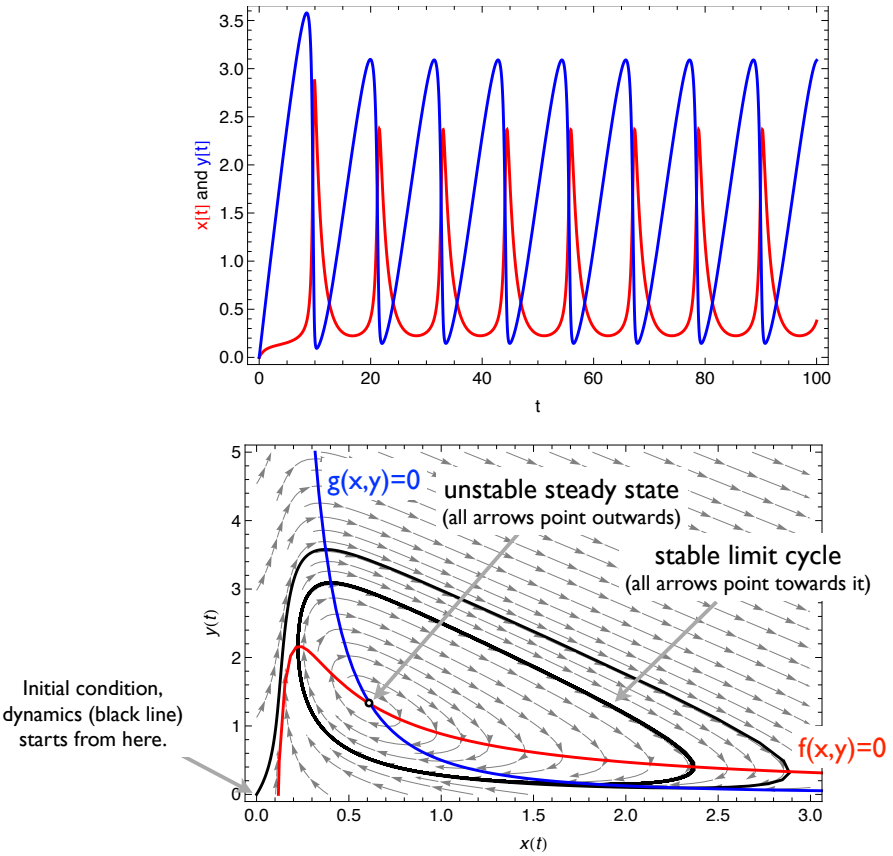

Figure 7.5: **Explanation of phase plane analysis.** We consider the dynamics of the molecules X and Y that engage in the chemical reactions given in equation 7.12. We choose $k_1^+ a = \frac{1}{5\sqrt{3}}$, $k_1^- = 1$, $k_2 = 0.5$ and $k_3 = 1$ to obtain figures A and B. In A the oscillatory dynamics of the concentrations of X and Y is shown as function of time. In Figure B, the phase plane is shown. Three curves are visible. In black the oscillatory dynamics of x and y as function of time is shown in the (x, y)-plane. In red, the equation $dx/dt = 0$ is shown; in other words on this line lie values of x and y that together make $dx/dt = 0$. The blue curve achieves the same for $dy/dt = 0$.

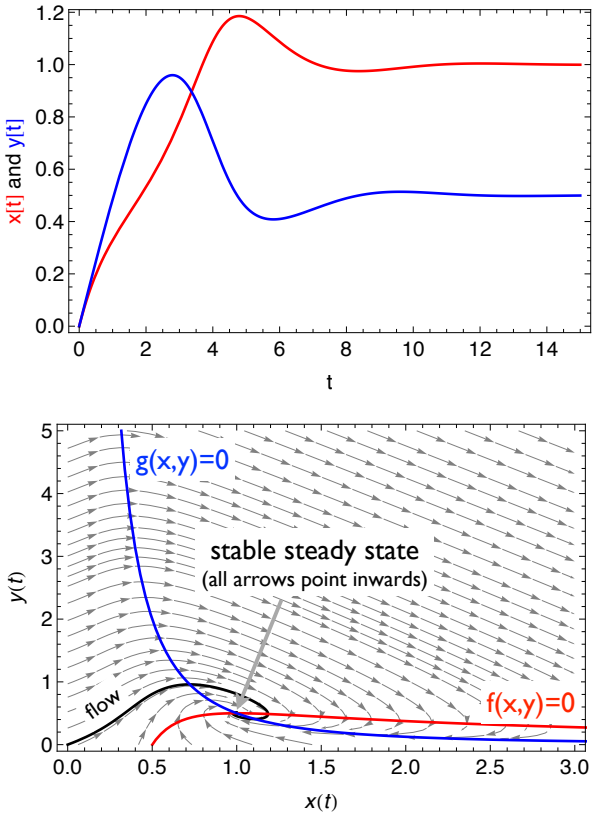

Figure 7.6: **Disappearance of oscillations upon parameter change.** A change in kinetic parameter ak_1^+ to 0.5 leads to disappearance of oscillatory behavior and the steady state become stable.The black line indicates the dynamics ("flow") of the system as shown in the upper figure.

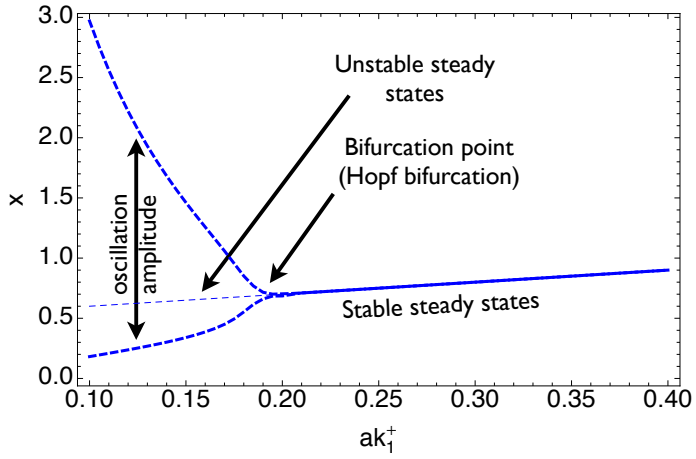

Figure 7.7: **Bifurcation diagram.** The kinetic parameter ak_1^+ was varied and the stability of the steady states were checked. Below a critical value of this parameter oscillations were found and above this bifurcation point stable (non-oscillating states were found). The thick-dashed lines indicate the amplitude of the oscillations. A similar figure can be obtained for y as function of ak_1^+.

wise stable. For two variables we have to consider two eigenvalues. The number of eigenvalues always equals the number of (independent) variables.

To explain the stability of the steady states of dynamic systems with two variables we have to start a bit technical. We consider again the mass balances of the molecules,

$$\begin{aligned} \frac{dx}{dt} &= f(x,y) \\ \frac{dy}{dt} &= g(x,y) \end{aligned} \tag{7.18}$$

The steady state is defined as the combination of concentrations (x_s, y_s) such that,

$$\begin{aligned} 0 &= f(x_s,y_s) \\ 0 &= g(x_s,y_s) \end{aligned} \tag{7.19}$$

To assess the stability of this steady state we have to determine whether a small change in the value of x_s to $x_s + \delta x$ and/or y_s to $y_s + \delta y$ will lead to dynamics where the system returns to x_s and y_s such that the steady state is stable or not in case of instability. Again we can write,

$$\begin{aligned} \frac{dx_s}{dt} + \frac{d\delta x}{dt} &= f(x_s,y_s) + \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y \\ \frac{dy_s}{dt} + \frac{d\delta y}{dt} &= g(x_s,y_s) + \frac{\partial g}{\partial x}\delta x + \frac{\partial g}{\partial y}\delta y \end{aligned} \tag{7.20}$$

All these derivatives are evaluated at $x = x_s$ and $y = y_s$. This set of equations can be simplified to,

$$\begin{aligned} \frac{d\delta x}{dt} &= \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y \\ \frac{d\delta y}{dt} &= \frac{\partial g}{\partial x}\delta x + \frac{\partial g}{\partial y}\delta y \end{aligned} \tag{7.21}$$

This set of equation is often written in terms of vectors and matrices,

$$\begin{pmatrix} \frac{d\delta x}{dt} \\ \frac{d\delta y}{dt} \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix}}_{\text{Jacobian matrix}} \cdot \begin{pmatrix} \delta x \\ \delta y \end{pmatrix} \tag{7.22}$$

The matrix that contains the partial derivatives is termed the Jacobian matrix. In order to assess stability we have to determine the eigenvalues of this equation. The properties of this matrix are sufficient to assess stability of a steady state. This means that for the steady state of interest the jacobian entries have been evaluated.

Exercises

1. Draw the reaction network and determine the jacobian matrices of the following dynamic systems:

   (a)

   $$\begin{aligned} \frac{dx}{dt} &= k_1^+ s - k_1^- x - k_2^+ x + k_2^- y \\ \frac{dy}{dt} &= k_2^+ x - k_2^- y - k_3^+ y - k_3^- p \end{aligned}$$

   (b)

   $$\begin{aligned} \frac{dx}{dt} &= k_1 y - k_2 \cdot x \cdot y \\ \frac{dy}{dt} &= k_3 - k_4 y - k_2 \cdot x \cdot y \end{aligned}$$

   (c) The Schnakenberg model (considered in the main text):

   $$\begin{aligned} \frac{dx}{dt} &= k_1^+ a - k_1^- x + k_3 x^2 y \\ \frac{dy}{dt} &= k_2 b - k_3 x^2 y \end{aligned} \tag{7.23}$$

   (d) The Brusselator:

   $$\begin{aligned} \frac{dx}{dt} &= k_1 a + k_2 x^2 \cdot y - k_3 b \cdot x - k_4 x \\ \frac{dy}{dt} &= k_3 b \cdot x - k_2 x^2 y \end{aligned} \tag{7.24}$$

(e) Determine the steady state of the Brusselator model in terms of kinetic parameters.

7.5.1 Analysis of the 2x2 jacobian matrix

Two properties of the jacobian matrix are very insightful when evaluating the stability properties of the steady state of a 2-variable dynamic system,

$$\begin{aligned} \text{trace}: \quad T &= \frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} \\ \text{determinant}: \quad D &= \frac{\partial f}{\partial x}\frac{\partial g}{\partial y} - \frac{\partial f}{\partial y}\frac{\partial g}{\partial x} \end{aligned} \tag{7.25}$$

The eigenvalues associated with a particular steady state can be expressed in terms of the trace and the determinant of the jacobian matrix (how to do this can be found in standard linear algebra books),

$$\begin{aligned} \lambda_1 &= \frac{1}{2}(T + \sqrt{(T^2 - 4D)}) \\ \lambda_2 &= \frac{1}{2}(T - \sqrt{(T^2 - 4D)}) \end{aligned} \tag{7.26}$$

The line $D = \frac{1}{4}T^2$ (derives from $T^2 - 4D = 0$) in the (T,D)-plane divides this plane into six regions (figure 7.8). Those regions define all the qualitatively different dynamics of the system around the steady state. Therefore, from these regions the stability of the steady state can be assessed and the kind of instability can be identified. This means that you can now classify all the types of steady state a two variable dynamic system can have. On the basis of the values of D and T determined from the jacobian matrix of a dynamic system; this jacobian matrix should be calculated at the steady state of interest.

Let's use the information in figure 7.8 by working out an example[3]. The example is given by,

$$\begin{aligned} \frac{dx}{dt} &= x(x(1-x) - y) = f(x,y) \\ \frac{dy}{dt} &= k(x - 1/\mu)y = g(x,y) \end{aligned} \tag{7.27}$$

This example is a bit artificial but will nonetheless indicate a number of useful points. First, we determine the nullclines. The nullcline given $dx/dt = 0$ corresponds to $y = x(1-x)$ and the line $x = 0$. The nullcline corresponding to $dy/dt = 0$ equals the lines $x = 1/\mu$ and $y = 0$. Solving $dx/dt = 0$ and $dy/dt = 0$ for x and y leads to three steady states: $(0,0)$, $(\mu^{-1}, \mu^{-1}(1-\mu^{-1}))$, and $(1,0)$. See figure 7.9.

Next, we will determine the stability properties of these steady states and whether they correspond to centers, spirals, or nodes. To achieve this we need to determine the jacobian matrix,

[3]Here we follow the appendix in *Mathematical models in molecular and cellular biology* edited by L.A. Segel.

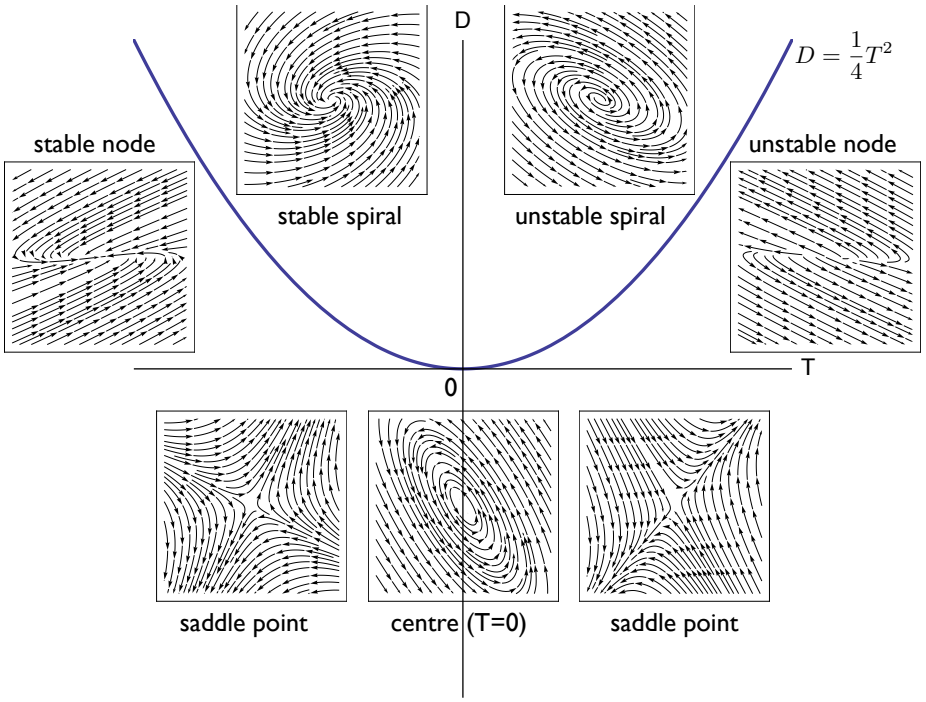

Figure 7.8: **Classification steady states and their stability.** The line $D = 1/4T^2$ and the regular axes divide the figure into six regions. Each of these regions corresponds to a particular steady state class, which differ in stability and the nature of the dynamics around the steady state point. The steady state point always lie exactly in the middle of the the inset plots, which display the qualitative dynamics. Clearly stability occurs when $D > 0$ and $T < 0$: two types of stable steady state are possible a stable node (or fixed point) and a stable spiral (damped oscillations). When the determinant changes sign from stable node to a saddle point, a saddle-node bifurcation occurs. This kind of bifurcation is associate with bistability. An unstable spiral is associated with oscillations.

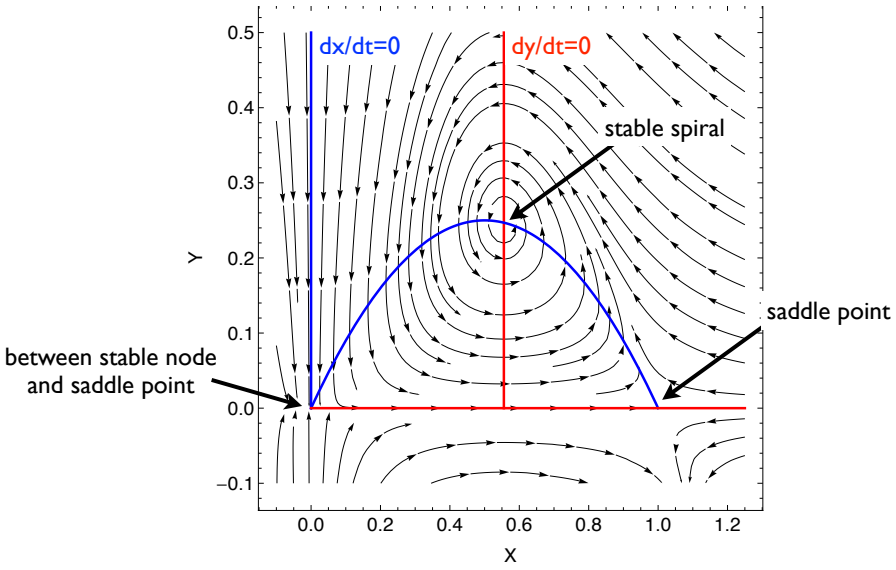

Figure 7.9: **Stability analysis of a complicated example system.** The blue and red lines correspond to the null cline that results from $dy/dt = 0$ and $dx/dt = 0$. Three intersections are found between the nullclines so three steady states occur. The arrows already hint at the nature of those steady states, a node, a centre and a saddle point. This is confirmed in the main text with a calculations of the determinant and the trace to be able categorize the steady states on the basis of figure 7.8.

$$\mathbf{M} = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix} = \begin{pmatrix} 2x - 3x^2 - y & -x \\ ky & k(x - \mu^{-1}) \end{pmatrix} \tag{7.28}$$

We choose $k = 1$ and $\mu = 1.8$. Evaluation of the jacobian matrix at the three steady states gives:

$$\begin{aligned} (0,0) \quad &\Rightarrow \quad \mathbf{M} = \begin{pmatrix} 0 & 0 \\ 0 & -0.56 \end{pmatrix} \Rightarrow \quad D = 0,\ T = -0.56 \\ (0.56, 0.25) \quad &\Rightarrow \quad \mathbf{M} = \begin{pmatrix} -0.062 & -0.56 \\ 0.25 & 0 \end{pmatrix} \Rightarrow \quad D = 0.13,\ T = -0.062 \\ (1,0) \quad &\Rightarrow \quad \mathbf{M} = \begin{pmatrix} -1 & -10 \\ 0 & 0.44 \end{pmatrix} \Rightarrow \quad D = -0.44,\ T = -0.56 \end{aligned} \tag{7.29}$$

Inspection of figure 7.8 then shows that the steady state $(0, 0)$ is lies on the $D =$ left from the origin ($T < 0$) and the state lies in between a stable node and a saddle point. For the second steady state at $(0.56, 0.25)$, $D > 1/4T^2$, the determinant is positive and the trace is negative, so this state is a stable spiral.

7.5.2 Exercises

1. Confirm the calculations done in the previous paragraph, determine yourself:
   (a) The nullclines
   (b) The steady states
   (c) The jacobian matrix
   (d) Determine the determinant and the trace of the jacobian matrix at the three steady states
   (e) Assess the type of steady state
   (f) Determine the eigenvalues of the steady state. When the real part of the eigenvalue (the number without the i in front, which indicates the imaginary part) is negative the steady state is stable.
2.

Bibliography

[1] A Becskei, B Séraphin, and L Serrano. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J*, 20(10):2528–2535, May 2001.

[2] H C Berg and E M Purcell. Physics of chemoreception. *Biophys J*, 20(2):193–219, Nov 1977.

[3] N Blüthgen, F J Bruggeman, S Legewie, H Herzel, H V Westerhoff, and B N Kholodenko. Effects of sequestration on signal transduction cascades. *FEBS J*, 273(5):895–906, Mar 2006.

[4] F J Bruggeman, J L Snoep, and H V Westerhoff. Control, responses and modularity of cellular regulatory networks: a control analysis perspective. *IET Syst Biol*, 2(6):397–410, Nov 2008.

[5] F J Bruggeman, H V Westerhoff, J B Hoek, and B N Kholodenko. Modular response analysis of cellular regulatory networks. *J Theor Biol*, 218(4):507–520, Oct 2002.

[6] D C Clarke and X Liu. Decoding the quantitative nature of tgf-beta/smad signaling. *Trends Cell Biol*, 18(9):430–442, Sep 2008.

[7] W W Cleland. The kinetics of enzyme-catalyzed reactions with two or more substrates or products. i. nomenclature and rate equations. 1963. *Biochim Biophys Acta*, 1000:213–220, 1989.

[8] O Decroly and A Goldbeter. Birhythmicity, chaos, and other patterns of temporal self-organization in a multiply regulated biochemical system. *Proc Natl Acad Sci U S A*, 79(22):6917–6921, Nov 1982.

[9] 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, Aug 2002.

[10] M T Flikweert, J P van Dijken, and J T Pronk. Metabolic responses of pyruvate decarboxylase-negative saccharomyces cerevisiae to glucose excess. *Appl Environ Microbiol*, 63(9):3399–3404, Sep 1997.

[11] A Goldbeter and D E Koshland. An amplified sensitivity arising from covalent modification in biological systems. *Proc Natl Acad Sci U S A*, 78(11):6840–6844, Nov 1981.

[12] A Goldbeter and R Lefever. Dissipative structures for an allosteric model. application to glycolytic oscillations. *Biophys J*, 12(10):1302–1315, Oct 1972.

[13] I Golding, J Paulsson, S M Zawilski, and E C Cox. Real-time kinetics of gene activity in individual bacteria. *Cell*, 123(6):1025–1036, Dec 2005.

[14] I V Grigoriev, Y A Makhnovskii, A M Berezhkovskii, and V Y Zitserman. Kinetics of escape through a whole. *Journal of Chemical Physics*, 116:9574, 2002.

[15] J H Hofmeyr and A Cornish-Bowden. The reversible hill equation: how to incorporate cooperative enzymes into metabolic models. *Comput Appl Biosci*, 13(4):377–385, Aug 1997.

[16] J H Hofmeyr and B G Olivier. The regulatory design of an allosteric feedback loop: the effect of saturation by pathway substrate. *Biochem Soc Trans*, 30(2):19–25, Apr 2002.

[17] J S Hofmeyr and A Cornish-Bowden. Regulating the cellular economy of supply and demand. *FEBS Lett*, 476(1-2):47–51, Jun 2000.

[18] C Y Huang and J E Ferrell. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc Natl Acad Sci U S A*, 93(19):10078–10083, Sep 1996.

[19] R U Ibarra, J S Edwards, and B O Palsson. Escherichia coli k-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, 420(6912):186–189, Nov 2002.

[20] 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, Mar 2000.

[21] 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, Oct 1999.

[22] B N Kholodenko, J B Hoek, H V Westerhoff, and G C Brown. Quantification of information transfer via cellular signal transduction pathways. *FEBS Lett*, 414(2):430–434, Sep 1997.

[23] D E Koshland, A Goldbeter, and J B Stock. Amplification and adaptation in regulatory and sensory systems. *Science*, 217(4556):220–225, Jul 1982.

[24] D E Koshland, G Némethy, and D Filmer. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*, 5(1):365–385, Jan 1966.

[25] D C LaPorte, K Walsh, and D E Koshland. The branch point effect. ultrasensitivity and subsensitivity to metabolic control. *J Biol Chem*, 259(22):14068–14075, Nov 1984.

[26] F J Nicolás, K De Bosscher, B Schmierer, and C S Hill. Analysis of smad nucleocytoplasmic shuttling in living cells. *J Cell Sci*, 117(Pt 18):4113–4125, Aug 2004.

[27] S V Popova and E E Sel'kov. Generalization of the model by monod, wyman and changeux for the case of a reversible monosubtrate reaction sr,tp. *FEBS Lett*, 53(3):269–273, May 1975.

[28] B Schmierer and C S Hill. Tgfbeta-smad signal transduction: molecular specificity and functional flexibility. *Nat Rev Mol Cell Biol*, 8(12):970–982, Dec 2007.

[29] B Schmierer, A L Tournier, P A Bates, and C S Hill. Mathematical modeling identifies smad nucleocytoplasmic shuttling as a dynamic signal-interpreting system. *Proc Natl Acad Sci U S A*, 105(18):6608–6613, May 2008.

[30] I H Segel. *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*. Wiley-Interscience, 1993.

[31] H Shankaran, D L Ippolito, W B Chrisler, H Resat, N Bollinger, L K Opresko, and H S Wiley. Rapid and sustained nuclear-cytoplasmic erk oscillations induced by epidermal growth factor. *Mol Syst Biol*, 5:332–332, 2009.

[32] A Singer, Z Schuss, D Holcman, and R S Eisenberg. Narrow escape. part i. *Journal of statistical physics*, 122:437, 2006.

[33] B Teusink, J Passarge, C A Reijenga, E Esgalhado, C C van der Weijden, M Schepper, M C Walsh, B M Bakker, K van Dam, H V Westerhoff, and J L Snoep. Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? testing biochemistry. *Eur J Biochem*, 267(17):5313–5329, Sep 2000.

[34] B Teusink and E J Smid. Modelling strategies for the industrial exploitation of lactic acid bacteria. *Nat Rev Microbiol*, 4(1):46–56, Jan 2006.

[35] B Teusink and H V Westerhoff. 'slave' metabolites and enzymes. a rapid way of delineating metabolic control. *Eur J Biochem*, 267(7):1889–1893, Apr 2000.

[36] K Walsh and D E Koshland. Branch point control by the phosphorylation state of isocitrate dehydrogenase. a quantitative examination of fluxes during a regulatory transition. *J Biol Chem*, 260(14):8430–8437, Jul 1985.